UPRT 2006

Empirical Studies in English Applied Linguistics

Edited by Marianne Nikolov and József Horváth

University of Pécs Roundtable 2006: Empirical Studies in English Applied Linguistics

Edited by Marianne Nikolov and József Horváth

Pécs, Lingua Franca Csoport

www.xanga.com/lingua_franca

Introduction

This edited volume is the outcome of a conference called *University of Pécs Roundtable 2006: Empirical Studies in English Applied Linguistics.* The event was held on January 20, 2006 at the Department of English Applied Linguistics, Faculty of Humanities, University of Pécs with thirty enthusiastic participants from five countries. It is our pleasure to make the written papers available to a wider audience. As you will see, the contributions cover a wider range of themes in applied linguistics and language pedagogy.

The first five papers discuss what roles individual differences, more exactly, motivation, anxiety, creativity and working memory play in foreign language learning. The study by Jelena Mihaljević Djigunović examines the relationships between affective factors and the productive skills of year 8 and 12 Croatian learners. The qualitative enquiry conducted by Zsuzsa Tóth explores Hungarian English majors' perceptions of the effects of foreign language anxiety on their oral performance. Katalin Piniel's study uses mixed methodology to tap into classroom foreign language classroom anxiety of Hungarian secondary-school EFL learners, whereas Polish dyslexic language learners' foreign language anxiety is the focus of Ewa Piechurska-Kuciel's longitudinal study. Two further studies involved Hungarian students: the relationships between creativity, language aptitude and English language proficiency are in the focus of Ágnes Albert's research on advanced learners, whereas the role of working memory is examined in intensive programmes of secondary-school students' groups by Judit Kormos and Anna Sáfár.

The next four studies enquire into the learning of foreign language vocabulary and rules. Brigitta Dóczi's study maps the mental lexicon of pre-intermediate learners, while Magdolna Lehmann discusses how advanced learners' vocabulary is tested in a proficiency exam. The effects of hypertextual input are examined in Gyula Sankó's longitudinal study involving intermediate level secondary-school learners. Ildikó Furkóné Banka's paper addresses the issues of parameter resetting and the role of transfer.

Assessment projects are in the centre of the next three papers. Marianne Nikolov and Krisztián Józsa examine relationships between Hungarian learners' achievements in English and German and classroom-related variables. Gloria Vickov gives an account of an experimental study aiming to examine young Croatian learners' writing skill in English and the mother tongue. Gábor Szabó analyzes the impact of anchor items' number and difficulty range on item difficulty calibrations in use of English tests.

The last six papers examine classroom processes and teachers' development. Francis J. Prescott's classroom study explores the importance of the group in a first-year university class. Teachers' discourse is analyzed and compared in two corpora collected in Croatian primary and secondary schools by Sanja Čurković Kalebić. Gordon Dobson's classroom research study examines how exploratory practice has contributed to his own professional development in tertiary education. The role of cooperation in pre-service education and in-service development is examined in Stefka Barócsi's

qualitative study. Réka Lugossy provides insights into how teachers' beliefs developed about using stories in their English classes, and finally, Éva Szabó examines the process of Hungarian school teachers' lesson planning practices.

Marianne Nikolov & József Horváth

Role of Affective Factors in the Development of Productive Skills

Jelena Mihaljević Djigunović

Faculty of Philosophy, University of Zagreb, Croatia

jmihalje@ffzg.hr

Introduction

In the past, it was often considered that language learning was primarily linked to the learner's cognitive abilities to understand, reproduce and create messages in a way intelligible to other speakers of that language. By now, however, not only have the competences been redefined, but it is commonly accepted nowadays that during the foreign language (FL) learning process both cognitive and affective learner qualities are activated.

Affect

Affective aspects of FL learning are a complex area whose importance is now well established. The number of affective learner factors considered in research is on the increase. New learner emotional characteristics are emerging as potentially important in order to understand and explain the process of language learning.

Affective learner characteristics started to be more systematically studied and measured rather late (from mid twentieth century). They were more difficult to define and measure because they seemed to be more elusive as constructs. Interest in the affective aspects of learning was prompted, among other things, when it was realised that the whole personality of the learner needs to be involved in education and that learners do not automatically develop emotionally as they may intellectually. Affect came to be considered as a very important contributing factor to success in learning. Some even went so far as to stress that affect was more important than cognitive learner abilities because without, say, motivation to learn cognitive learner abilities would not even start to be engaged in the process of learning.

In the next few subsections we will touch upon only those affective learner characteristics that we included in our research study. These are attitudes and motivation, anxiety and self-concept.

Attitudes and motivation

The importance of attitudes and motivation in FL learning is not questioned any more. Numerous studies (e.g., Dörnyei, 2001; Gardner, 1985; Lambert & Gardner, 1972; Mihaljević Djigunović, 1998) have confirmed that it is not possible to fully understand what happens in FL learning or to interpret research results without taking them into consideration. In fact, besides language learning aptitude, motivation is considered to be the best predictor of FL achievement.

In contemporary theories of language learning, attitudes are taken as a basis on which motivation for learning is formed or established. Attitudes are commonly defined as acquired and relatively durable relationships the learner has to an object. Lambert and Gardner (1972) differentiate between several groups of attitudes connected to language learning motivation: attitudes towards the community whose language is being learned; attitudes towards the FL classes, towards the FL teacher, towards language learning as such etc. While Lambert and Gardner take the view that, of all the relevant types of attitudes, those that refer to the FL community and its speakers are the most responsible for FL learning motivation, other researchers (e.g., Dörnyei, 2001; Nikolov, 2002) stress that in FL learning contexts attitudes towards different aspects of the teaching situation take precedence.

Recent trends in motivational research seem to be rooted in a broader perspective, in what Dörnyei (1994) describes as the language level, the learner level and the learning situation level. Motivation is increasingly approached as a multi-faceted construct (e.g., Clément & Gardner, 2001; Csizér & Dörnyei, 2005; Dörnyei , 2005, Dörnyei & Ottó, 1998, Ushioda, 2003), that is as a phenomenon that includes trait-like, situation-specific and state-like elements and that changes during the different stages of the language learning process. Also, like other individual difference variables, motivation is nowadays seen as a learner characteristic that interacts with other individual difference variables, as well as with contextual factors.

Anxiety

There are different approaches to the phenomenon of language anxiety (Scovel, 1991). According to one, it is essentially a manifestation of more general types of anxiety such as communication apprehension, test anxiety or of apprehensiveness as a personality trait. According to a different approach, language anxiety is quite a distinct type of anxiety. The fact that some of the first studies on the effect of anxiety on SLA produced conflicting results prompted a need to distinguish between different types of anxiety. While it is true that conceptual foundations for the phenomenon are provided by the concepts of communication apprehension, test anxiety and fear of negative social evaluation, it is nowadays widely accepted that language anxiety is "a distinct complex of self-perceptions, beliefs, feelings, and behaviours related to classroom language learning arising from the uniqueness of the language learning process" (Horwitz, Horwitz & Cope, 1991, p. 31).

MacIntyre and Gardner defined it as 'the feeling of tension and apprehension specifically associated with second language contexts, including speaking, listening and learning' (1994, p. 284).

Determining the causal direction in the negative relationship between anxiety and language achievement has engendered lively debate (Horwitz, 2000; MacIntyre, 1995a, 1995b; Sparks & Ganschow 1995, 2000). The basic issue has been whether anxiety causes poor performance or poor performance causes anxiety. On the one hand, the consistent negative relationship between language anxiety and language achievement in numerous studies has been explained as pervasive effects of language anxiety on cognitive processing (e.g., MacIntyre & Gardner, 1994). Anxiety arousal is thought to be associated with self-related thoughts that compete with task-related thoughts for cognitive resources. Due to the fact that information processing capacity in humans is limited, the self-related cognition emerges as a distracter or hindrance during cognitive performance. On the other hand, some experts (Sparks & Ganschow, 1991, 1993, 1995, 2000; Sparks, Ganschow & Javorsky, 1995) believe that language aptitude causes difficulties in linguistic coding in L1 (particularly in the coding of its phonological and syntactic aspects), which causes FL learning difficulties, which then give rise to anxiety. Anxiety, like other affective variables, is then the consequence and not the cause of poor FL performance.

Self-concept

Another learner factor we focused on in our study is self-concept. It is usually defined as a store of self-perceptions that emerge through experiences and reflect the perceived reactions of other people (Laine, 1987). Authors usually distinguish the following three aspects:

- the actual self – a person's notions, beliefs and cognitions of what he or she actually is
- the ideal self – what we would like to be, reflects our wants and aspirations, defines our goals for the future; an optimal discrepancy may contribute to one's motivation
- the social self – the way we perceive other people see us.

Aspects of self-concept have been shown to be connected with learning achievement (Burns, 1982; Sinclair, 1987). Self-concept changes with age (Wittrock, 1986). It is also related with attributions of success and failure in language learning. Ushioda (1996) has stressed the great value of the ability of positive motivational thinking, which helps the learner to maintain a positive self-concept in spite of negative experiences during language learning.

Productive language skills

Speaking

Most learners consider speaking the most important language skill. Researchers (e.g., Bygate, 2002) often describe it as a complex and multilevel skill. Part of the complexity is explained by the fact that speakers need to use their knowledge of the language and activate their ability to do this under real constraints.

Psycholinguistic models of speech production, focusing on ways in which speakers plan and monitor their speech production, recognize that variability is both socially and psycholinguistically motivated. In Levelt's model of speech production (1983) the socially motivated variability is connected to message generation in the 'conceptualiser', while the psycholinguistically motivated sources of variability are present at all levels: in the 'conceptualiser' (when speakers decide which language variety to use and which communicative intentions they want to realize through speech); in the 'formulator' (where the 'pre-verbal message' is turned into a speech plan through word selection and application of grammatical and phonological rules); in the 'articulator' (where the created speech plan is converted into actual speech); and in the 'speech comprehension system' (which offers speakers feedback on the basis of which they can make the necessary adjustments in the 'conceptualiser'). It is argued that while first language (L1) production is to a large extent automatic, second language (L2) production in general is not. This is why research on L2 variability has often concentrated on the effect of 'planning time'. It is assumed that, generally, L2 speakers need more time to plan the processing stages and this is highly likely to affect L2 speech. During speech production, speakers may pay conscious attention to different utterance elements so that they could improve them.

A lot of controversies still surround the teaching of speaking skills although a lot of research has already been done on speaking both within the second language acquisition field and theory of language teaching. Different approaches (Brumfit, 1984; Littlewood, 1981; Skehan, 1998; Widdowson, 1998) to how the different parts of the speaking skill hierarchy should be practised show that a lot more work needs to be done before a general agreement is reached.

Writing

The importance of writing in FL learning has been perceived differently throughout history. In the past it was only viewed, as Rivers (1968) nicely put it several decades ago, as a *handmaid* to the other language skills, it was considered to be useful for reinforcing the knowledge of vocabulary and grammar acquisition. It has gone from not even being viewed as a skill which should be taught to a highly important skill which gives us access to knowledge, power and resources. Recently, writing has been recognized as a skill that is an important and compulsory part of FL teaching for which teachers, as Silva (1993) points out, need more training in order to teach well.

In the past 20 years or so, a number of authors (e.g,: Cumming, 1989; Raimes, 1985; Zamel, 1983) have investigated various aspects of the writing skills. The most relevant finding points to the need of reaching a threshold level of proficiency in the FL before FL learners can engage the efficient processes they use while writing in L1.

Study of the relationship between affect and productive skills

The study to be reported here was carried out as part of a national project called *English in Croatia* that started in 2003. The project aimed to find out about the communicative competence levels Croatian learners of English as a foreign language (EFL) achieved by the end of primary and by the end of secondary education. More than 2,000 learners' competence was tested using communicative tests developed and validated in Hungary in 2002 and piloted in Croatia to check their validity for the new context. The Hungarian tests were used because they were considered potentially valid for the Croatian context, due to many socio-educational similarities, and because the same tests would also allow comparisons between the two neighbouring countries.

The overall findings of the project point to a good mastery of EFL by both primary and secondary school Croatian EFL learners at the level of the receptive skills. In listening comprehension and writing both primary and secondary school participants performed above the expected *Council of Europe Framework of Reference (CEFR,* 2001) levels. The results for the productive skills were, however, less impressive. Below are the general descriptors for the two levels of communicative language competence that Year 8 and Year 12 Croatian EFL learners are expected to reach. Year 8 learners are expected to be at the A2 CEFR level and Year 12 learners should reach the B1 CEFR level.

A2 level:
Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his\her background, immediate environment and matters in areas of immediate need. (*CEFR,* 2001, p. 24).

B1 level:
Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans. (*CEFR,* 2001, p. 24)

During the testing of communicative language competence, learners' affective characteristics were also measured and this allowed a look at the relationship of affect and development of productive skills.

Aim of the present study

The aim of the present study was to look into the relationship of affective learner characteristics and development of speaking and writing competence of Croatian learners of EFL. Although many studies have been carried out in order to see the relationship of affect and language achievement, most have considered success in language learning as a general construct including all the skills. It is our belief that it might be useful to look into this relationship differentially, since language learners are often stronger in some skills than others.

By considering this relationship in Year 8 and Year 12 we hoped to not only get an insight into the relationship itself, but also to be able to conclude about the development of speaking and writing skills with reference to affect.

Participants

A total of 2,086 EFL learners participated in the study. There were 1,430 Year 8 and 656 Year 12 participants. These two years were chosen as they represent the school-leaving years. Year 8 is the final year of primary education in Croatia, when students transfer to secondary education or leave the education system altogether. Year 12 marks the end of secondary education after which students look for a job or go on to the university. The Year 8 sample was drawn from village, small town and big town schools. Year 12 participants came from small and big town schools. The number of learners that took part in various parts of the testing varied, though, since the testing was done in three turns per class. Their communicative competence in English was tested by means of a battery of tests consisting of two test booklets (one on reading comprehension and one on listening comprehension and writing) and a speaking test.

Instruments

Measures of affect

In order to collect data on the affective profile of learners a 13-item questionnaire (see Appendix) was used. Each item was accompanied by a 5-point Likert scale. The instrument was designed and validated in Hungary and piloted in Croatia before it was used in the project. The 13 items elicited information on the following: attitudes to English, attitudes to EFL classes, motivation, self-concept and language anxiety. The scale was homogeneous, with $\acute{\alpha}$ = .833 and $\acute{\alpha}$= .787 for Year 8 and Year 12 respectively.

Measures of speaking skills

The oral tests consisted of three tasks. The first two tasks were the same for both groups of participants.

Task 1 lasted 2-3 minutes and consisted of the interlocutor asking nine questions: the first three were general questions (*What's your name? Could you spell your name, please? How old are you?*), the other six could be selected from the remaining nine. In Task 2 participants were first to choose one of six pictures spread out on the table, describe it and explain the similarities and differences between the scene in the picture (e.g., a busy street, a garden) and the same place in their own life. The task lasted 4-5 minutes.

In Task 3, which also lasted 4-5 minutes, Year 8 participants were to choose two of six situations and act them out with the interlocutor. For example:

Your friend is coming to visit you. Give him/her directions from the nearest station or bus stop to your home.

You would like to cook something nice with your friend. Discuss what you like or dislike and why.

In the first situation the interlocutor initiated conversation, while in the second one the interviewee was to initiate it.

In their Task 3, Year 12 participants were asked to choose one of five offered statements and say why they agree or disagree with it. The statements referred to issues (e.g., using mobile phones or watching soap operas) that young people have strong feelings about.

The oral test was administered to six students from each school. The interlocutors were trained prior to going to the schools. The interviews were carried out individually and audiotaped. The test lasted for up to 15 minutes and was strictly structured timewise.

Measures of writing skills

Year 8 participants were asked to describe two pictures by writing about the ten differences in the pictures. Prompts were given on what to describe.

Year 12 participants were asked to write a letter to the editor of a youth magazine and give reasons why their friend should get the best friend award. The letter was supposed to include about 150 words and there were five subtopics that had to be included.

Procedure

Writing tests were administered to whole classes, while oral tests were done on an individual basis, out of class, and with only six students randomly chosen from each school.

Assessment of speaking and writing performance was done by means of specially designed assessment scales. The speaking assessment scale was constructed along the following criteria: task achievement, vocabulary, accuracy and fluency, pronunciation and intonation. The scale included five bands (0-4). The writing assessment scale comprised the following criteria: task achievement, vocabulary, accuracy and text structure. There were five bands, four of which included double scores (0, 1-2, 3-4, 5-6, 7-8).

The assessors of both writing and speaking were trained. Since such training has to focus on the actual tasks, four sets of training (two for speaking and two for writing) were conducted. Length of the training depended on how much time the assessors needed to standardize their criteria.

Results

In this section we will first present descriptive statistics for measures of affect, speaking skills and writing skills. Then we will look into correlations between affect and the assessed aspects of speaking and writing.

Descriptive statistics

The affective profile of the two age groups of participants did not show much difference. Both groups tended to be positive about EFL learning and about themselves as language learners (see Table 1).

Table 1: Means and standard deviations in affect for Year 8 and Year 12 participants

	Affect (max. 65)
Year 8	M=48.59
(n=1,316)	SD=9.19
Year 12	M=49.51
(n=531)	SD=7.59

The means on orals tests show that Year 8 participants scored highest on Task 1 and lowest on Task 3, while Year 12 participants also scored highest on Task 1 but lowest on Task 2, as presented in Table 2. Overall, younger participants showed better results on the speaking test.

Table 2: Means and standard deviations in three oral tasks for Year 8 and Year 12 participants

	Task 1 (max. 16)	Task 2 (max. 16)	Task 3 (max. 16)	Total (max. 16)
Year 8 (n=191)	M=14.52 SD=2.54	M=13.15 SD=3.33	M=12.96 SD=3.71	M=40.64 SD=9.08
Year 12 (n=80)	M=12.80 SD=3.30	M=11.92 SD=3.53	M=12.13 SD=3.54	M=36.91 SD=9.82

In terms of the assessed aspects of EFL oral competence (Table 3), both groups scored higher on task achievement and fluency than on vocabulary and, particularly, than on accuracy. Year 8 participants scored higher on all the four aspects than Year 12 participants.

Table 3: Means and standard deviations in four aspects of speaking competence for Year 8 and Year 12 participants

	Task achievement (max. 12)	Vocabulary (max. 12)	Accuracy (max. 12)	Fluency (max. 12)
Year 8 (n=191)	M=10.41 SD=2.28	M=10.27 SD=2.46	M=9.87 SD=2.33	M=10.08 SD=2.36
Year 12 (n=80)	M=9.68 SD=2.39	M=8.95 SD=2.66	M=8.72 SD=2.64	M=9.47 SD=2.55

Year 8 participants showed a higher total score on the writing test than Year 12 students, as Table 4 shows.

Table 4: Means and standard deviations in writing task for Year 8 and Year 12 participants

	Writing total (max. 32)
Year 8 (n=1,306)	M=18.23 SD=8.45
Year 12 (n=421)	M=16.94 SD=9.05

As can be seen in Table 5, Year 8 participants were best at task achievement and worst at composing the text. While Year 12 participants also scored highest in task achievement, their text composing skill was not the least developed aspect of their writing skills; the biggest problem for them was accuracy.

Table 5: Means and standard deviations in four aspects of writing competence for Year 8 and Year 12 participants

	Task achievement (max. 8)	Vocabulary (max. 8)	Accuracy (max. 8)	Text (max. 8)
Year 8 (n=1,306)	M=5.39 SD=2.40	M=4.55 SD=2.22	M=4.28 SD=2.19	M=4.01 SD=2.14
Year 12 (n=421)	M=4.52 SD=2.36	M=4.31 SD=2.46	M=3.97 SD=2.30	M=4.24 SD=2.25

4.5.2 Correlations

Correlation coefficients were computed between scores on the affect measure and on the speaking and writing tests and the individual aspects of the two skills.

Table 6: Correlations between affect and totals on speaking and writing tests

	Speaking	Writing
Affect Year 8	.57*	.58*
Affect Year 12	.50*	.39*

*p< .01

As can be seen from Table 6, the computed correlations are higher for both speaking and writing scores in the Year 8 group. The difference is especially prominent in the case of writing.

When correlations were computed separately for the three oral tasks (Table 7), a different pattern emerged in the two groups. With Year 8 participants the coefficients did not range as widely as with Year 12. In the younger group affect showed the highest connection with picture description, while this was the just the opposite with the older group, where this presented the weakest relationship. Year 12 participants showed the highest correlation between affect and argumentative talk. In each of the three oral tasks the coefficients were lower in Year 12 than in Year 8.

Table 7: Correlations between affect and totals for three speaking tasks

	Task 1	Task 2	Task 3
Affect (year 8)	.52**	.57**	.53**
Affect (year 12)	.47**	.45*	.51

*p<.01; **p<.001

The relationship of affect and the individual aspects of writing is, generally, also stronger in Year 8 than in Year 12 (see Table 8). In Year 8, the strongest relationship of affect was found with task achievement and vocabulary use. In Year 12, quite interestingly, the strongest relationship was found with accuracy.

Table 8: Correlations between affect and four speaking aspects

	Task Achievement	Vocabulary	Accuracy	Fluency
Affect (year 8)	.57**	.57**	.52**	.55**
Affect (year12)	.48**	.44**	.52**	.49**

**p<.001

In the case of writing aspects, coefficients were higher for Year 8 than for Year 12. Affect in younger participants was more strongly connected with accuracy and successful vocabulary use than with task achievement and text-composing skills. With Year 12 participants the same pattern emerged, only – as already mentioned – the coefficients were lower in each case than those of Year 8 participants. These coefficients are presented in Table 9.

Table 9: Correlations between affect and four writing aspects

	Task Achievement	Vocabulary	Accuracy	Coherence
Affect (year8)	.53**	.57**	.58**	.52**
Affect (year12)	.26**	.39**	.42**	.37**

**p<.001

Conclusions

On the basis of the results we obtained it can be concluded that there is a significant relationship between affect and productive skills of speaking and writing. It is particularly prominent in Year 8 learners. With Year 12 learners there seems to be an important difference in significance between the two skills: success in speaking seems to be more strongly related to affect than success in writing skills.

With respect to the type of speaking activity it is interesting to note that with older and more proficient learners (Year 12) success in argumentative talk was more highly correlated with affect than the less complex activities of answering questions and picture description. Year 8 learners did not show differences in this respect and, overall, it seems that success in all types of speaking activities in their group was connected with positive affect.

In terms of the four assessed aspects of speaking Year 8 learners, again, did not show much difference in the strength of the relationship of affect and individual aspects of speaking skills. With older learners the differences seem prominent and we can conclude that affect was most highly correlated with accuracy, the aspect

that these learners had the lowest success in. It is also interesting that with Year 12 learners affect was less strongly related than with Year 8 learners to all speaking skill aspects except accuracy.

As has been mentioned, success in writing was significantly less correlated with affect in the Year 12 group than with Year 8 learners. If we consider the four criteria that writing was assessed along, we can conclude that in both groups the highest correlation with affect was found where the scores were low: with accuracy.

Our findings seem to point to two general conclusions about the relationship between affect and success in productive language skills. The first conclusion may be considered to be of developmental nature: the relationship is stronger for younger and less proficient learners. The second conclusion is connected to the complexity and difficulty of using productive skills: affect is more strongly connected with more complex activities.

If we interpret the relationships evidenced by the significant correlation coefficients in terms of affect as a cause of success, the teaching implications of these findings are quite apparent: we should help FL learners to create and maintain a positive affective profile.

Implications for further study

In the present study we used an instrument for measuring the general affective profile of participants. Perhaps more meaningful relationships could be obtained if, along with such an instrument, more state-like measures are taken as well. A particular task may be more or less motivating for a learner and trait and situation-specific items may not tap into state motivation, or state anxiety, that may also significantly influence learner behaviour and success on the task.

It might also be a good idea to get information on how learners themselves assess their performance on tasks. Insights into which aspects of speaking and writing performance learners find more or less difficult may be very valuable for conclusions about the role and impact of affective factors in language learning.

Since our younger learners (Year 8) had also learnt EFL for a shorter period of time than older ones (Year 12), it is possible that the developmental nature in our first conclusion in fact reflects the stage of language learning and not age. Which of the two possibilities is true could be found by studying participants of the same age but at different stages of learning.

References

Brumfit, C. J. (1984). *Communicative methodology in language teaching*. Cambridge: Cambridge University Press.

Burns, R. (1982). *Self-concept development and education*. London: Holt, Rinehart & Winston.

Bygate, M. (2002). Speaking. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 27-38). Oxford: Oxford University Press.

Clément, R., & Gardner, R. C. (2001). Second language mastery. In H. Giles & W. P. Robinson (Eds.), *The new handbook of language and social psychology* (2nd ed., pp. 489-504). London: Wiley.

Common European framework of reference. (2001). Cambridge: Cambridge University Press.

Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning*, *39*, 81-141.

Csizér, K., & Dörnyei, Z. (2005). The internal structure of language learning motivation and its relationship with language choice and learning effort. *Modern Language Journal*, *89*, 19-36.

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *Modern Language Journal*, *78*, 273-84.

Dörnyei, Z. (2001). *Teaching and researching motivation.* London: Longman.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Erlbaum.

Dörnyei, Z., & Otto, I. (1998). Motivation in action: A process model of L2 motivation. *Working Papers in Applied Linguistics 4*, 43-69.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold Publishers.

Horwitz, E.K. (2000). It ain't over 'til it's over: On foreign language anxiety, first language deficits, and the confounding of variables. *Modern Language Journal*, *84*, 256-259.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal*, *70*, 125–32.

Julkunen, K. (1989). *Situation- and task-specific motivation in foreign language learning and teaching*. Joensuu: University of Joensuu.

Laine, E. J. (1987). Affective factors in foreign language learning and teaching: a study of the "filter." *Jyvaskyla Cross-Language Studies, No. 13.*

Gardner, R.C., & Lambert (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*,*14*, 41-104.

Littlewood, W. T. (1981). *Communicative language teaching: An introduction*. Cambridge: Cambridge University Press.

MacIntyre, P.D. (1995). How does anxiety affect second language learning? A reply to Sparks and Ganschow. *Modern Language Journal*, *79*, 90-99.

MacIntyre, P.D. (1995). On seeing the forest and the trees: a rejoinder to Sparks and Ganschow. *Modern Language Journal*, *79*, 245-248.

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, 44, 283-305.

Mihaljević Djigunović, J. (1998). *Uloga afektivnih faktora u učenju stranoga jezika.* Zagreb: Filozofski fakultet Sveučilišta u Zagrebu.

Nikolov, M. (2002). *Issues in English language education.* Bern: Peter Lang.

Raimes, A. (1985). What unskilled ESL students do as they write?: A classroom study of composing. *TESOL Quarterly*, *19*(2), 229-258.

Rivers, W. (1968). *Teaching foreign language skills.* Chicago: University of Chicago Press.

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing. *TESOL Quarterly, 27* (4), 657-677.

Scovel, T. (1991). The effect of affect on foreign language learning: A review of the anxiety research. In E. K. Horwitz & D. J. Young (Eds.), *Language anxiety. From theory and research to classroom implications* (pp.15-23). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Sinclair, J. McH. (1987). Classroom discourse: Progress and prospects. *RELC Journal*, *18*, 1-14.

Skehan, P. (1998). Task-based instruction. In W. Grabe et al. (Eds.), *Annual review of applied linguistics 18: Foundations of second language teaching* (pp. 268-286). Cambridge: Cambridge University Press.

Sparks, R. L., & Ganschow, L. (1991). Foreign language learning difficulties: Affective or native language aptitude differences? *Modern Language Journal*, *75*, 3-16.

Sparks, R., & Ganschow, L. (1993). Searching for the cognitive locus of foreign language learning difficulties: Linking first and second language learning. *Modern Language Journal, 77*, 289–302.

Sparks, R. L. & Ganschow, L. (1995). A strong inference approach to causal factors in foreign language learning: A response to MacIntyre. *Modern Language Journal, 79*, 235-244.

Sparks, R. L., Ganschow, L., & Javorsky, J. (1995). I know one when I see one (Or, I know one because I am one): A response to Mabbot. *Foreign Language Annals, 28*, 479-487.

Sparks, R. L., Ganschow, L., & Javorsky, J. (2000). Déjà vu all over again: A response to Saito, Horwitz, and Garza. *Modern Language Journal, 84*, 251-255.

Ushioda, E. (1996). *Learner autonomy 5: The role of motivation*. Dublin: Authentik.

Ushioda, E. (2003). Motivation as a socially mediated process. In D. Little, J. Ridley, & E. Ushioda (Eds.), *Learner autonomy in the foreign language classroom: Teacher, learner, curriculum, assessment* (pp. 90-102). Dublin, UK: Authentik.

Widdowson, H. G. (1998). Skills, abilities, and contexts of reality. In W. Grabe et al. (Eds.), *Annual review of applied linguistics 18: Foundations of second language teaching* (pp. 323-333). Cambridge: Cambridge University Press.

Wittrock, M. C. (1986). Students' thought processes. In M. C. Wittrock (Ed.), *Handbook on research on teaching* (pp. 297-313). New York: Macmillan.

Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly, 17* (2), 165-186.

Appendix

Items in the Affective profile questionnaire:

1. I like English very much.
2. Knowledge of English is useless to me.
3. My parents think that it is important for me to know English.
4. People who speak English are interesting to me.
5. I'm interested in films and pop music in English.
6. I find English lessons extremely boring.
7. I have no feeling for languages, I'm a hopeless case for learning languages.
8. I find it easy to learn English.
9. It would take much more effort and will for me to be more successful at English.
10. No matter how much I study I can't achieve better results.
11. I like to use English in my free time.
12. I often fail while learning English.
13. I feel anxious when speaking English during English lessons.

First-Year English Majors' Perceptions of the Effects of Foreign Language Anxiety on their Oral Performance

Zsuzsa Tóth

Pázmány Péter Catholic University, Piliscsaba, Hungary

zsuzsat@yahoo.co.uk

Introduction

The paper reports on a qualitative study investigating the role of Foreign Language Anxiety (FLA) experienced by advanced-level learners during oral production of English. It examines Hungarian first year English majors' perceptions of how anxiety affected their performance in a semi-formal conversation with a native-speaker of English. The unique feature of the study is that it provides an emic perspective on how anxiety interferes with various cognitive processes implicated in second language (L2) speech production. The findings lend support to a cognitive interpretation of anxiety and are analysed from a psycholinguistic perspective.

A growing body of research since the 1980s has documented the fact that FL learning and communication induces anxiety in many learners, a distinct type of anxiety specifically related to second language (L2) contexts. This situation-specific anxiety termed Foreign Language Anxiety is assumed to affect learners' performance and achievement in the L2 (Gardner, 1985; Gardner & MacIntyre, 1993a; Horwitz, 1986; Horwitz, Horwitz & Cope, 1986; MacIntyre & Gardner, 1989, 1991a). Most research into the role of anxiety in language learning examined the relationship between learners' anxiety level and some *global* measures of TL achievement, usually operationalised as course grades, standardised tests, self- or teacher-ratings of proficiency, and the majority of studies reported a moderately negative relationship suggesting poorer performance for learners with high levels of anxiety (Gardner & MacIntyre, 1993a, b; Horwitz, 2001; MacIntyre & Gardner, 1991b; MacIntyre, 1995, 1999, 2002; Young, 1991, 1994).

Although studies investigating the sources and characteristics of FLA have revealed that, of all the L2 skills, it is speaking that learners consider as the most anxiety-provoking (Bailey, 1983; Cohen & Norst, 1989; Gregersen & Horwitz, 2002; Gregersen, 2003; Hilleson, 1996; Koch & Terrel, 1991; McCoy, 1979; Price, 1991; Young, 1990), relatively few studies looked at the effect of this anxiety specifically on *oral* performance. Early research examined the relationship between students' anxiety scores and some indicators of their TL *output*. Young (1986) and Scott (1986) found a negative relationship between anxiety level and oral test scores,

indicating that the oral performance of anxious students was weaker than that of learners with low anxiety. This result was replicated by Phillips (1992), who found negative correlations between students' anxiety scores and their oral exam grades. In addition, she also examined the effect of FLA on performance variables related to the quantity and quality of the oral output and found that students with high anxiety said less, produced shorter communication units, and used fewer dependent clauses and target structures than those whose anxiety level was lower. MacIntyre and Gardner (1994a) found similar differences in the output quality of learners who experienced various levels of anxiety. Anxious students in their study received lower ratings for their L2 fluency, sentence complexity, as well as accent. Finally, a study by Steinberg and Horwitz (1986), which aimed to investigate the more subtle effects of anxiety on the content of L2 speech, found that the oral descriptions of subjects in the anxiety arousal group were significantly less complex or interpretive than those of the more relaxed group.

An important line of research targeted more than just overt performance. In a series of laboratory studies MacIntyre and Gardner (1991a, 1994a, b) examined the effects of anxiety on specific *cognitive processes* involved in language learning and communication from an information-processing perspective, applying insights from general anxiety research to L2 learning (Tobias, 1979, 1980, 1986). They found negative correlations not only between anxiety and the performance of tasks specific to the output stage but also those specific to the input and processing stage, which they saw as evidence that anxiety can negatively affect not only learners' output but also the processing of FL input. MacIntyre and Gardner (1991a, 1994a, b) explained these findings from a cognitive psychological perspective (Sarason, 1980; Schwarzer, 1986). Within a cognitive framework, anxiety is conceptualised as a cognitive response marked by self-concern, feelings of inadequacy, worry, and self-blame. These task-irrelevant preoccupations are thought to represent cognitive interference, preventing the anxious individual from directing full attention to the task at hand. The processing of task-irrelevant information is believed to preempt processing resources and some of the limited capacity of working memory, making cognitive performance less efficient, which provides an explanation for anxiety's potential to hinder performance in cognitive tasks (Eysenck, 1979; Sarason, 1984; Wine, 1971, 1982). MacIntyre and Gardner (1991a, 1994a, b) propose that this theory is also able to explain the negative relationship observed between anxiety and FL performance. They assume that anxiety can reduce learners' capacity to receive and encode the target language, to organise, store, and assimilate the received input, as well as to hinder students' ability to use the FL.

As the review of research into anxiety's impact on oral performance shows, previous studies in this area tended to be quantitative inquiries. Whether they examined the effect of anxiety on output quality, or on cognitive processes associated with L2 learning, they did it from an etic (or outsider's) perspective. Consequently, what is currently known, or rather assumed about how anxiety affects FL speaking is largely based on the results of correlations between learners' anxiety scores and measures of their oral performance. What *learners themselves* think about the role of FLA in speaking their target language, however, appears to be a hitherto unresearched area. Therefore, the present study aims to fill this gap in the

literature by investigating the hypothesised effects of FLA on L2 speech production from an emic perspective: through learners' retrospections after conversing with a native-speaker of English. Exploring learners' feelings and self-perceptions about how anxiety affects FL speech production, the study aims to test the relevance of cognitive theories of anxiety in the domain of FL learning and performance.

Method

Participants

The participants of the study were 16 students selected from a larger group of 117 first-year English majors from a Hungarian university, identified as the eight most anxious and the eight least anxious students based on their scores on a Hungarian version of the Foreign Language Classroom Anxiety Scale (Horwitz et al., 1986; Tóth, 2003). The mean FLA score for the 117 students was 84.59 with a standard deviation of 19.34. The FLA scores of the eight high-anxious participants ranged from 108 to 136, while those of the eight low-anxious ones from 49 to 59. Of the 16 participants, 13 were female and 3 were male, and they had studied English for an average of 9.5 years. All participants were between the ages of 18 and 22. The interlocutor was a new lecturer from England, whom interviewees had never met before.

Procedures

Participants took part in a 10-15 minute, one-on-one conversation with the native speaking interlocutor. They performed three tasks of different nature such as (1) exchanging information about themselves and the interlocutor in a lead-in free phase, (2) expressing their opinion on a controversial issue presented on a situation card, and (3) describing and interpreting an ambiguous picture. The three tasks were designed to measure spontaneous communication-, argumentative-, and interpretive skills, respectively.

Following this, they were requested to think aloud: give vent to their feelings concerning the conversation, describe any thought that crossed their mind while performing the tasks, and reflect on their own performance. The post-task interviews were conducted in the participants native Hungarian and were tape-recorded with their consent. After the post-task interviews participants also completed a four-item questionnaire inquiring into their affective reactions to the semi-formal conversation.

Results and discussion

In the questionnaire students completed after the post-task interviews, they were directly asked whether they thought anxiety had any effect on their performance. Responses in the two anxiety categories turned out to show important differences. While in the high-anxiety group six out of the eight students said anxiety interfered with their performance, one felt it did not influence it, and one said it depended on the strength of her anxiety, which tended to fluctuate during the interview, answers in the low-anxiety group revealed just the opposite pattern: six out of the eight students felt anxiety did not influence their performance, one felt it interfered with it, and one said it was task-dependent. No student in either category said anxiety aided his/ her performance.

The post-task interviews provided an insight into interviewees' perceptions as to *how* anxiety affected their oral performance. In general terms, students tended to see anxiety as an impediment to successful performance, preventing them from performing to the best of their ability: as well as they ideally could were it not for their anxiety. As one of them put it:

> I feel I know much more than what I managed to show here, in a more relaxed state of mind it would have been all different. (Quotes from the interviews are given in the researcher's translation.)

Participants were also able to identify specific problems in their own oral performance which they attributed to anxiety. These were related to the following three processes involved in L2 speech production and comprehension: (1) comprehending FL input, (2) generating ideas, and (3) vocabulary retrieval.

Comprehending FL input

As for comprehension, students said anxiety hindered them in following the native speaking interlocutor's speech. As one student describes it:

> As I felt anxious, I didn't really understand what he was asking precisely, I think I would have understood him better if I had managed to stay calm.

One of the students with high anxiety even saw a direct link between her actual anxiety level at a given moment and her ability to understand the interlocutor. She said her anxiety mounted during the problem-solving task, as she felt she did not understand the prompt card, as a result of which, she "was simply unable to pay attention to the interlocutor's questions, and did not understand half of what he was trying to say". By contrast, when doing the picture description, which she felt more comfortable with, her anxiety "all but disappeared", and she had the impression she "understood his reactions completely, and what he was asking, too".

Problems with reading comprehension, specifically, understanding the hypothetical situation on the prompt card (Task 2), were also attributed to an anxiety effect. Anxious interviewees said the task of having to read a text and offer a personal opinion on it on the spot filled them with anxiety of such intensity that they found it difficult to concentrate on the task at hand. As evidenced by the next excerpt, feeling under pressure seemed to produce an adverse effect: the more quickly and efficiently they wanted to read and respond to the prompt, the less they succeeded in their efforts.

> I wanted to read it fast so that I wouldn't take up the interlocutor's time, and I tried hard to understand it, but I felt I just couldn't concentrate. All I had on my mind was wanting to understand it, in fact, 80 % of my attention was occupied by this, so it's not surprising I didn't get on. The words looked familiar, but I just didn't know what they meant.

What anxious students' comments appeared to have in common was that they felt anxiety interfered with their ability to concentrate, to pay attention to what is said or written in the FL, i.e. FL input. Students' perceptions seem to lend support to cognitive psychologists' conceptualization of anxiety in terms of cognitive and attentional processes, according to which anxiety functions as a distraction of attention or cognitive interference (Eysenck, 1979; Wine, 1971, 1982). As shown by the excerpt above, appearing in the form of self-related intrusive thoughts, anxiety kept this anxious student from directing full attention to the task, as she had her attention divided between task-related cognition (i.e. processing the text) and task-irrelevant cognition (like not wanting to take the interlocutor's time, wanting to do the task quickly, and worrying about the consequences of not being able to do so), which may have contributed to the difficulty she experienced in understanding the text.

Students' subjective feelings that anxiety hindered them in comprehending what is said or written in the FL can also be explained from a psycholinguistic perspective, relying on theoretical models of L1 vs. L2 speech production and comprehension (de Bot, 1992; Kormos, 2006; Levelt, 1989, 1993, 1995). According to these theories, one of the important differences between speech comprehension and production in the first vs. in second languages is that while in the L1, with the exception of conceptual processing (i.e. planning messages, self-monitoring, and discourse interpretation), the lower level processes involved in the production and understanding of speech (such as lexico-grammatical-, morpho-phonological-, phonetic *encoding* and articulation when producing speech, and acoustic-phonetic analysis, phonological and grammatical *decoding* when comprehending it) are automatic, they are not fully automatic in the L2, that is, they require conscious attention (Kormos, 2006; Levelt, 1989, 1993, 1995; Temple, 1997). Consequently, if learners' attentional capacity is diminished by anxiety, it is clear they will experience difficulty in comprehending FL messages, as with reduced attentional resources and working memory capacity the processing of the received input will be less efficient.

Let me return to the last sentence of the previous extract: "The words looked familiar, but I just didn't know what they meant". This comment shows that the student recognized the words as lexemes, i.e. word forms, but could not retrieve their sense or meaning (lemma level). Similar problems seemed to have been encountered by other anxious participants. One of them said she had to read the passage two or three times, as it took her a while to settle down and bring herself to grasp what the text was about.

> I didn't understand a word, when I read it for the first time. When I'm very anxious I don't understand what I'm reading, the letters are blurred before my eyes, I can't see words or sentences, just letters, then gradually I'm beginning to recognize words, but only one or two, they stick in my mind but I still don't understand anything, as if I hadn't read the text, just these one or two words, then after a while, when I go through the whole thing for the second or third time, the message is beginning to come through.

As evidenced by comments like this, anxiety interfered with the process by which the incoming letter or sound stream is converted into a meaningful interpretation through a complex series of mental operations, beginning with the segmentation of the input string into linguistic units and ending with the recovery of the intended message (Harrington, 2001, Tokowicz & Perfetti, 2005). The excerpt suggests that the anxious student experienced difficulty with word recognition/ lexical access, as anxiety slowed down the process of retrieving lexeme, lemma, and conceptual level information about words on the basis of the input letter string.

Generating ideas

The second problem area students ascribed to anxiety during the conversation was what they described as some kind of "failure" in cognitive functioning, which they realised as a momentary inability to (1) come up with ideas, (2) think up pros and cons for the controversial issue on the prompt card (task 2), or (3) give alternative interpretations of the ambiguous picture (task 3). They generally complained about, "going blank", "being unable to think", "feeling as if their mind was paralysed", "not being able to use their imagination". To give a sample comment:

> It was not merely a language problem: that I didn't remember how to say something in English, but that I had no idea whatsoever, nothing came to my mind about this paper [i.e. the situation card] at all, nor about the picture, I was so anxious about being able to say something that maybe this is what prevented me from coming up with anything.

The student's remark "it was not merely a language problem" shows that she felt anxiety interfered with her speech production well before the actual verbal phase, i.e. before actually giving a linguistic form to ideas. This phase corresponds to what is called *conceptualisation* in models of speech production, which is a pre-verbal/ pre-formulation stage, in which message generation takes place, i.e. the conceptualisation of what we want to say. Conceptualising/ message construction is believed to involve highly controlled processing (Levelt, 1989, 1993, 1995), i.e. it demands attentional resources. Anxious students' retrospective comments suggested that anxiety may have hindered the retrieval of information to be expressed, which was particularly evident during the argumentative and picture description tasks (No. 2 and 3).

> It was the picture task that I found the most difficult. It was O.K. just to describe it, but to interpret it, what's more, speculate on various possibilities as to what may be happening, who or what could be off the picture, etc., this would have required a lot of empathy and openness, the ability to put myself into others' shoes, but that's next to impossible in such a stressful situation, at least for me. I think you need to be very relaxed to be able to do that.

As evidenced by the sample comment, students felt the anxious state of mind they were in during the conversation was not conducive to deep and creative thinking, required by abstract or imaginary topics such as the controversial issue and the ambiguous picture in the second and third task, respectively. In contrast, talking about themselves in the lead-in task was perceived as less severely affected by anxiety. One of the anxious interviewees provided the following explanation for this:

> I didn't have to think, as it wasn't the first time I had talked about these things and I knew what to say, so I was thinking in English and spoke in English. As for the other two [i.e. the situation card and the picture], I was thinking in Hungarian and spoke in English, which is very difficult under stress.

The perceived difference in difficulty between talking about themselves vs. expressing an opinion on a controversial issue or interpreting an ambiguous picture can again be explained from a psycholinguistic perspective. The task of having to talk about themselves did not put as heavy demands on students in the conceptualisation phase as the other two tasks did, because students did not have to invent what to say, i.e. did not have to conceptualise their messages on the spot, and in all probability they were able to retrieve whole memorised phrases, clauses or even sentences in English from their long-term memory used on previous occasions when talking about themselves and getting information about someone else (which may account for the feeling "I was thinking in English and spoke in English"). In contrast, their verbal reactions to the situation card or the ambiguous

picture had to be conceptualised, planned, and constructed from scratch, which obviously required much more attention, consequently must have been affected by anxiety more severely.

Vocabulary retrieval

The interviews also provided evidence for anxiety's interference with lower level processes involved in the production of L2 speech, by which the conceptual/ pre-verbal message is encoded, i.e. translated into a linguistic form (*formulation* phase in models of speech production). The most frequently mentioned problem interviewees ascribed to anxiety concerned the retrieval of vocabulary items from the mental lexicon, in other words, finding the words students needed. To give a sample comment:

> When I'm in such an anxious state, words simply don't come. Actually, I think they don't come to mind because I'm afraid they won't.

Interviewees generally talked about a "mental block", due to which they "did not find"/ "forgot" words. Interestingly enough the words which tended to "slip their mind"/ "go out of their head" were what they called "very simple" or "basic" ones, such as *machine*, *drop*, or *library*, for instance, which they "definitely knew under normal circumstances", meaning, when they did not experience anxiety. To quote some remarks:

> What I find very embarrassing is that these are words you're supposed to know even if you're aroused from your sleep, the simplest words and phrases, still you won't recall them, which is extremely embarrassing.
> I have no idea as to how I couldn't remember a word like 'drop', it's very annoying, by the same token I could have forgotten any word [laughs…].

What is interesting about these mysterious forgettings of words is that the needed vocabulary items were either (1) incompletely, or (2) incorrectly retrieved from long term memory, which suggests that anxiety may have interfered at some point with the procedures for accessing them. An example for the first case is the following. One of the students with high anxiety could not recall the word *library*, so what she said was: "I want to work in a a a …… there are books in it." Her sentence nicely shows that she was able to activate the lemma corresponding to the concept she needed, which is evidenced by her correctly encoding it syntactically (the lexical gap in her sentence was preceded by a preposition and an indefinite article), however she was not able to retrieve the lexeme, i.e. the actual word form: *library*. As for the other case, not being able to recall vocabulary items in their appropriate forms, students said they felt as if their mind or memory "did not work properly", as sometimes they simply failed to find the words they were looking

for, even when they were very close to finding them. One student, for instance, said she wanted to use the word *researcher* (saying she wanted to become a researcher), she was absolutely certain she knew this word, still she couldn't recall it in its proper form.

> I felt *searcher* was not the right word, as there's no such word, but I just couldn't find the word *researcher*. I knew *searcher* was very close to what I was looking for, and also that I was searching in the right place, still I wouldn't find it [laughs…], although it would have been so logical: *search, research, researcher*, when he [i.e. the interlocutor] helped me out with *researcher*, it was so obvious this was the word I had been looking for.

The word *searcher* provides evidence for a lapse in the encoding process. The anxious speaker successfully accessed the lemma corresponding to the concept she wanted to verbalise (semantic and syntactic information was correctly retrieved), she even produced a word form very similar to *researcher*, but although she felt what she came out with was definitely not what she was looking for, she could not retrieve the accompanying lexeme fully and correctly.

The same extract, documenting an instance of pre-articulatory self-monitoring, shows that word search in this case was consciously attended to, rather than being an automatic process. The same is suggested by what another anxious student said about her not being able to recall the word *horseshoe*. She said she even knew the word she was trying to find had the name of the animal in it, which means, not only did she retrieve the lemma matching the selected concept, but she even accessed information about the inner composition of the word form (lexeme), as she knew it was a compound having the word *horse* in it. However, she said she "seemed to have gone blank and couldn't recall *horse* either", which she felt puzzled by, as this again was "such a basic word everyone knows". As vocabulary retrieval for these participants was not fully automatic, as indicated by the examples suggestive of word search being under conscious control, it is quite likely that anxious students, having their attentional resources diminished by anxiety, could not allocate enough attention to retrieving words from the mental lexicon, which explains the difficulties they reported to experience.

Conclusion

The article reported on the findings of a qualitative inquiry into the role of FLA in advanced learners' speech production, exploring what learners themselves thought about anxiety's role in their own performance in the context of communicating with a native speaker of English. It revealed that participants with high levels of FLA tended to feel anxiety hindered them in communicating with the native-speaking interlocutor, as it appeared to interfere with various aspects of their performance. They said anxiety made it difficult for them to concentrate, thereby adversely affecting their ability (1) to understand what is said or written in the FL,

(2) to generate ideas, to reason, and suggest hypotheses, as well as (3) to retrieve words.

As has been shown, the findings of the study lend support to a cognitive interpretation of anxiety and can be explained from a psycholinguistic perspective. Participants' retrospective comments provided evidence for anxiety manifesting itself in the form of task-irrelevant cognition, preventing them from paying full attention to the task of communicating in the FL. Having their attentional capacity diminished by anxiety, highly anxious students, in all probability, could allocate less attention to the execution of non-automatic processes involved in L2 speech production and comprehension. This provides a plausible explanation for the difficulties they experienced in conceptualising messages, as well as in encoding and decoding them in their L2.

Acknowledgements

References

Bailey, K. M. (1983). Competitiveness and anxiety in adult second language learning: Looking *at* and *through* the diary studies. In H. W. Seliger & M. H. Long (Eds.), *Classroom-oriented research in second language acquisition* (pp.67-102). Rowley, MA: Newbury House.

Cohen, Y., & Norst, M. J. (1989). Fear, dependence and loss of self-esteem: Affective barriers in second language learning among adults. *RELC Journal, 20*, 61-77.

de Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Pscycholinguistics, 13,* 1-24.

Eysenck, M. W. (1979). Anxiety, learning, and memory: A reconceptualization. *Journal of Research in Personality, 13*, 363-385.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation.* London: Edward Arnold.

Gardner, R. C., & MacIntyre, P. D. (1993a). On the measurement of affective variables in second-language learning. *Language Learning, 43*, 157-194.

Gardner, R. C., & MacIntyre, P. D. (1993b). A student's contributions to second-language learning. Part II: Affective variables. *Language Teaching, 26*, 1-11.

Gregersen, T. S. (2003). To err is human: A reminder to teachers of language-anxious students. *Foreign Language Annals, 36*, 25-32.

Gregersen, T. S., & Horwitz, E. K. (2002). Language learning and perfectionism: Anxious and non-anxious language learners' reactions to their own oral performance. *Modern Language Journal, 86*, 562- 570.

Harrington, M. (2001). Sentence processing. In P. J. Robinson (Ed.), *Cognition and second language instruction* (pp. 91-124). Cambridge: Cambridge University Press.

Hilleson, M. (1996). "I want to talk with them, but I don't want them to hear": An introspective study of second language anxiety in an English-medium school. In K. Bailey, & D. Nunan (Eds.), *Voices from the language classroom* (pp. 248-275). Cambridge: Cambridge University Press.

Horwitz, E. K. (1986). Preliminary evidence for the reliability and validity of a Foreign Language Anxiety Scale. *TESOL Quarterly, 20,* 559-562.

Horwitz, E. K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics, 21*, 112-126.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal, 70*, 125-132.

Koch, A. S., & Terrel, T. D. (1991). Affective reactions of foreign language students to natural approach activities and teaching techniques. In E. K. Horwitz, & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 109-126). Englewood Cliffs, NJ: Prentice-Hall.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, N.J.: Lawrence Erlbaum.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1993). Language use in normal speakers and its disorders. In G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, & C.-W. Wallesch (Eds.), *Linguistic disorders and pathologies* (pp. 1-15). Berlin: de Gruyter.

Levelt, W. J. M. (1995). The ability to speak: From intentions to spoken words. *European Review, 3,* 13-23.

MacIntyre, P. D. (1995). How does anxiety affect second language learning? A reply to Sparks and Ganschow. *Modern Language Journal, 79*, 90-99.

MacIntyre, P. D. (1999). Language anxiety: A review of the research for language teachers. In D. J. Young (Ed.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere* (pp. 24-45). Boston: McGraw-Hill.

MacIntyre, P. D. (2002). Motivation, anxiety and emotion in second language acquisition. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 45-68). Amsterdam/Philadelphia: John Benjamins Publishing Company.

MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second language learning: Toward a theoretical clarification. *Language Learning, 39*, 251-275.

MacIntyre, P. D., & Gardner, R. C. (1991a). Language anxiety: Its relationship to other anxieties and to processing in native and second languages. *Language Learning, 41*, 513-534.

MacIntyre, P. D., & Gardner, R. C. (1991b). Methods and results in the study of anxiety and language learning: A review of the literature. *Language Learning, 41*, 85-117.

MacIntyre, P. D., & Gardner, R. C. (1994a). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44*, 283-305.

MacIntyre, P. D., & Gardner, R. C. (1994b). The effects of induced anxiety on cognitive processing in the second language. *Studies in Second Language Acquisition, 16*, 1-17.

McCoy, I. R. (1979). Means to overcome the anxieties of second language learners. *Foreign Language Annals, 12*, 185-189.

Phillips, E. M. (1992). The effects of language anxiety on students' oral test performance and attitudes. *The Modern Language Journal, 76*, 14-26.

Price, M. L. (1991). The subjective experience of foreign language anxiety: Interviews with anxious students. In E. K. Horwitz, & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 101-108). Englewood Cliffs, NJ: Prentice-Hall.

Sarason, I. G. (1980). *Test anxiety: Theory, research and applications*. Hillsdale, NJ: Erlbaum.

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology, 46,* 929-938.

Schwarzer, R. (1986). Self-related cognitions in anxiety and motivation: An introduction. In R. Schwarzer (Ed.), *Self-related cognition in anxiety and motivation* (pp. 1-17). Hillsdale, NJ: Erlbaum.

Scott, M. L. (1986). Student affective reactions to oral language tests. *Language Testing, 3,* 99-118.

Steinberg, F. S., & Horwitz, E. K. (1986). The effect of induced anxiety on the denotative and interpretive content of second language speech. *TESOL Quarterly, 20*, 131-136.

Temple, L. (1997). Memory and processing modes in language learner speech production. *Communication and Cognition, 30* (1/2), 75-90.

Tobias, S. (1979). Anxiety research in educational psychology. *Journal of Educational Psychology, 71*, 573-582.

Tobias, S. (1980). Anxiety and instruction. In I. G. Sarason (Ed.), *Test anxiety: Theory, research and applications* (pp. 289-309). Hillsdale, NJ: Erlbaum.

Tobias, S. (1986). Anxiety and cognitive processing of instruction. In R. Schwarzer (Ed.), *Self-related cognition in anxiety and motivation* (pp. 35-54). Hillsdale, NJ: Erlbaum.

Tokowitz, N., & Perfetti, C. A. (2005). Introduction to Part II: Comprehension. In J. F. Kroll, & A. M. B. de Groot (Eds.), *Handbook of bilingualism* (pp. 173-179). Oxford: Oxford University Press.

Tóth, Zs. (2003). Foreign language anxiety and first-year English majors at Pázmány Péter Catholic University. Unpublished seminar paper. Eötvös Loránd University, Budapest.

Wine, J. D. (1971). Test anxiety and direction of attention. *Psychological Bulletin, 76,* 92-104.

Wine, J. D. (1982). Evaluation anxiety: A cognitive-attentional construct. In H. W. Krohne & L. Laux (Eds.), *Achievement, stress, and anxiety* (pp. 207-219). Washington, D. C: Hemisphere.

Young, D. J. (1986). Th.e relationship between anxiety and foreign language oral proficiency ratings. *Foreign Language Annals, 19,* 439-445.

Young, D. J. (1990). An investigation of students' perspectives on anxiety and speaking. *Foreign Language Annals, 23*, 539-553.

Young, D. J. (1991). Creating a low-anxiety classroom environment: What does language anxiety research suggest? *Modern Language Journal, 75*, 426-439.

Young, D. J. (1994). New directions in language anxiety research. In C. A. Klee (Ed.), *Faces in a crowd: The individual learner in multisection courses* (pp. 3-45). Boston: Heinle & Heinle Publishers.

Foreign Language Classroom Anxiety: A Classroom Perspective

Piniel Katalin

Eötvös University, Budapest, Hungary

piniel@freemail.hu

Introduction

The aim of the present research is to investigate foreign language classroom anxiety by questioning its stability across different foreign languages among secondary school students and to explore its causes stemming from the foreign language classroom. Participants of the study included 61 9th-year students of a grammar school in Hungary who are learning two foreign languages simultaneously. Validated Hungarian translation of standardized anxiety scales, namely the Foreign Language Classroom Anxiety Scale (FLCAS) of Horwitz, Horwitz & Cope (1991) and State Trait Anxiety Inventory (STAI) (Sipos, Sipos & Spielberger, 1994), were used to investigate the relationship between the differences of Foreign Language Classroom Anxiety (FLCA) experienced in the two different foreign language classrooms, taking into consideration and controlling for students' levels of trait anxiety. Interviews were conducted with six participants who do not possess high levels of trait anxiety, yet demonstrate high levels of anxiety in one foreign language classroom but not in the other. Findings suggest that FLCA may develop irrespective of low levels of trait anxiety and independently of the level of anxiety experienced in the other foreign language classroom. As such, the origins of FLCA may be sought among the classroom factors. In the particular cases presented here, the role of the instructor proved to be dominant in influencing students' levels of FLCA.

Learning at least one foreign language is compulsory in almost all secondary schools of Europe. Today's teenagers learn two or sometimes even three foreign languages simultaneously as obligatory subjects of the curriculum. Often, however, the case is that some students are successful in learning one of the foreign languages but are lagging behind their classmates in the other foreign language class. These learners, when attending lessons of the foreign language, feel less successful and may experience extensive nervousness while having to speak in front of their peers. Some may even have the urge to skip class rather than risk being laughed at. Others can become so anxious that they begin to sweat, and feel nausea. The aim of this research is to investigate this phenomenon known as Foreign Language Anxiety by questioning its stability across different foreign languages among secondary school students. Another objective of the study is to explore what learners perceive as possible causes of their feelings of FLCA.

Foreign language classroom anxiety

Foreign Language Anxiety, recognized as an affective factor in foreign language learning and normally discussed alongside other individual learner differences (Gardner & MacIntyre, 1992, 1993), is still considered to be a relatively new and developing area within foreign language research. In order to understand its nature, it is necessary to present an overview of anxiety in general and consider the different forms it may manifest itself in.

Different researchers have approached FLA from different aspects. From a broader perspective, anxiety itself is defined by psychologists as "the subjective feeling of tension, apprehension, nervousness, and worry associated with an arousal of the autonomic nervous system" (Spielberger, as cited in Horwitz, Horwitz & Cope, 1991, p. 27). Literature usually differentiates between three types of anxiety:

- trait anxiety – which is a personality trait (Eysenck, 1979)
- state anxiety – which is apprehension experienced at a particular moment in time
- situational anxiety – which is anxiety experienced in a well-defined situation (MacIntyre & Gardner, 1991a)

Foreign Language Anxiety, or more precisely, Foreign Language Classroom Anxiety (FLCA) is considered to be a situational anxiety experienced in the well-defined situation of the foreign language classroom (MacIntyre & Gardner, 1991a, 1991b, 1994). As such, Horwitz, Horwitz and Cope (1991) view FLCA as "a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning process" (in Horwitz & Young, 1991, p.31) In all of these specifications, the context or situation dependent nature of foreign language anxiety is emphasized.

According to Horwitz, Horwitz and Cope (1991) possible causes of FLCA are communication apprehension, test anxiety and fear of negative evaluation. MacIntyre and Gardner (1991a) describe FLCA stemming from the negative expectations in foreign language learning. On the other hand, Price (1991) and Piniel (2000) found classroom related factors to play a role in learners' developing FLCA. Correlational studies have also sought to establish relationships between individual learner variables and FLA. (It must be noted however, that these relationships do not imply cause-effect relationships.) FLA has been shown to correlate with age (Bailey, Onwuegbuzie & Daley, 2000), gender (Aida, 1994; Baker & MacIntyre, 2000), certain personality traits (such as perfectionism, Gregersen & Horwitz, 2002), and with negative self-perceptions (Onwuegbuzie, Bailey & Daley, 1999).

FLA itself can have contradictory effects on language learning, and as such literature mentions both debilitating and facilitating types of anxiety, where debilitating anxiety poses an obstacle to language learning, whereas facilitating anxiety facilitates or fosters it. Although both types exist, studies have mostly concentrated on the former (Scovel, 1978 in Horwitz & Young, 1991, p.15-23; MacIntyre & Gardner, 1991a; 1991b)

In spite of the growing number of research dealing with FLA, the majority of the studies mentioned above have involved the participation of mainly college or university students in a second language or a foreign language setting, in all cases dealing with the acquisition of one foreign/second language (Horwitz, Horwitz & Cope, 1991; Horwitz & Young, 1991; MacIntyre & Gardner, 1994). In other words, there are limited numbers of studies involving secondary school students albeit it is in this milieu that most learners are compulsorily introduced to studying foreign languages. Something else that has been given little attention are the possible differences between the levels of FLCA of one person studying two foreign languages simultaneously (Deweale, 2005; Piniel, 2000; Rodriguez & Abreu, 2003) Unfortunately, the results of the relatively few studies conducted on the issue proved to contradict each other, which calls even more so for further investigation.

Research questions and hypotheses

In view of the unexplored areas of FLCA, the following research questions were formulated to guide the study:

- Does an individual's level of Foreign Language Classroom Anxiety vary significantly in view of the two foreign languages he is learning at the same time irrespective of his trait anxiety?
- If it does, what are the elements that learners attribute to the development of foreign language anxiety in one foreign language classroom but not in the other?

Based on previous research the following hypotheses were drawn up as possible answers to the research questions presented.

- As the different foreign language learning situations vary, in some learners they may produce different levels of a specific type of situational anxiety, i.e. Foreign Language Classroom Anxiety.
- The attributed causes of FLCA are likely to lie in the learning situation i.e. the language classroom.

In order to prove these hypotheses correct a two-phased study consisting of a quantitative survey and a qualitative case study was carried out.

Method

Design

In order to seek answers to the research questions guiding this study, a two-phased within-participants research design was developed that included both quantitative and qualitative elements. The quantitative part of the study, with the help of surveys, sought to focus on the differences in the levels of FLCA a student may experience in two different language classes. Whereas the qualitative part, consisting of a case study, provided more insight into the possible perceived sources of the differences in the levels of foreign language anxiety an individual may experience.

Participants

The participants in the study were *N*=61 ninth-year students studying at a grammar school in Budapest, Hungary. According to the school's curriculum, all ninth-year students study two foreign languages. The pair of languages students took were the following: English-French, English-German, English-Italian, German-French, German-Italian, German-English.

Furthermore, the school's curriculum distinguishes between first and second foreign languages by calling the language students had learnt before entering secondary school (i.e. in elementary school) as the first foreign language (FL1). Normally, the second foreign language (FL2) students choose and begin to learn at the secondary school without having had any previous training in that given language. Hence English and German can appear both in the position of FL1 as well as FL2. Another note worth mentioning here is the fact that school policy for this particular group of ninth-year students requires that they learn the FL1 and FL2 in the same number of hours per week (i.e. three). This regulation allowed the variable of the number of lessons per week to be controlled for.

Using the results of the instruments involved in the first phase, participants who demonstrated different levels of anxiety in their two different language lessons were selected for participation in the qualitative phase. Based on the results of the quantitative phase of the study, the purposive sample of critical cases were selected (students with low levels of trait anxiety, low levels of FLCA for FL1 but high levels of FLCA for FL2.). This allowed for the exploration of the perceived reason behind learners' increased feelings of discomfort in one language lesson but increased feelings of comfort in the other.

Instruments

Three questionnaires were used for data collection in the first phase. The first two surveys were the Hungarian versions of the Foreign Language Classroom Anxiety Scale (Tóth, 2003) administered for the two different foreign languages students were learning. With their help, the aim was to determine the level of anxiety stu-

dents experience in the two different language classrooms. The Foreign Language Classroom Anxiety Scale (FLCAS) was developed by Horwitz, Horwitz and Cope (1991) to assess the degree to which a respondent feels anxious in a foreign language classroom based on the construct of foreign language classroom anxiety being a composite of communication apprehension, fear of negative evaluation and test anxiety. The scale is a 33-item 5-point Likert scale which includes 24 positively worded and 9 negatively worded items (for the interpretation of results see data analysis below).

Participants were administered the Hungarian version of the scale which was translated and validated by Tóth (2003) for use in the college EFL classroom (α=93). As such, beside the introduction of the survey, one minor adjustment in wording was made for the convenience of secondary school students. (In item 8, The term 'zh' for tests in college was substituted by 'dolgozat' to refer to tests in the secondary school context. Reliability was additionally ensured by way of a think-aloud with a potential respondent.

Besides the two FLCAS, a third questionnaire was administered to purposively select respondents to participate in the qualitative phase of the study. For this purpose, the second part of the State Trait Anxiety Inventory (STAI) was used which aims at determining an individual's level of trait anxiety. The STAI was translated into Hungarian, validated and published by Sipos, Sipos and Spielberger (1994). The second part of the STAI, which refers to trait anxiety, is a 20-item 4-point Likert scale, consisting of sentences referring to participants' general emotional state, of which 13 are positively and 7 are negatively worded items. With the help of this survey, those students who demonstrated high levels of foreign language anxiety due to their high levels of trait anxiety could be filtered out, thus suggesting that foreign language classroom anxiety can develop in individuals irrespective of a low level of trait anxiety.

In the qualitative phase of the study, the instrument used was an interview schedule adapted from Price's study (1991 in Horwitz & Young, 1991), focusing on the possible sources of foreign language classroom anxiety. The logic behind employing an instrument of a previous qualitative study was mainly to ensure the comparability of results. Thus, the five open-ended questions concentrating on how students feel in the foreign language classroom and the role of the instructor were merely supplemented with three more items: Item (1) inquired about the student's experiences of the target culture, item (2) investigated the possible personal, day-to-day contact with the target language in an informal setting, and item (8) was included in order to differentiate between the debilitating and the facilitating effect of anxiety on language learning.

Procedure

Data collection

Data collection was organized in two phases. In the first quantitative data were gathered using the FLCAS and the STAI. Students filled in the FLCAS referring to their English lessons during their English class, with only the researcher being present. Two weeks later, in their homeroom class, all ninth-year students filled in the FLCAS for the other foreign language they were learning along with the second part of the STAI. These surveys too were administered without the presence of the language teacher.

In order to preserve anonymity to a certain extent, participants had the option of signing their questionnaires using their names or symbols they chose. However, in the case of using symbols, respondents were also asked to give their name and the symbol they used on a separate sheet of paper, which was kept in a sealed envelope until all quantitative data had been analyzed, thus those students eligible for the interviews could be easily traced. All three questionnaires were paired according to the symbols or names used, they were then given code numbers for easier manageability and their answers were subjected to statistical analysis (described in the data analysis section below).

Students demonstrating a high level of anxiety in one foreign language classroom but low in the other, and scoring low on the trait anxiety scale of the STAI were asked to participate in a case study. The objective of this second phase was to gather qualitative data concerning the possible sources of FLCA. Interviews were conducted, recorded, and transcribed. The transcriptions were then coded with the corresponding numbers adapted from phase one, making matching of the surveys' answers to the interview data more manageable. In order to emphasize that individual personal experiences formed the basis of the case studies, the learners were also allotted pseudonyms to refer to their interview data. The data gathered in the interviews were finally analyzed and compared to those of the surveys.

Data analysis

The first scale administered in the quantitative phase of the study was the Foreign Language Classroom Anxiety Scale. Although there are no precise guidelines as to the analysis of the responses given to the FLCAS, literature suggests adding the points of the 33 item five-point Likert scale by giving a value of 1 to the response 'strongly disagree', and a value of 5 to 'strongly agree'. In the case of negatively worded items (such as item 2, 5, 8, 11, 14, 18, 22, 28, and 32), the values were reversed.

In the present study, the scores of the FLCAS were interpreted as follows: Scores lower than 99 denote lower levels of anxiety, while scores higher than 99, denote higher levels of anxiety. The score of 99 was the cut-off line. Participants' overall scores of the FLCAS for both foreign languages were calculated and compared by calculating correlations to determine similarities and employing the

paired samples t-test to test for the significance of the differences in the levels of FLCA experienced in one foreign language classroom and in the other.

For the four-point 20-item Likert scale of trait anxiety in the STAI, following previous studies, participants' scores were calculated by assigning a value of 1 to answers of 'never' and '4' to answers 'always' respectively. The points of the Likert scale were added, with the scores of the negatively worded items (items 1, 6, 7, 10, 13, 16, and item 19) reversed. Based on Sipos, Sipos and Speilberger (1994), respondents who reached a score higher than 40 were considered to have a higher level of trait anxiety.

With the help of the SPSS (2004) statistical program two filters were set up to select the cases where the following three conditions were satisfied:

1. one of the FLCAS scores demonstrates a high level of foreign language classroom anxiety
2. but the other does not, what is more,
3. the participant does not possess a high level of trait anxiety

The data analysis of the interviews followed the steps of qualitative content analysis, seeking common patterns in the responses. Although in qualitative research a priori hypotheses are not set up, based on previous studies concerning students' ideas of the sources of their feelings of discomfort in language class, some things could be predicted (Piniel, 2000; Price, 1991). Thus, categories involving group (group size, peers), classroom management, time studied, ambiguity (of expectations, evaluation of the teacher, clarity of input) were expected to emerge from the interview data.

Results

Phase one

The validity and reliability of the results of the survey data were ensured by the validity and reliability of the Hungarian version of the two questionnaires (Tóth, 2003 and Sipos, Sipos & Spielberger, 1994 respectively). The reliability of the responses was further enhanced by the absence of the language teacher and the option for the participants to use symbols to sign the questionnaires.

Upon examining the two FLCAS scores for the two foreign languages (FL1 and FL2) do not seem to co-vary, Pearson's product-moment correlations were not found to be significant (r= .108, N=61). However, differences in the two scores for the two foreign languages of the sample (N=61) proved to be significant, with t=4,055 at the level of significance p < .01 .

These results suggest that it is possible that a learner studying two foreign languages simultaneously develops different levels of FLCA. Furthermore, calculating with the possible affects of trait anxiety on FLCA, cases demonstrating high levels of trait anxiety were filtered out. Hence those learners were purposefully selected for the qualitative phase of the study who did not prove to possess the

personality trait of anxiety (based on their STAI scores) and alongside this, showed a significant difference in terms of FLCA in the two different foreign language classrooms. (For filters see Table 1, for filtered cases see Table 2).

Out of the 61 participants, 7 students met the criteria outlined above. Unfortunately, of the six, only five were available for the remaining part of the research. (For the cases selected via the filtering procedure described above see Table 2.)

Table 1. Filters used to select cases for the second phase of the study.

Filter 1: SUM1>99&SUM2<99&SUM3<40
Filter 2: SUM1<99&SUM2>99&SUM3<40
SUM 1: sum of scores for the FLCAS for the first foreign language

Table 2: Cases that were filtered for the second phase of the study and their SUM1, SUM2, and SUM3 values.

CODE NUMBERS	CODE	SUM1	SUM2	SUM3
2	Anna	69	125	36
14	Brigi	45	125	28
20	Csaba	65	100	39
39	Dóra	67	116	35
42		91	127	39
51	Erika	69	121	39

SUM 1: sum of scores for the FLCAS for the first foreign language (FL1)
SUM 2: sum of scores for the FLCAS for the second foreign language (FL2)
SUM 3: sum of scores for the STAI

3.2 Phase two

The second half of the study involved five of the selected six cases (due to problems of availability) mentioned above. All of these students learned English as FL1 and had just started learning French as FL2. Triangulation of methods of data collection added to the dependability of the data gathered in the qualitative phase. The responses to the FLCAS of the higher score, i.e. representing their levels of anxiety during the French lessons, were analyzed based on the content of the items. The answers to interview questions were subjected to content analysis. Finally, the data of the questionnaires and the interviews were compared and grouped into categories. For the convenience of retrieval an audit trail was kept of the research process.

Anna

Anna studies English and French at grammar school. She has been learning English for eight years, has spent twice ten days abroad studying English and she has a certificate of the intermediate level state exam of the English language. She feels very positive about her knowledge of English as opposed to French which she started to learn three months ago. Her feelings about the two foreign language classes are rather different, as her scores of the FLCAS show (see Table 2)

Anna generally experiences discomfort during French lessons as her answers of the relevant items of the FLCAS depict. She is more tense in French classes than in other classes at school (item 26), she feels quite nervous (item 28) and many times would rather avoid going to class altogether (item 17). In the course of the interview Anna did not declare her feelings quite outright, but alluded to things she did not feel comfortable with or would rather change if she could (e.g. *"the teacher should motivate us somehow"* (line 32), *"we shouldn't learn texts by heart, but talk and discuss interesting topics"* (line 36)).

Both in the FLCAS and in the interview, Anna showed evidence of low self-esteem. In the questionnaire referring to the French lessons, Anna gave a score of four to items 13, 18, and 16, and a score of five for item 1 and declared in the interview that she was not good at all at French (line 20). She considers herself to be one of the weaker students: *"because I'm one of the worst students there"* (line 16), which probably makes her self-conscious when speaking in front of others (as her answers to the FLCAS items 7, 23 and 24 depict).

Besides her uneasiness in front of her peers, Anna especially fears negative evaluation form the part of the teacher. Her answers to the questionnaire suggest that she is not ready to offer answers in class (item 13), tends to become nervous if the teacher calls on her (item 20) or corrects her for an unknown reason (item 15). Anna seems to be simply afraid of the teacher correcting her (item 19), she feels great pressure when preparing for a lesson (item 22), and fears the situation where she has no time to prepare before answering the teacher's questions (item 33).

It follows that Anna is nervous when speaking in class (item 27) and also feels distressed when she cannot understand the teacher (item 29). This more often refers to the explanations the teacher gives as the following quotes clearly show: *"the teacher does not explain things"* (line 22) and *"…I don't understand a word of what the teacher says"* (lines 25). This maybe the reason why the more Anna claims to study for tests, the more confused she becomes (item 21 and item 8).

Anna added in the interview that although she is not a diligent student, she would probably learn more if the material taught would be more colorful and interesting (including topics of culture, famous sights, history, stories, curiosities etc. (line 36-39) Although she refrains from accusing the teacher for the way she feels at a French lesson, Anna does say that a lot depends on the teacher-student relationship and the teacher's effort to motivate students (line 66).

Brigi

Brigi, as Anna, started learning French three months ago in contrast with English which she has been learning for six years. Brigi has a sister-in-law who is French, and thus she had the chance to familiarize herself with the French language and culture via direct experience of spending time in France and also through keeping contact with her relation (although English is used as the means of communication). Despite these circumstances, Brigi's score on the FLCAS of French (SUM 2) demonstrated a significantly higher level of anxiety than what she experiences during English (SUM 1) lessons (see Table 3.).

Brigi scored a maximum of 5 on 19 of the 33 items of the FLCAS. The most prominent were the six items (5, 11, 12, 17, 26, 28) referring to the general state of nervousness she experiences during a French lesson, each item receiving a score of 5. In the interview she used the expression *"there is great stress"* several times to hint at the atmosphere of the lessons.

For Brigi it seems that the low level of her self-confidence and peer opinion do not have such a subtle effect on her level of anxiety as in Anna's case. As Brigi claims in the interview:

> there are certain people who it is very difficult to take it from if they laugh at you. But, as for me, I don't care, because I don't think they have a good enough reason to laugh at me, so I don't care whether they do. (line103-104).

Brigi is more affected by the teacher's evaluation, as in the survey she admits to worrying about making mistakes (item 2), fearing to be called on by the teacher (items 3 and 20), having to answer without preparation (items 9 and 33) and being corrected by the teacher (items 19 and 15). Considering that all spoken contributions to the lesson are evaluated by the teacher, it does not come as a surprise that Brigi fears such situations. As she puts it:

> if someone doesn't know something it's a minus or a one (the worst grade you can receive), and everything, and well, this is awful, everyone is really stressed (line 51-52).

In view of this, it is reasonable that Brigi admits to being nervous when speaking in class (item 27).

What Brigi claims to cause anxiety for her in the French lesson is, as she explains, the fact that teacher's expectations are not clear, *"And this is all because she doesn't tell us what she expects exactly"* (line 53-54), there is little time allotted to practice, and there is immediate evaluation of the student's performance:

"she gives short deadlines for memorization tasks" (line 54)

"you have to learn it at home and on the lesson you have to answer right away when she asks you"(line 89)

"it's awful if I have to say something no one has said before me and it is right at that moment that it turns out whether what I'm saying is something stupid." (line 92-93).

What further contributes to Brigi's increased level of anxiety is the fact that students at the French lessons are asked one by one to produce correct chunks of language while there is complete silence in the classroom ("S*he asks everyone one by one. That person gets complete silence, and he/she has to answer. And I think this is very stressful"* (line 85-86).

Another source of uncertainty is the incomprehensibility of the language output of the teacher as well as that of peers, and the limited pool of linguistic tools at Brigi's disposal to express her thoughts in French: *"Well, you see, English lessons are a lot better because I understand what the teacher is saying, and what the others are saying too, and also I can express myself somehow in English"* (line 28-29).

All in all, according to Brigi, the teacher plays a key role in determining how she feels at a French lesson. Brigi expects the instructor to add something to the language lessons besides just reading the textbook:

> Well, everything depends on it [the instructor's role]. Because, without it we might as well stay at home and learn French from a book. Theoretically, this is why I come to school, so that the teacher can add something to this, she can explain and tell us etc. So it all depends on how someone feels, I think. (line 97-98)

Brigi also suggests a few practical ideas how a lesson could be made more interesting on the part of the instructor. She lists group work, using pictures and watching films as things that contribute to a good lesson: *"a lesson is good if there are, for example, group tasks, or watching a film, pictures etc, these make a good lesson"* (line 78).

It is interesting, however, that for Brigi the age of the teacher is a criteria that determines student-teacher relationships and thus ultimately underlies the success of the above mentioned activities *"From a teacher who is a lot older than us, these things are not so relaxed, so o.k. she would do it let's say, but it wouldn't be honest, she can't sympathize with us, I think"* (line 79-80).

Csaba

Csaba is another student who experiences feelings of nervousness and anxiety in the French classroom (see Table 2.). As the two students mentioned before him, he too studies English and French at grammar school. He has been learning English for eight years, whereas he has just took up French three months ago. Csaba did spend some time in France but did not seem to return with a clearly positive ex-

perience. All he could say about his impressions was that French people cannot speak English (line 11).

As for his French lessons in school, Csaba only gave a maximum of 5 points on the Likert scale of the FLCAS to items regarding his general discomfort (items 26 and 28). He further reflected on this in the interviews saying that *"French is a nerve-racking lesson"* (line 17).

Csaba, unlike Anna, is not so much afraid of his peers' criticism (only scoring 4 on one item related to the fear of being left behind by the others), but does share the worries of negative evaluation of the teacher (items 2, 20, 22, and 33). In the interview Csaba elaborated on his worries especially in view of oral tests. It seems that no matter how much he or his classmates study, oral tests are the chief sources of stress:

> well, there is a kind of pressure on people, in French, so everyone looks nervous, or in spite of having studied they are stressed. And when a question is asked and they are called on to answer it, it's like hearing your death sentence. So it's pretty awful. (line 29-31)

Which is not surprising if one considers Csaba' account of his personal experience:

> …she called on me, but [a fellow student] is just before me on the roll and he was before me and I knew that I would be next, and out of the three texts I had only learnt two. So I told her [the teacher] that I knew only two of the texts. And she answered that that's a huge problem, and I thought that I was going to get a one [a fail grade] but finally I recited text three. At home I could recite it from memory very fluently, but then in class I stammered, after every other word…it was horrible. (line 58-63)

Although not to the extent of causing physical symptoms to appear, written tests in French are also potential sources of anxiety for Csaba: *"Well, the tests, last time we wrote a three and a half page test and I got a two (a grade better than a 1, on a 1-5 scale where 1 is a fail grade) Well, I guess it's not that bad, because half the group failed, so. "* (line 25-26).

When inquired about the perceived causes of anxiety he experiences in French class, Csaba was baffled: "I don't know what I'm afraid of, because my mum tells me not to be nervous, nothing will happen if I accidentally get a fail grade in French. Well, I don't know." (line 65).

Later, Csaba talked about the role the teacher plays in influencing his feelings, as he elaborated on the importance of the teacher's personality (line 47), the way the teacher-student relationship develops (line 50), the teaching methodology she employs (line 47-48), with the consequences of confusion of the material taught (line 34-39). The weight of these is clearly demonstrated in Csaba' description of the physical symptoms of anxiety: *"I can't really talk to her [the teacher] because, it's awful, my hands start to shake, and everything. So it's quite horrible"* (line 55-56).

Dóra

Dóra is yet another student who is learning English and French simultaneously. From her biographical data it is clear that she has been to both English speaking and French speaking countries from where she returned with positive experiences. Despite this, her scores in the FLCAS for English and French classes differed greatly (see Table 3.) Dóra, similarly to Brigi, gave a score of five to sixteen of the items on the scale that contributed to her negative feelings of the foreign language classroom and gave a score of four to one additional item.

In view of the FLCAS scores, Dóra too emphasized her general feelings of discomfort during French classes (assigned a core of five to five items), along with claiming that she fears negative evaluation of the instructor (assigned a score of five to seven items dealing with the issue).

In the interview Dóra described the atmosphere of the French classroom as rather tense: *"It's not that I don't feel well, but the atmosphere is so tense and nervous."* (line59) supporting her responses on the FLCAS. Fear of negative evaluation arose in the course of the interview as fear of being called on by the teacher and not knowing what the correct answer to her question may be: *"During French lessons I don't feel well (comfortable) because I'm nervous whether we will have oral tests or not. "* (line 14-15).

It is interesting to note that Dóra does not feel responsible for her nervousness, but rather she attributes her negative feelings to the inappropriate teaching methods that are characteristic of her French lessons. Inconsistency, unclear expectations, demands impossible for students of their level to meet all suggest the inappropriacy of the methods the instructor employs. The above seem to contribute to Dóra's higher level of foreign language anxiety:

Inconsistency:

"It's like as if it were our fault that we have never heard of this before in our lives" (line 63)

Unclear expectations:

"…the teacher doesn't teach us what she demands of us. She just tells us that we have to know these things, but not even on which page of the textbook can we find it" (line 19)

Demands impossible for students of their level to meet:

"she asks too much, and we have to study a lot" (line 17).

Finally, Dóra states that as a result of the way she feels in her French lessons she is willing to take extra lessons in the summer and learn French in order to be better at it. In Dóra's case this seems to demonstrate more a facilitating type of anxiety than a debilitating one.

Erika

Erika has been studying English for nine years, and started to learn French this academic year. Like the four participants before, her scores of the FLCAS for English (SUM1) and French (SUM2) differed, showing a higher level of anxiety in the case of French (see Table 2.). Besides giving a score of four or five to four items (items 5, 17, 26, and 28) on the FLCAS that relate to the feeling of general discomfort, she too marked five items (items 3, 10, 15, 19, and 33) related to the fear of negative evaluation as representing her emotional state during French lessons.

In the interviews she describes the source of her fears as stemming from the teaching methods that dominate the lesson. As Dóra, Erika also mentions the problem of inconsistency: *"because we learn the grammar like, the teacher explains something and then says it's in the book. But then it's not what is in the book that she wants to hear",* (line 64-65). Expectations are often unclear: "*only if we knew what exactly is required of us, what are we supposed to learn, then it wouldn't be like having to look for things in the dictionary, in our books, or notebooks and we get totally confused*", (line 152-153). Moreover, input is frequently incomprehensible and confusing for her: "*I don't understand French that well, I like it less because, because somehow French lessons are a little incomprehensible*" (line 24).

In the interview, Dóra explained in detail the points of teaching methods that she suggests for reconsideration. In terms of the lessons' content Dóra would prefer dealing with more topics in general and not merely conducting oral tests:

> I know that this is a language lesson, and we don't have time for it, but if something is connected to the book or the topic anyway, we could talk about it instead of just having oral tests (evaluation), what's the next lesson about, what do we have to learn for the next day and that's it. (line 82-83)

As for the topics themselves, they should be relevant and useful for the language learners later on:

> And I would like it better, if there was something that is about the fact that we really will use French for communication, and not just memorizing texts word for word. It's ok that we know these things and get fives (get excellent grades) and it's important, but it would be more important if we could use the language. (line 121-123)

In terms of task type, Dóra would like to have speaking exercises (line 75), and clear explanation of the material to be learned: "*I think things should be taught in more detail and with more care (attention), so that everyone would be able to understand*" (line 108-109). Dóra feels that besides teaching methods, the teacher's personality and the time spent studying a language also have an influence (line 102-104) on how she feel at a lesson (line 22).

Based on the evaluation of the responses to the FLCAS and interview question, several topic and issues raised could be grouped in categories and slowly a pattern of perceived factors influencing the development of FLCA emerged. The fol-

lowing section will focus on how the findings can be linked with the research questions and how they relate to results of previous studies conducted in the field.

Discussion

The first, quantitative, phase of the study sought to answer the research question of whether an individual's level of Foreign Language Classroom Anxiety vary significantly in view of the two foreign languages he is learning at the same time irrespective of his proneness to anxiety. Based on the results and the statistical analysis presented, the study of a within-participants design found that in a sample of 61 secondary school students studying two foreign languages simultaneously demonstrated significantly different levels of foreign language classroom anxiety. Furthermore, in the case of six students, these result proved to be independent of students' level of the permanent characteristic of trait anxiety.

These findings supported and extended those presented in an earlier study (Piniel, 2000) conducted with the participation of a smaller group of students indicating that foreign language anxiety can develop independently of trait anxiety. However, Rodriguez and Abreu (2003) reported no significant difference between the levels of foreign language anxiety in college students majoring in two foreign languages simultaneously, namely English and French with prospects of becoming foreign language teachers. The fact that the context of the studies mentioned above differ does not justify the generalization of their results but warrants further investigation.

Controlling for trait anxiety as a variable affecting foreign language classroom anxiety, the second phase involving five students' case studies investigated the perceived causes of the feeling of anxiety in the foreign language classroom. All five students were selected based on their differing scores of FLCAS for English (FL1) and French (FL2), and their low scores on the trait anxiety scale of the STAI. The five participants scored higher on the FLCAS of the FL2 (French), which all of them have been learning for three months, whereas the FL1 they have been studying for at least three and maximum for nine years (English). The difference of time spent studying a language was attributed several times in the courses of the interviews to the feelings of discomfort. For this MacIntyre, Baker, Clément, and Donovan, (2002) have also found support, stating that anxiety tends to be dependent on the age and perceived competence of the learner.

By analyzing the items of the FLCAS receiving a high score of four or five, two main categories of the causes of anxiety emerged: that of general discomfort felt in the foreign language classroom, and students' fear of negative evaluation form the part of the teacher rather than their peers. The unsure posture, the feeling of discomfort arising at the idea of having to sit through more lessons, the understanding of lessons as something to be afraid of and nervous about were among the ideas the selected respondents all tended to opt for when describing French lessons (e.g., items 1, 4, 5, 29). Points directly related to concrete events in the foreign language classroom included the following: the teacher does not clarify why the student is being corrected, the feeling of fear of being corrected by the teacher

every time a mistake is made, the teacher posing questions students have not pre-pared for, and the fast pace of the lessons (e.g., items 9, 15, 19, 25).

The results of the interviews seem to support the above with the role of the teacher surfacing as the overall influential factor behind the general feelings of anxiety and fear of negative evaluation. Within the role of the teacher several aspects were found to be perceived sources of foreign language classroom anxiety (see Figure 1).

The outcome of Price's (1991) and Piniel's (2000) research indicating the instructor's significant role in developing language learner's anxiety very much resemble those presented here. However, in both studies the teacher's role arose as merely one classroom factor from among the others such as fear of peers' negative evaluation, group size, number of previous teachers, classroom management etc., whereas in the present study no or very limited allusion were made to other situational factors than the teacher and the material to be learnt.

It is also interesting to note that FLCA can appear in spite of positive experiences with the target culture or the target language prior to formal instruction as in Brigi's case. It may be fruitful to further investigate to what extent and in which context do situational factors of the formal learning situation override perception and positive experiences with the target language group and language. This is the more interesting if we consider the model of Willingness to Communicate put forth by MacIntyre, Dörnyei, Clément and Noels (1998) where the learner's attitude towards the target language group is claimed to have a determining role in the learner's WTC. Yet it seems, that in the foreign language context (as opposed to the second language context) situational elements may override the perceptions one has of the target language and the target culture.
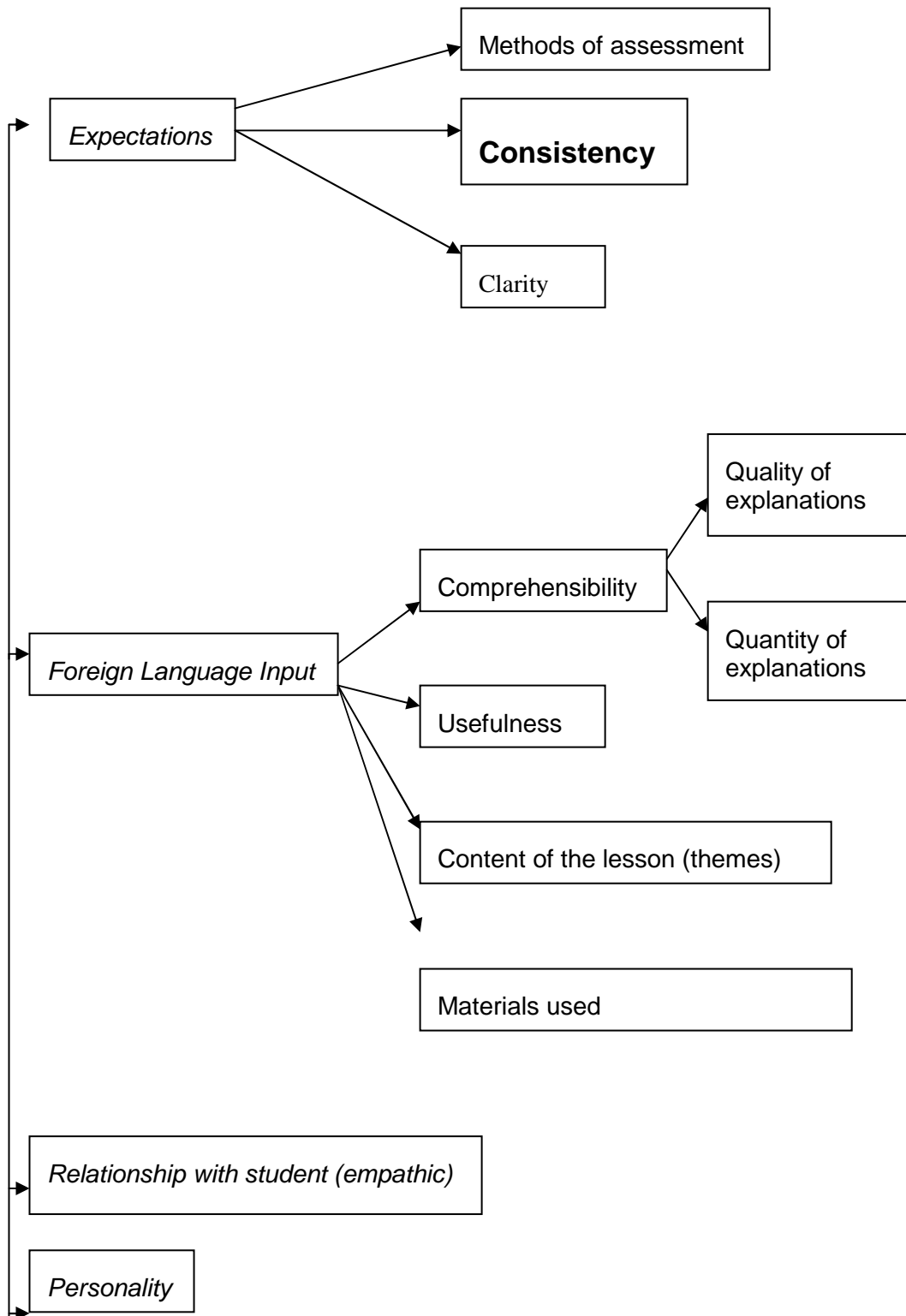
Figure 1: The role of the instructor as a potential source of foreign language class-room anxiety

Conclusion

Although this research does not put forth generalizable results, it does draw the reader's attention to the phenomena of Foreign Language Classroom Anxiety and its possible causes as perceived by learners themselves. The possibility of foreign language anxiety developing in spite of a low level of trait anxiety emphasizes the role of the context of foreign language learning, which in an EFL situation is the classroom itself. Thus such a study has pedagogical implications as to the importance of the factors playing a role in a foreign language classroom. In the case studies presented here, the teacher's role was especially highlighted as a potentially key factor in inducing students' foreign language anxiety. However, these results must be taken with caution. It should be kept in mind that these are relevant in the particular setting the research was conducted in, namely the environment of a school in Hungary where students are obliged to take two foreign language courses simultaneously, where oral tests in all subjects are common, and where the 'excellent' grade is the goal to aim at. As seen from literature, other sources of anxiety in the classroom do exist, thus this study is best viewed as one that simply raises awareness of the importance of the instructor's role in view of learners' foreign language anxiety.

Besides producing merely thought provoking rather than generalizable results, the present research was not without limitations. Problems concerning research methodology did occur, which may affect the reliability of the quantitative result. The timing of administering the second part of the questionnaire in the last lesson of the day may have caused students to merely rush through the survey questions as soon as they could, without giving themselves time to consider the questions.

Another problem was presented by the fact that a complete class of 35 people learning Latin as one of the foreign languages was excluded, decreasing the number of participants and indirectly affecting the generalizability of the results. Furthermore, reliability of students' responses may have been negatively influenced by the fact that the researcher herself is a teacher at the school where the study was conducted.

As only five students were selected to participate in the case study, statistical analysis of gender and its relationship to the development of foreign language anxiety could not be determined. Similarly, correlations of the amount of time spent studying a language, or the low level of proficiency with the level of foreign language anxiety could not be calculated.

As this field of research has not yet been fully explored, further investigations are suggested in determining the reasons of discrepancy between the two levels of foreign language anxiety if an individual is learning two foreign languages simultaneously. As a follow-up to the interviews, the case studies should be extended with observing students' behavior in the foreign language classroom. Moreover, as part of a longitudinal study, returning to these students in two years' time would open new realms in the investigation into the nature of foreign language anxiety.

References

Aida, Y. (1994). Examination of Horwitz, Horwitz and Cope's construct of foreign language anxiety: The case of students of Japanese. *The Modern Language Journal, 78*, 155-168.

Bailey, P., Onwuegbuzie, A. J., & Daley, C. (2000). Correlates of anxiety at three stages of the foreign language learning process. *Journal of Language and Social Psychology, 19*, 474-490.

Baker, S., & MacIntyre, P. D. (2000). The role of gender and immersion in communication and second language orientations. *Language Learning, 50*, 311-341.

Dewaele, J.-M. (2005). Sociodemographic, psychological and policultural correlates in Flemish students' attitudes towards French and English. *Journal of Multilingual and Multicultural Development, 26*, 118-134.

Eysenck, M. W. (1979). Anxiety, learning, and memory: A reconceptualization. *Journal of Research in Personality, 13*, 363-385.

Gardner, R. C., & MacIntyre, P. D. (1992). A student's contributions to second language learning. Part I: Cognitive variables. *Language Teaching, 25*, 211-220.

Gardner, R. C., & MacIntyre, P. D. (1993). A student's contributions to second language learning. Part II: Affective variables. *Language Teaching, 26*, 1-11.

Gregersen, T., & Horwitz, E. K. (2002). Language learning and perfectionism: Anxious and non-anxious language learners' reactions to their own oral performance. *The Modern Language Journal, 86*, 562-570.

Horwitz, E.K., Horwitz, M.B., & Cope, J. .A. (1991). Foreign language classroom anxiety. In E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 27-36). Englewood Cliffs, New Jersey: Prentice Hall.

Horwitz, E. K., & Young, D. J. (1991). *Language anxiety: From theory and research to classroom implications.* Englewood Cliffs, NJ: Prentice-Hall.

MacIntyre, P. D., Baker, S. C., Clément, R., & Donovan, L. A., (2002). Sex and age effects on willingness to communicate, anxiety, perceived competence, and L2 motivation among junior high school French immersion students. *Language Learning, 52,* 537-564.

MacIntyre P. D., Clément R., Dörnyei Z., & Noels K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *Modern Language Journal, 82,* 545-562.

MacIntyre, P. D., & Gardner R. C. (1991a). Anxiety and second language learning: Towards a theoretical clarification. In: E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 41-54). Englewood Cliffs, New Jersey: Prentice Hall.

MacIntyre, P. D., & Gardner, R.C. (1991b). Methods and results in the study of anxiety and language learning: A review of literature. *Language Learning, 41*, 85-117.

MacIntyre, P. D., & Gardner, R. C., (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44,* 283-305.

Onwuegbuzie, A. J., Bailey, P., & Daley, C. E. (1999). Factors associated with foreign language anxiety. *Applied Psycholinguistics, 20*, 217-239.

Piniel, K. (2000). *Foreign language anxiety: The role of classroom factors in the development of foreign language classroom anxiety.* ELTE: Unpublished MA thesis.

Price, M. L. (1991). The subjective experience of foreign language anxiety: Interviews with highly anxious students. In: E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp.101-108). Englewood Cliffs, New Jersey: Prentice Hall.

Rodriguez, M., & Abreu, O. (2003). The stability of general foreign language classroom anxiety across English and French. *The Modern Language Journal, 87*, 365-374.

Scovel, T. (1978). The effect of affect on foreign language learning: A review of the anxiety research. In: E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp.15-23). Englewood Cliffs, New Jersey: Prentice Hall.

Sipos, K., Sipos, M., & Spielberger , C. D. (1994). A state-trait anxiety inventory (STAI) magyar változata. In F. Mérei & F. Szakács (Eds.), *Pszichodiagnosztikai Vademecum 2. rész.* Budapest: Nemzeti Tankönyvkiadó.

Tóth, Zs., (2003). Validation of the Hungarian foreign language classroom anxiety scale. ELTE: Unpublished PhD seminar paper.

Foreign Language Anxiety in Polish Dyslexic Secondary School Students

Ewa Piechurska-Kuciel

Opole University, Opole, Poland

epiech@uni.opole.pl

Introduction

The aim of the present paper is to investigate language anxiety levels perceived over three years of study at a Polish secondary school within the context of the developmental dyslexia influence. (The research was financed by the Polish State Committee for Scientific Research, 2004-07.)First, a brief outline of developmental dyslexia with its cognitive and affective consequences will be given, followed by a presentation of the phenomenon of language anxiety. Then empirical findings on the relationship between language anxiety and developmental dyslexia will be examined and discussed.

According to the latest definition of developmental dyslexia, it is

> a specific learning disability that is neurobiological in origin. It is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language that is often unexpected in relation to other cognitive abilities and the provision of effective classroom instruction. Secondary consequences may include problems in reading comprehension and reduced reading experience that can impede growth of vocabulary and background knowledge. (Lyon, Shaywitz & Shaywitz, 2003, p. 2)

The definition acknowledges the importance of latest findings of functional imaging or genetic studies for the neural basis of dyslexia. Thanks to the most recent advances in neurobiology, left hemisphere posterior brain systems are found not to work properly in reading or other visual task processing (Bednarek, 2002). Dyslexic students are characterized by the inability to read fluently, even despite mature experience with reading (Lefly & Pennington, 1991). The central difficulty in dyslexia is connected with the lack of phonemic awareness defined as 'the ability to isolate and manipulate consciously the sounds of the language and relate them to the appropriate written letters or letter combinations' (Nijakowska, 2000, p. 248). Dyslexics are unable to decompose words into phonological segments and to decipher the reading code based on the process of blending or analyzing sound segments and rearranging phonetic elements (Downey, Snyder & Hill 2000). Con-

sequently, dyslexics' chronological age or academic development does not correspond to their reading age (Lyon et al., 2003). Phonological difficulties lead to reading and background knowledge problems. Nevertheless, effective classroom instruction provided as early as possible leads to avoiding many reading problems created not by developmental dyslexia but by the lack of foundational skills in some disadvantaged children (Torgesen, 2000).

Dyslexia significantly affects the individual's learning abilities as such students are found to be easily distracted, uncoordinated; they may also have poor time-management skills (Barga, 1996). Developmental dyslexia is correlated with numerous negative effects, like difficulty with motor skills, poor working memory, low self-esteem or slow speed of information processing (Crombie, 2000). Apart from difficulties at the orthographic/phonological level, dyslexic students have problems understanding language rules or reflecting on language (Ganschow & Sparks, 2000). With their defective phonemic awareness that may be sufficient for L1 but deficient for L2 (Ganschow & Sparks, 1995), they experience the process of foreign language learning as even more stressful than their unimpaired peers. Consequently, it may be presumed that any negative emotions connected with L2 learning that may be experienced by non-dyslexic students may constitute a remarkably significant threat to their dyslexic peers.

Many negative effects of developmental dyslexia are connected with a high level of frustration for the pupil, which in turn leads to emotional and behavioural problems (Alcott, 1997). These problems include sensory deficit (motion sensitivity or flicker), motor (balance) and cognitive (phonological, working memory or speed) (Nickolson, 2001). It is clearly visible in the process of learning a foreign language inducing feelings of apprehension and doubt when learners need to express their mature thoughts in a language they have not fully mastered (Young & Kimball, 1995). Their self-perception of genuineness is threatened by a limited linguistic repertoire (Horwitz, 1999).

There is not much evidence concerning emotional issues related to developmental dyslexia (MacNulty, 2003), nevertheless, it is claimed that this learning disability has a potential negative impact on academic achievement (Reynolds et al., 1996) and social life (Kappers & Veerman, 1995). Such students are likely to display symptoms of anxiety and depression, together with low self-esteem, self-doubt and emotional insecurity (MacNulty, 2003).

Anxiety is an emotional symptom very frequently reported by dyslexics (Ryan, 1992). In the foreign language classroom when 'a situation requires the use of a second language with which the individual is not fully proficient' (Gardner & MacIntyre, 1993, p. 5), a different type of apprehension, *foreign language anxiety*, often comes into being. It constitutes an influential variable in a FL acquisition process for all students, so it may be presumed that language anxiety is very strongly pronounced with dyslexics.

Language anxiety is generally viewed as an emotion connected with a general use of a foreign language (Gardner et al., 1993). There is also another definition of language anxiety stressing the impact of classroom conditions for the process of L2 acquisition. Here language anxiety is seen as 'a distinct complex of self-perceptions, beliefs, feelings, and behaviours related to classroom language learning

arising from the uniqueness of the language learning process' (Horwitz et al., 1986, p. 128). For the purpose of this paper the definition is adopted as the working one due to its focus of the specificity of the learning situation and context.

MacIntyre and Gardner (1991) offer a model of language anxiety development, according to which students who experience a string of negative emotions connected with the L2 acquisition process have a tendency to attribute their fearful experiences with the foreign language they study. Even though at the beginning of a FL course a student may not necessarily suffer from high levels of language anxiety due to their motivation and language aptitude (MacIntyre et al., 1989), later, with more negative experiences accumulating, high levels of language anxiety are likely to appear (Gardner et al., 1993). But in the course of language study, with FL proficiency expanding, language anxiety levels may decrease due to the occurrence of learning experiences that are more positive (Mihaljević Djigunović, 2004).

Nevertheless, it may also be presumed that in some cases high language anxiety levels may prevail throughout the whole language learning process due to some intervening factors (Ganschow et al., 1998). It may then be expected that in spite of similar tuition and study time some students, like dyslexics, do not benefit from classroom instruction as their unimpaired peers.

According to the Linguistic Coding Differences Hypothesis, 'the primary causal factors in successful or unsuccessful FL learning are linguistic; that is, students who exhibit FL learning problems have overt or subtle native language learning differences that affect their learning of a foreign language' (Ganschow et al., 1998, p. 248-9). Skills in the native language components are believed to provide the basic foundation for foreign language learning (Spolsky, 1989), so L1 and L2 learning processes are interrelated, as they reflect basic language functions (Ganschow et al., 1995). In other words, general language skills affect the acquisition of the foreign language (Sparks et al., 1991).

It then follows that in a dyslexic student a deficit in the phonological component of a native language is even more threatening when a foreign language is to be studied. Then dyslexic students are likely to have more FL learning problems and be anxiety-prone with their long-term school failures and emotional disturbances (Jurek, 2004). While a high-anxiety learner may suffer from a variety of academic, cognitive, social and personal language anxiety effects (MacIntyre, 1999), a dyslexic student is even more prone to affective disorders (Gindrich, 2002). Developmental dyslexia may then be a significant cause for experiencing high language anxiety levels (Ganschow et al., 1998) that may not easily decrease in the course of time due to anxiety being a consequence 'of differences in basic language competence' (Ganschow & Sparks, 1996, p. 208), leading to 'an impairment in the representation and manipulation of phonemes' (Fisher & DeFries, 2002, p. 767).

Method

Many learning theories view growing proficiency or learning effects as a function of time, ability and motivation (Johnston & Aldridge, 1985; Karweit, 1985). Consequently, it is often believed that more study time leads to 'a high level of sophistication in the language' (Smythe, Stennett & Gardner, 1974, p. 21). Nevertheless, the research in the field of foreign language acquisition shows that in the natural setting it may not be the case beyond a five-year term exposure (Cummins, 1981; Patkowski, 1980). Still, it is expected in the present research that language anxiety levels generally decrease in the course of language learning experience (time effect). Hence the measurements of the language anxiety levels over three years of study at the secondary school will regularly lower due to the significant investment of time, as far as students with no learning disabilities are concerned.

Nevertheless, the main issue to be investigated is the relationship between developmental dyslexia and language anxiety. It is then presumed that developmental dyslexia is connected with relatively high language anxiety experience throughout the course of secondary school study. Due to the interplay of developmental dyslexia with emotional problems, it is hypothesized that learners suffering from dyslexia symptoms will experience higher level of language anxiety irrespective of the length of their experience with English, which should not be the case of their non-dyslexic peers. Accordingly, the first hypothesis is as follows:

H1: Dyslexic FL students experience higher language anxiety levels than their unimpaired peers.

Another factor taken into consideration while measuring language anxiety levels is gender. The investigation of its role in the way language anxiety is experienced has led to inconsistent results. In some studies boys are found to display higher language anxiety levels during a language course (Campbell, 1999; Campbell & Shaw, 1994). Other studies show no significant differences between FL anxiety levels in boys and girls (Aida, 1994) or prove that girls generally demonstrate higher general anxiety levels (Jose & Ratcliffe, 2003). Nevertheless, the relationship between developmental dyslexia and language anxiety is the main focus of the study. According to Jurek (2003), dyslexic girls declare more acute emotional problems. Accordingly, the next hypothesis is:

H2: Dyslexic girls declare higher language anxiety levels than dyslexic boys.

The participants of the study were 393 students (N=393) coming from 17 classes of the six secondary grammar schools in Opole. There were 127 boys and 266 girls. At the beginning of the study their average age was 16 with the minimum of 15 and maximum of 18 years of age. They all attended classes with typical English programmes: three to five hours a week. English was the second obligatory foreign language (L3) for them with French or German chosen as a priority (L2).

The design of the study is a longitudinal, time-series design using multiple measures taken at different points of time (Graziano & Raulin, 1993). The basic

instrument used in the study was a questionnaire. It consisted of the 33-item Foreign Language Classroom Anxiety Scale (FLCAS) by Horwitz, Horwitz and Cope (1986). The maximum number of points in the scale is 165 on a 5-point Likert scale (1 – I strongly disagree to 5 – I strongly agree). Its reliability was assessed by means of Cronbach's α = .9354 in the first wave, then .9393 and .941 in the second and third, respectively.

Another instrument was the Adult Dyslexia Checklist (Vinegrad, 1994, translated into Polish by Bogdanowicz and Krasowicz, 1996) with *yes* and *no* answers, was also applied (Cronbach's α = .7271). Some additional items explored gender, self-assessment of FL skills with a Likert scale of 1 to 6 for the highest grade, semester grades (similarly 1 – 6) and the length of learners' experience with English.

On the basis of the Revised Adult Dyslexia Checklist results (Vinegrad, 1994), the sample was divided into quartiles. The lower quartile (<=22) accommodated 142 students with no dyslexia symptoms (103 girls and 39 boys) and the upper quartile (>=26) comprised a group of 105 students with developmental dyslexia symptoms (63 girls and 42 boys).

The data collection procedure took place over three years between 2002 and 2005. The language anxiety measurements were taken in December 2002, December 2003 and January 2005. In the second wave of the study dyslexia symptoms were checked. Each time, in each class, the students were asked to fill in a questionnaire. The time slotted for the activity was 15 to 45 minutes, depending on the speed at which the participants worked. They were asked to give sincere answers without taking time to think. Each part of the questionnaire was preceded by a short statement introducing a new set of items in an unobtrusive manner.

Three kinds of variables are identified in the study. The dependent one is language anxiety. The independent variables are gender and symptoms of developmental dyslexia, while the moderator variable is the length of one's experience with English. All the variables are operationally defined as questionnaire items.

The data were computed by means of a statistical programme STATISTICA, with the main operations being descriptive statistics (means – arithmetic average and SD showing how far individuals vary from the mean), together with correlation. That is 'an estimate of the degree to which two sets of interval scale scores go together' (Brown, 1988, p. 132), represented by the Spearman rank-order correlation coefficient *R*. There is also the correlated student's t-test and t-test for independent samples, exploring mean differences between two groups of subjects (Graziano et al., 1993).

Results

As far as the length of the students' exposure to English is concerned, in the case of dyslexic students the average is 5.7 with the minimum of 1 year and the maximum – 12 years. In the case of non-dyslexic learners the average is 6.3 years (min. 1 and max. 13 years). There is no significant difference between the results (t = -1.5; p = .1116), which means that both groups are comparable in relation to the length of their experience with English.

The levels of language anxiety of dyslexic students were respectively: 93.17 (SD = 23.49), and 91.3 (SD = 24.25) and 88.72 (SD = 24.68) in the three waves. The correlated t-test results show that there was a statistically significant decrease of language anxiety levels only when comparing the first and third wave results (t_{W1-2} = 1.06, p = .29; t_{W2-3} = 1.79, p = .077; t_{W1-3} = 2.19, p = .03). The summary of descriptive results is presented in Table 1 below.

Table 1: Summary of descriptive statistics results

	DYSLEXICS (N = 105)		NON-DYSLEXICS (N = 142)	
	Mean	SD	Mean	SD
Language anxiety Wave 1	93.17	23.49	80	22.73
Language anxiety Wave 2	91.3	24.25	76.78	21.27
Language anxiety Wave 3	88.72	24.68	74.65	22.17

As far as the students free from dyslexia symptoms were concerned, their language anxiety levels were: 80 (SD = 22.73) in the first wave, 76.78 (SD = 21.27) in the second and 74.65 (SD = 22.17) in the third wave. The correlated t-test measurement showed a statistically significant decrease in the second wave (t = 2.51, p = .013) but not a significant one between Wave 2 and 3 (t_{W2-3} = 1.49, p = .139) (see Fig. 1 below).
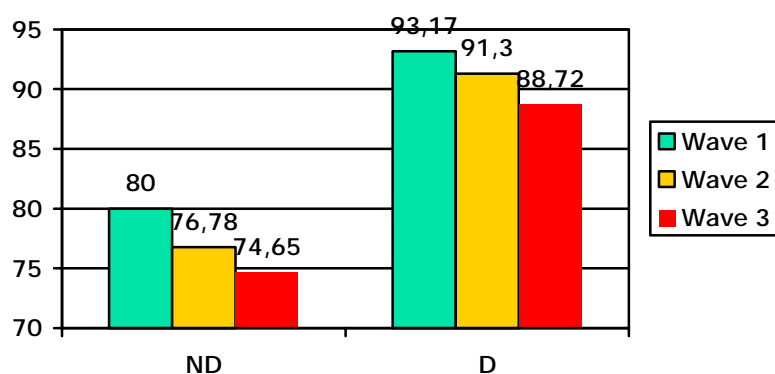
Figure 1: Language anxiety results for dyslexic (D) and non-dyslexic (ND) students

The differences between the results of language anxiety measurements in the two groups in the three waves are all significant: t_{W1} = 4.42, p = .0000; t_{W2} = 5, p = .0000; t_{W3} = 4.7, p = .0000.

As far as girls with dyslexia symptoms are concerned, their anxiety levels at the three points of measurement were 96.11 (SD = 26.65), 96.59 (SD = 26.12) and 93.95 (25.21). No statistically significant decrease of language anxiety levels can be observed in time with t_{W1-2} = -.2, p = .8405 and t_{W2-3} = 1.47, p = .146.

Girls without dyslexia symptoms declared their language anxiety at the following levels: 82.2 (SD = 23.57) in the first wave, 78.31 (SD = 21.52) in the second and 77.47, SD = 22.65) in the third one. The decreasing tendency for language anxiety results follows the general pattern of the whole non-dyslexic group, i.e., there is a statistically significant fall in Wave 2 (t_{W1-2} = 2.37, p = .019), while such a decrease is not observed in the next wave (t_{W2-3} = .464, p = .643). Obviously, there is a statistically significant reduction found while comparing Waves 1 and 3 (t_{W1-3} = 2.43, p = .017).

In all the three measurements the girls with dyslexia symptoms displayed significantly higher language anxiety results when compared to the girls without them (t_{W1} = 3.5, p = .000; t_{W2} = 4.9, p = .0000; t_{W3} = 4.35, p = .0000).
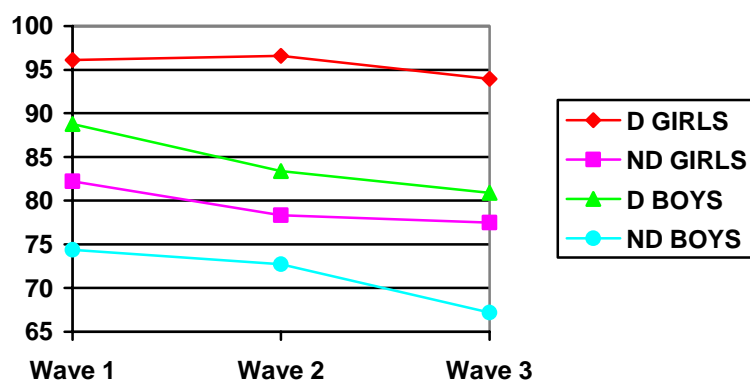
Figure 2: Language anxiety levels of dyslexic and non-dyslexic girls and boys

On the other hand, boys with dyslexia symptoms obtained 88.76 (SD = 17.1), 83.38 (SD = 18.78) and 80.88 (21.89). In their case a decrease can be found with the measurements of language anxiety levels in Wave 1 and 2 (t_{W1-2} = 2.12, p = .04) but not in Wave 2 and 3 (t_{W2-3} = 1.02, p = .311).

As far as boys without dyslexia symptoms are concerned, they scored 74.38 (SD = 19.49), 72.74 (SD = 20.29) and 67.2 (SD = 19.19), respectively. In their case there is no statistically significant fall between Waves 1 and 2 (t_{W1-2} = .84, p = .4), while there is one between Waves 2 and 3 (t_{W2-3} = 2.68, p = .01) and obviously between Waves 1 and 3 (t_{W1-3} = 3.38, p = .002). In all the three measurements the two groups differed in a statistically significant manner (t_{W1} = 3.5, p = .000; t_{W2} = 2.4, p = .016; t_{W3} = 3, p = .004).

Finally, it is extremely interesting to investigate the relationship between language anxiety levels and the length of English language study. In the case of all the participants there is quite a strong and statistically significant correlation, although the results of the non-dyslexic group show a slightly stronger relationship. Generally speaking, these correlations, quite stable for the three waves, show that both dyslexic and non-dyslexic students who had a longer training in English experience lower levels of language anxiety (for the summary of the results see Table 2 below).

Table 2: Correlations between language anxiety and the length of EFL study

	DYSLEXICS (N = 105)		NON-DYSLEXICS (N = 140)	
	R	p	R	p
Wave 1	-.35	.000	-.42	.000
Wave 2	-.38	.000	-.36	.000
Wave 3	-.32	.001	-.42	.000

Discussion

The findings of the present study generally corroborate the model of language anxiety development proposed by MacIntyre and Gardner (1991). In the three waves of the research language anxiety levels decreased significantly in the group of secondary school students who were free from dyslexia symptoms. In the fist wave, three months after entering the secondary school, their language anxiety levels were highest, in spite of anxiety playing 'a negligible role in proficiency' (MacIntyre et al., 1991, p. 110) in the case of beginner students. The finding can be attributed to the fact that apart from facing a new challenging course, the participants had just experienced a transition from a different school type, which was likely to increase their worry and anxiety (Pappamihiel, 2001). Even though the sample comprised a group of students whose experience with the EFL process was relatively short (less than a year), it was found that they still suffered from very high language anxiety levels. That may be attributed to the fact that although in all the studied classes the course of English started as a beginner one, a majority of the students had already had a substantial exposure to formal teaching of English, either in the previous school (junior high) or in private tutorials. In this situation the real beginner students were forced to catch up with their false beginner peers, which might negatively influence their emotions and attitudes about language learning. This provides evidence that false beginners may have relaxed at the beginning of their foreign language experience but already suffered from high anxiety when the study began. Real beginners should have experienced less anxiety, as MacIntyre and Gardner suggest (1993), but, like in the case of false beginners, their anxiety levels were already high three months after the school started. This would mean that lower anxiety levels may occur at the very beginning of language exposure, in less than three months.

Nevertheless, almost all the participants' language anxiety levels lowered towards the end of the secondary school study. That may lead to an optimistic conclusion that the students were able to encounter more favourable experiences and develop positive emotions accompanied by growing proficiency and experience connected with the FL learning process.

It yet needs to be added that as far as the dynamics of the language anxiety decrease is concerned, no significant fall was observed in the last measurement (the last grade of the secondary school). That finding can be attributed to the fact that the students were facing the obligatory maturity examinations in five months. That may have constituted a very influential stressor. Nevertheless, it may generally be concluded that the secondary school offers a general language education for all students in which past or initial negative language experiences are regularly eliminated due to occurrence of positive experience and increased achievement (Young, 1994), as deduced from the lowering language anxiety levels observed throughout the whole length of secondary school experience.

It is also worth mentioning that all the mean language anxiety results obtained by the non-dyslexic group in the three grades of the Polish secondary schools are considerably lower than the measurements of language anxiety levels presented in the literature of the field so far (e.g., Aida, 1994; Horwitz et al., 1991). It can be ex-

plained by the vast popularity of the English language in Poland. It is apparent that a majority of Polish students choose English as a compulsory foreign language at school. Moreover, English in the case of the participants of the study was their L3, as the first foreign language was either French or German. As a low-stakes language, English might not constitute a debilitative challenge. Nevertheless, that finding may also be attributed to the transfer of study skills from L2 to L3.

It can also be inferred that anxiety levels also decrease through the process of experience growth directly connected with time because a significant investment of time, in this case – three years – leads to effective learning (Fredrick & Walberg, 1980) irrespective of talent in a particular area (Ericsson & Charness, 1993) (see Hwang et al., 2004 for a review of research on learning time and learning effects).

Almost all the participants' language anxiety levels fell remarkably throughout the length of their secondary school study. As far as students with developmental dyslexia are concerned, their language anxiety levels underwent a notable decrease only in relation to the comparisons between their initial and final measurement (Waves 1 and 3). Yet, they remained significantly higher in comparison to the results obtained by non-dyslexic students.

According to Young (1999), there are various sources of language anxiety. Some of them may arise from individual differences, where poor native language skills can be placed. As empirical research on the origins of language anxiety is still scarce (MacIntyre, 1999), it then seems worthwhile to shed more light on the relationship between language anxiety and developmental dyslexia as foreign language acquisition is affected by general language skills (Sparks et al., 1991). The results obtained in the research allow for corroborating the hypothesis, according to which *dyslexic FL students experience higher language anxiety level*. Students with symptoms of developmental dyslexia were found to suffer from very high language anxiety levels. It is believed that developmental dyslexia is connected with emotional problems whose onset is evoked by frustration centering on dyslexics' inability to meet expectations (Ryan, 1992). According to a dyslexic adult:

> When you have a learning problem it puts a lot of stress and pressure on you. It robs you of being a whole person since you are not able to reach your potential. Your life is hellish since you think you are intelligent, yet you are not accomplishing anything. As you watch time just slipping by, you feel very depressed, realizing your life is going nowhere. (Stein, 1987, p. 409)

It then follows that the present research results allow for accepting the Linguistic Coding Differences Hypothesis, according to which successful FL learning can be attributed to linguistic factors deeply rooted in one's learning disabilities. The study shows that in the case of students with developmental dyslexia symptoms their significantly higher language anxiety levels are possibly caused by their native language learning differences. Developmental dyslexia negatively affects the FL acquisition process (Schneider, 1999), thus without proper correction and compensation intervention, dyslexic students do not get a chance of combating its

negative influences on their own (Schneider & Crombie, 2003). Still, a prolonged exposure to the foreign language seems to be helpful in more positive language anxiety management of non-dyslexic students in general and dyslexic girls in particular.

In the case of dyslexic students, it is expected that skills mastered through correction compensation intervention offered in one language can be transferred to another (Petrus & Bogdanowicz, 2004), thus the students who get such feedback in English or in Polish may more effectively combat negative effects of language anxiety. Still, extra classes applying multisensory language instruction are found to offer substantial help to dyslexic students (Bogdanowicz, 1997; Nijakowska, 2004).

The next hypothesis focuses on the role of gender in experiencing language anxiety (*dyslexic girls declare higher language anxiety levels than dyslexic boys*). Following the claim made by Oxford, Ehrman and Lavine (1991) about gender differences in reactions in the FL classroom, it is expected that boys and girls differ also in ways of perceiving language anxiety. The present research shows that throughout the whole length of the study the female participants, no matter whether dyslexic or not, declared higher language anxiety than males. It corroborates general findings that female students are in the main more worried and anxious than boys (Gierl & Rogers, 1996). Non-dyslexic girls reported decreasing language anxiety levels in the second wave of the study, when they attended their second grade. It is assumed that their attitudes to language learning became more positive (MacIntyre et al., 2002), while they grew familiar with the learning environment and the foreign language learning experience. Still, in their last grade, with the final exam approaching, their language anxiety levels remained relatively stable.

Unfortunately, that is not the case of girls with dyslexia symptoms. Although they are found to study more and work harder (Jurek, 2003), due to their learning disabilities, they are described to show more emotional and behavioural problems (Hellendoorn & Ruijsenaars, 2000). Consequently, in the present study their experience of language anxiety is found to be a painful affliction with its extremely high and stable levels. That finding stands in agreement with a general gender research strand according to which females are more vulnerable to stress, depression and anxiety, although it is not clear yet whether gender differences are due to environmental or psychological factors in the case of students with dyslexia (Riddick, Sterling, Farmer & Morgan, 1999).

Boys, on the other hand, are described to be able to reduce fear and anxiety (Byrne, 2000) and are generally less anxious than girls (Ginsburg & Silverman, 2000). In the present research males' experience of language anxiety was not as severe as girls' – they had a general tendency to appear less anxious.

Boys without dyslexia symptoms seemed to be the group least affected by language anxiety. Throughout the length of the research procedure their language anxiety levels dropped significantly. It is also worth noting that, contrary to those of non-dyslexic girls' and boys' with dyslexia symptoms, these levels remained relatively stable towards their second grade, which seems to lead to a conclusion that they did not feel at ease with the foreign language yet and needed more time to work out more positive strategies allowing them to cope with negative emo-

tions involved with the process of FL acquisition. Nevertheless, in the course of time, even despite the final exams approaching, they constituted the only group where a significant decrease of language anxiety levels could be observed. It seems then that non-dyslexic boys needed more time for adapting to the academic requirements but later did not experience much further pressure connected with foreign language learning.

On the other hand, the management of language anxiety by dyslexic boys resembles the pattern displayed by non-dyslexic girls – they quickly adapt to the new school requirements and successfully manage their language anxiety. Yet, with the final exams approaching, their language anxiety does not decrease significantly. Generally, they are found to perceive language anxiety in a more acute manner than students without dyslexia symptoms.

The results of the research can be summed up in the following points:

- Dyslexic secondary school learners suffer from significantly higher language anxiety than their non-dyslexic peers.
- Language anxiety levels decrease with growing FL mastery gained over time.
- Real beginner students may suffer from high language anxiety levels due to intervening factors, like school transition or false beginners' pressure.
- Students with dyslexia symptoms, especially girls, need more time to adapt to the conditions of a new school. Their language anxiety levels decrease in the last grade.
- Girls suffer from significantly higher language anxiety levels than boys.
- Girls with dyslexia symptoms experience highest language anxiety levels that remain stable throughout their FL language experience in the secondary school.

The study is certainly not free from limitations. The instrument used for measuring symptoms of developmental dyslexia is the Revised Adult Dyslexia Checklist (Vinegrad, 1994). It is a preliminary tool whose indications need to be confirmed by a team of specialists, which was impossible in the research due to practical reasons. Hence, the dyslexia measurements are tentative and must be treated with care. Although the research design is longitudinal, only plausible influences can be inferred.

As a final comment it is worth stressing that these findings are not against the introduction of a foreign language to children as early as possible. The results of the study show that, irrespective of their language deficits, the students who have a longer experience with English suffer from lower anxiety levels. Consequently, their proficiency is more likely to grow when anxiety decreases. After a longer FL exposure such students are able to eliminate the negative effects of language anxiety more successfully, regardless of their developmental dyslexia handicap. Moreover, with an early treatment of dyslexia comprising multi-sensory techniques employing various modes through visual, auditory and kinaesthetic-tactile activities, they will be able to cope with the foreign language more effectively (Schneider et al., 2003). The long-term effects of anxiety lead to an increased risk of

learned helplessness or lower expectations of success (Riddick et al., 1999), hence an early FL exposure and diagnosis of developmental dyslexia seem to constitute a valuable tool of developing an efficient and well-adjusted student.

References

Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's constructs of foreign language anxiety: The case of students of Japanese. *Modern Language Journal, 78*, 155-168.

Alcott, M. (1997). *Children with special educational needs*. London: Hodder & Stoughton.

Barga, N. K. (1996). Students with learning disabilities in education: Managing a disability. *Journal of Learning Disabilities, 29*, 413-421.

Bednarek, D. (2002). Specyficzne trudności w czytaniu w świetle najnowszych badań [Specific reading difficulties in the light of the latest research]. *Kosmos, 51*, 57-67.

Bogdanowicz, M. (1997). *Integracja percepcyjno-motoryczna: teoria – diagnoza – terapia.* [*Perception and motor integration: Theory – dignosis – therapy*] Warszawa: Centrum Metodyczne Pomocy Psychologiczno-Pedagogicznej Ministerstwa Edukacji Narodowej.

Bogdanowicz, M. & Krasowicz, G. (1996). Kwestionariusz objawów dysleksji u dorosłych (KODD) [Inventory of dyslexia symptoms in adults]. Retrieved September 20, 2003 from http://dysleksja.univ.gda.pl

Brown, J. D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design*. New York: Cambridge University Press.

Byrne, B. (2000, March). Relationships between anxiety, fear, self-esteem and coping strategies in adolescence. *Adolescence*. Retrieved December, 27, 2004 from http://www.highbeam.com/library/docfree.asp

Campbell, C. M. (1999). Language anxiety in men and women: Dealing with gender difference in the language classroom. In D. J. Young (Ed.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom* (pp. 191-215). Boston: McGraw-Hill.

Campbell, C. M., & Shaw, V. M. (1994). Language anxiety and gender differences in adult second language learners: Exploring the relationship. In C. A. Klee (Ed.), *Faces in a crowd: The individual learner in multisection courses* (pp. 215-246). Boston: Heine & Heine.

Crombie, M. A. (2000). Dyslexia and the learning of a foreign language: Where are we going. *Dyslexia, 6,* 112-123.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics, 2*, 132-149.

Downey, D. M., Snyder, L. E., & Hill, B. (2000). College students with dyslexia: Persistent linguistic deficits and foreign language learning. *Dyslexia*, 6, 101-111.

Ericsson, K. A., & Charness, N. (1993). Expert performance: Its structure and acquisition. *American Psychologist, 49,* 725-745.

Fisher, S. E., & DeFries, J. C. (2002). Developmental dyslexia: Genetic dissection of a complex cognitive trait. *Nature Reviews. Neuroscience, 3,* 767-780.

Fredrick, W. C., & Walberg, H. J. (1980). Learning as a function of time. *Journal of Educational Research, 73,* 183–194.

Ganschow, L., & Sparks, R. L. (1995). Effects of direct instruction in Spanish phonology on the native-language skills and foreign language aptitude of at-risk foreign language learners. *The Journal of Learning Disabilities, 28,* 107-120.

Ganschow, L., & Sparks, R. L. (1996). Anxiety about foreign language learning among high school women. *The Modern Language Journal, 80,* 199-212.

Ganschow, L., Sparks, R. L., & Javorsky, J. (1998). Foreign language learning difficulties: An historical perspective. *Journal of Learning Disabilities, 31,* 248-58.

Ganschow, L., & Sparks, R. L. (2000). Reflections on foreign language study for students with language learning problems: Research, issues and challenges. *Dyslexia, 6,* 87-100.

Gardner, R. C., & MacIntyre, P. D. (1993). A student's contribution to second-language learning. Part II: Affective variables. *Language Teaching, 26,* 1-11.

Gierl, M., & Rogers, T. (1996). A confirmatory factor analysis of the test anxiety inventory using Canadian high school students. *Educational and Psychological Measurement, 56,* 315-324.

Gindrich, P. A. (2002). *Funkcjonowanie psychospołeczne uczniów dyslektycznych* [Psychological and social functioning of dyslexic students]. Lublin: Wydawnictwo UMCS.

Ginsburg, G. S., & Silverman, W. K. (2000). Gender role orientation and fearfulness in children with anxiety disorders. *The Journal of Anxiety Disorders, 14,* 57-67.

Graziano, A. M., & Raulin, M. L. (1993). *Research methods: A process of inquiry.* New York: HarperCollins College.

Hellendoorn, J., & Ruijssenaars, W. (2000). Personal experiences and adjustment of Dutch adults with dyslexia. *Remedial and Special Education, 21,* 227-239.

Horwitz, E. K. (1999). Preface. In D. J. Young (Ed.), *Affect in foreign language and second language learning. A practical guide to creating a low-anxiety classroom atmosphere* (pp. xi – xiii). Boston: McGraw-Hill.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70,* 125-132.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1991). Foreign language classroom anxiety. In E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 27-36). Englewood Cliffs, NJ: Prentice Hall.

Hwang, W, Chang, C., & Chen, G. (2004). The relationship of learning traits, motivation and performance-learning response dynamics. *Computers & Education, 42,* 267-287.

Johnston, K. & Aldridge, B. (1985). Examining a mathematical model of mastery learning in a classroom setting. *Journal of Research in Science Teaching, 22,* 323-346.

Jose, P. E., & Ratcliffe, V. (2004). Stressor frequency and perceived intensity as predictors of internalizing symptoms: Gender and age differences in adolescence. *New Zealand Journal of Psychology, 33,* 145-154.

Jurek, A. (2003). Trudności w nauce języków obcych u uczniów z dysleksją Rozwojową [Foreign language learning difficulties in dyslexic students]. In E. Jędrzejowska (Ed.), *II Europejski Tydzień Świadomości Dysleksji. Dysleksja rozwojowa – rozpoznanie problemu i praca z uczniem dyslektycznym* (pp. 71-132). Opole: Opolgraf, S.A.

Jurek, A. (2004). ABC dysleksji [ABC of dyslexia]. *Języki obce w szkole, 5,* 18-40.

Kappers, E. J. & Veerman, J.W. (1995). The family environment of young adolescents with dyslexia. *Dyslexia, 1,* 108-119.

Karweit, N. (1985). *Time scales, learning event and productive instruction.* Washington, DC: National Education Commission on Time and Learning.

Lefly, D. L. & Pennington, B. F. (1991). Spelling errors and reading fluency in compensated adult dyslexics. *Annals of Dyslexia, 41,* 143-162.

Lyon, G. R., Shaywitz, S. E. & Shaywitz, B. A. (2003). Part I. Defining dyslexia, co-morbidity, teacher's knowledge of language and reading. A definition of dyslexia. *Annals of Dyslexia, 53,* 1-14.

MacIntyre, P. D. (1999). Language anxiety: A review of the research for language teachers. In D. J. Young (Ed.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere* (pp. 24-45). Boston: McGraw-Hill.

MacIntyre, P. D., Baker, S. C., Clément, R., & Donovan, L. A. (2002). Sex and age effects on willingness to communicate: Anxiety, perceived competence and L2 motivation among junior high school French immersion students. *Canadian Modern Language Review, 59,* 589-607.

MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second language learning: Toward a theoretical clarification. *Language Learning, 39,* 251-275.

MacIntyre, P. D., & Gardner, R. C. (1991). Methods and results in the study of anxiety and language learning: A review of the literature. *Language Learning, 41,* 85-117.

MacNulty, M. (2003). Dyslexia and the life course. *Journal of Learning Disabilities, 36,* 636-381.

Michaljević Djigunović, J. (2004). Language anxiety – An important concern in language learning. In N. Murray & T. Thornz (Eds.), *Multicultural perspectives on English language and culture* (pp. 42-51). Tallin, London: Tallin Pedagogical University, University of London.

Nickolson, R. I. (2001). Introduction. Developmental dyslexia: Into the future. In A. Fawcett (Ed.), *Dyslexia: Theory and good practice* (pp. 1-36). London, Philadelphia: Whurr.

Nijakowska, J. (2000). Dyslexia – Does it mean anything to a foreign language teacher? In L. Peer & G. Reid (Eds.), *Multingualism, literacy and dyslexia: A challenge for educators* (pp. 248-56). London: David Fulton.

Nijakowska, J. (2004). Usprawnianie umiejętności odczytywania i zapisywania wyrazów w języku angielskim młodzieży z dysleksją rozwojową [Facilitating the ability to read and write words in English in adolescents with developmental dyslexia]. In M. Bogdanowicz & M. Smoleń (Eds.), *Dysleksja w kontekście nauczania języków obcych*, (pp. 144-155). Gdańsk: Harmonia.

Oxford, R. L., Ehrman, M. E., & Lavine, R. Z. (1991). "Style wars": Teacher-student style conflicts in the language classroom. In S. Magnan (Ed.), *Challenges in the 1990s for college foreign language programs* (pp. 1-25). Boston: Heinle & Heinle.

Pappamihiel, N. E. (2001). Moving from the ESL classroom into the mainstream: An investigation of English language anxiety in Mexican girls. *Bilingual Research Journal, 25,* 1-8.

Patkowski, M. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning, 30,* 449-472.

Petrus, P., & Bogdanowicz, M. (2004). Sprawność fonologiczna dzieci w kontekście nauki języka angielskiego [Phonological ability in children in the context of English language learning]. In M. Bogdanowicz & M. Smoleń (eds), *Dysleksja w kontekście nauczania języków obcych*, (pp. 63-77). Gdańsk: Harmonia.

Reynolds, A. M., Elksnin, N., & Brown, F. R. III (1996). Specific reading disabilities: Early identification and long-term outcome. *Mental Retardation and Developmental Disabilities. Research Reviews, 2,* 21-27.

Riddick, B., Sterling, C., Farmer, M., & Morgan, S. (1999). Self-esteem and anxiety in the educational histories of adult dyslexic students. *Dyslexia, 5,* 227-248.

Ryan, M. (1992). The social and emotional effects of dyslexia. *The Education Digest, 57,* 68-71.

Schneider, E. (1999). *Multisensory structured metacognitive instruction.* Frankfurt am Main: Peter Lang.

Schneider, E. & Crombie, M. (2003). *Dyslexia and foreign language learning.* London: David Fulton.

Sparks, R. L., & Ganschow, L. (1991). Foreign language learning differences: Affective or native language aptitude differences? *The Modern Language Journal, 75,* 3-16.

Spolsky, B. (1989). *Conditions for second language learning.* Oxford: Oxford University Press.

Stein, N. L. (1987). Lost in the learning maze. *Journal of Learning Disabilities, 20,* 409-434.

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice, 1,* 55-64.

Vinegrad, M. (1994). A revised adult dyslexia check list. *Educare, 48,* 21-23.

Young, D. J. (1994). New directions in language anxiety research. In C. A. Klee (Ed.), *Faces in a crowd: The individual learner in multisection courses* (pp. 3-46). Boston: Heine & Heine.

Young, D. (1999). Giving priority to the language learner first. In D. J. Young (Ed.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere* (pp. 124-143). Boston: McGraw-Hill.

Young, D. J., & Kimball, M. (1995). Venerable voices. In T. Dvorak (Ed.), *Voices from the field: Experience and beliefs of our constituents* (pp. 194-228). Lincolnwood, IL: National Textbook Company.

Learner Creativity as a Potentially Important Individual Variable: Examining the Relationships Between Learner Creativity, Language Aptitude and Level of Proficiency

Ágnes Albert

Eötvös Loránd University, Department of English Applied Linguistics, Budapest, Hungary

albert.agnes@vnet.hu

Introduction

It seems to be a trend in individual differences research to examine variables that have already been found to be significant in general or in personality psychology. This is not really surprising because of the cross-disciplinary nature of applied linguistics, and it appears that besides well-established variables like aptitude, motivation, anxiety, learning styles and strategies, recently new concepts, such as extraversion (Dewaele, 2005; Dewaele & Furnham, 1999, 2000) and working memory (Miyaki & Friedman, 1998; Gathercole & Thorn, 1998; Sawyer & Ranta, 2001) are finding their way into this research area, as well. This trend is reflected well by the 2005 publication of Dörnyei's book *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*, which contains chapters on such traditional psychological variables as personality, self-regulation, self-esteem, and creativity.

For most people creativity is an intuitively appealing concept, and everyone seems to know, or feel, or at least have a hunch about what it means. The question arises, nevertheless, if this term is interpreted as a rare phenomenon observable only in the exceptionally talented, in which case its relevance for the millions of average people learning foreign languages is obviously negligible. If, however, creativity is hypothesised to be a special arrangement of those cognitive, motivational, personality or social characteristics that are present in everyone, its effects on second language acquisition cannot be disregarded. A number of researchers (Barkóczi & Zétényi, 1981; Cropley, 1972; Guilford, 1950; Harrington, Block & Block, 1983) believe that the underlying components of creativity are normally distributed in the population. Therefore, creativity, which implicitly involves imagination, unconventionality, risk-taking, flexibility, and creating new classifications and systematisations of knowledge (Sternberg, 1985), might be a factor affecting second language acquisition.

The changing methods of second language instruction, the prominence of methods of communicative and task-based language teaching, which in many cases employ tasks that require students to use their imagination, provide another reason why researching this variable should become imminent. Tasks that involve the use of imagination and the generation of new ideas might provide creative learners with more chance to practise, that is, to produce more comprehensible output, which could lead to greater success in second language acquisition (Swain, 1985). This might be even more so in a foreign language environment, where output is mainly produced in the classroom. Support for this line of argumentation was provided by Ottó (1998), who, in a small-scale study involving Hungarian secondary school learners instructed by communicative methods, found significant positive correlations between different measures of learner creativity and students' end-of year English grades.

Besides examining its effect on language learning success, it would also be interesting to see how creativity relates to other, well-established and well-researched variables. Therefore, the present article is going to investigate the effect of a so far neglected individual variable, creativity, on language proficiency, while also examining its relationship with one of the most influential and most widely researched variable, foreign language aptitude.

Review of literature

In the section below, I attempt to provide a brief review of literature on the two individual difference variables that are in the focus of this article. Concentrating on those points only that are relevant for our purposes, I attempt to identify possible points of interaction between the two variables.

Foreign language aptitude

When someone mentions the term language aptitude, even laypeople think they understand the concept that lies behind it: they tend to assume, probably rightly, that some people have a greater talent for learning foreign languages than others. However, the technical definition of foreign language aptitude is more restricted and more detailed at the same time. On the one hand, it does not imply that some people can learn foreign languages while others are incapable of it; it only concerns the *rate* of learning, that is, progress made over a given period of time, but not ultimate achievement. On the other hand, aptitude is not hypothesised as a unitary construct, but rather as a cluster of different cognitive traits that are advantageous as far as foreign language learning is concerned. The exact nature and the relative importance of these factors thus depend to a great extent on the theory of language aptitude proposed by different authors. Therefore, it seems justifiable to argue along with Dörnyei (2005) for an operational definition of foreign aptitude, who states that since the birth of commercial aptitude batteries "language aptitude is what language aptitude tests measure" (p. 35).

Foreign language aptitude is one of the most influential and most extensively researched individual variables as far as second or foreign language acquisition research is concerned. According to Dörnyei and Skehan (2003), correlations of aptitude and language learning success typically range between 0.20-0.60, which besides motivation and age of onset makes it one of the best predictors of language achievement. It is also one of those individual variables that have the longest research tradition, with investigations starting as early as the 1920s.

Despite these early beginnings, modern foreign language aptitude testing in fact started with John B. Carroll's and Stanley Sapon's work. In the 1950s, these authors devised the Modern Language Aptitude Test (MLAT, Carroll & Sapon, 1959) starting out from a purely empirical approach to test design. Having administered over 40 potentially important tests to learners, they collected data on learners' achievement at the end of a language course. Then they selected the best predictors of language learning success and compiled their test battery which is composed of five parts. The five sections of the test measure four underlying components of foreign language aptitude in a hybrid manner, that is, one subtest does not measure a single ability.

The first underlying factor proposed is *phonetic coding ability*, which is defined as "an ability to identify distinct sounds, to form associations between these sounds and symbols representing them, and to retain these associations" (Carroll, 1981, p. 105.), that is, it refers to the coding and memorising of phonetic material. Another component of language aptitude is *rote learning ability*, which is the "ability to learn associations between sounds and meaning rapidly and effectively and to retain these associations" (Carroll, 1981, p. 105), which refers to the ability to memorise language materials. The two other abilities, grammatical sensitivity and inductive language learning ability, are referred to by a single term linguistic ability by Skehan (1989), who believes that these two are very much similar in nature. In Carroll's original interpretation *grammatical sensitivity* is "the ability to recognise the grammatical functions of words (or other linguistic entities) in sentence structures" (1981, p. 105), while *inductive language learning ability* is "the ability to infer or induce the rules governing a set of language materials, given samples of materials that permit such inferences" (1981, p. 105). These definitions seem to support Skehan's line of argumentation that these two abilities might in fact refer to the passive and active manifestation of a single underlying ability respectively.

Carroll's work in aptitude research is influential for two main reasons: the MLAT is still used in research studies today, and although there are attempts from time to time to develop new instruments, they usually do not turn out to be better predictors of language learning success (Sparks & Ganschow, 2001). The MLAT has also served as a model for other aptitude tests; the Hungarian language aptitude test MENYÉT (Ottó, 2002) for example uses some tasks that are similar to those found in the MLAT, and the underlying components measured by the test are identical to those proposed by Carroll (1981).

The Hungarian language aptitude test MENYÉT was developed by Ottó (2002), and as stated above, it attempts to measure the same underlying components of language aptitude as the MLAT. The test itself is different in some respects, however: it consists of four tasks and each task is believed to measure a

single component of language aptitude unlike the five tasks in the MLAT. Therefore, the 'Hidden Sounds' subtest is intended to measure phonetic coding ability, the 'Language Analysis' subtest is assumed to tap inductive language learning ability, the 'Words in Sentences' subtest is believed to shed light on grammatical sensitivity, while the 'Vocabulary Learning' task is a test of rote learning ability.

Although there have been different foreign aptitude tests developed since the construction of MLAT which differ slightly in the emphasis placed on different factors and also somewhat in the nature of aptitude factors (e.g., the PLAB by Pimsleur, 1966; the Defense Language Aptitude Battery by Petersen & Al-Haik, 1976; the VORD by Parry & Child, 1990), they do not differ radically in their conceptualisation of foreign language aptitude. One exception is the recently developed Cognitive Ability for Novelty in Acquisition of Language as applied to foreign language (CANAL-F) theory and test (Grigorenko, Sternberg & Ehrman, 2000), which is not an empirically derived, rather a cognitive theory driven test of foreign language aptitude. It stresses the role of coping with novelty and ambiguity in foreign language learning. The theory describes five knowledge acquisition processes (selective encoding, accidental encoding, selective comparison, selective transfer, and selective combination) which operate at four levels of processing (lexical, morphological, semantic, and syntactic). It differentiates two modes of input and output (visual and oral), and two types of recall tasks (immediate and delayed) that can be used to test the encoding, storage, and retrieval of information.

For our purposes it seems justifiable to examine the knowledge acquisition processes in more detail, as these might serve as possible points of interaction with the other individual variable examined, creativity. In the CANAL-F theory (Grigorenko, et al., 2000), selective encoding refers to learners' ability to distinguish between information relevant or irrelevant for their purposes, while accidental encoding refers to encoding background or secondary information which can aid comprehension and production later on. Selective comparison is the process by which a learner determines the relevance of old information for a current task, and it is related to the personality variable tolerance of ambiguity, that is the person's "ability to hold contradictory, incomplete, or uninterpretable information in working memory without either rejecting it or coming to premature closure about it" (p. 392). Selective transfer concerns the process of applying decoded or inferred rules to new tasks and contexts, while selective combination refers to synthesising information gained through selective and accidental encoding with existing knowledge and modifying existing schemata when needed.

Creativity

In contrast to language aptitude, creativity is an individual variable that has been mostly ignored in second language acquisition research, except for some recent attempts (Albert & Kormos, 2004; Ottó, 1998). If we disregard the widely-held belief according to which we all know intuitively what creativity is, we will see that one plausible reason for this lack of interest lies in the nature of the scientific con-

cept itself. When trying to define the construct of creativity, the first difficulty encountered is that this concept covers a wide range of distinct but related phenomena: the creative performance or product, the creative person, the creative situation, the creative process and the creative potential (Brown, 1989; Lubart, 1994). Therefore, when attempting to define this concept, one of the first tasks should be restricting the scope of investigation and specifying the area or aspect of creativity that is to be examined. Therefore, the present article will focus on the creative potential only, that is, the cognitive underpinnings of the creative working of the mind. In a similar fashion to foreign language aptitude, we argue for an operational definition of creativity, stating that in this article creativity is understood as scores achieved on a standardised creativity test.

Theories of creativity, similarly to the wide range of issues covered by it, are numerous. Authors working within the psychodynamic (Freud, 1908/1959; Kris, 1952), the humanistic (Csikszentmihalyi, 1988; Maslow, 1968; Rogers, 1954) as well as the socio-psychological (Amabile, 1983, 1996) approaches have put forward theories in an attempt to account for this phenomenon. Although as proponents of recent models of creativity (Amabile, 1983, 1996; Sternberg, & Lubart, 1991, 1996) rightly point out, creativity is probably best hypothesised as a complex interplay of several cognitive, personality, motivational, and social factors, they also state that intellectual abilities are arguably among the most important components of creativity (Lubart, 1994). Therefore, the approach I chose which concentrates purely on the cognitive factors underlying creativity seems to be justifiable.

Guilford (1950) was among the first to put forward a list of cognitive processes involved in creativity. He believed that these processes include sensitivity to problems, creative fluency of production, ability to come up with novel ideas, flexibility of mind, synthesising ability, analysing ability, reorganisation or redefinition of organised wholes, a high degree of complexity of the conceptual structure, and evaluation. However, as Guilford (1959) later developed a comprehensive model of human intellect, he started to focus on divergent thinking, the ability to produce many different ideas in response to a problem, as the prime cognitive component of creativity. He suggested that it was an operation complementary to convergent thinking, the ability to find the correct solution to a problem; the cognitive process that he believed is tapped by the majority of intelligence tests. Divergent thinking is hypothesised to have four relatively independent facets: creative fluency, the ability to produce a large number of ideas, flexibility, the ability to produce a wide variety of ideas, originality, the ability to produce unusual ideas, and elaboration, the ability to develop or embellish ideas, to produce many details (Baer, 1993).

Today intellectual abilities considered to be relevant for creativity are usually grouped into two large categories: basic-level and high-level creativity-relevant abilities (Lubart, 1994). Basic-level creative abilities consist of two types: the above-described divergent thinking and different insight abilities comprising the capacities to notice relevant new information, to compare disparate information, to find relevant connections, and to combine information in a problem-relevant fashion. High-level abilities include problem-finding, problem-definition or re-definition, choosing a useful problem presentation, selecting an appropriate

problem-solving strategy, and evaluating the generated possibilities effectively. If we disregard the somewhat different terminology, it becomes obvious that some of these processes have already been discussed in connection with foreign language aptitude, within the framework of Cognitive Ability for Novelty in Language Acquisition - Foreign (CANAL-F) theory (Grigorenko, et. al 2000). Therefore, the processes that seem to be relevant both for foreign language aptitude and creativity are noticing relevant information (cf. selective encoding), comparing disparate information (cf. selective comparison), finding relevant connections (cf. selective transfer), and combining information in a problem-relevant fashion (cf. selective combination).

Current factor-analytic research suggests that factors of creativity-relevant intellectual abilities tend to load on a common higher-order factor, providing empirical evidence of the autonomous existence of this ability. Carroll (1993), having reviewed and reanalysed 121 datasets, found nine first-order factors: ideational fluency, naming facility, associational fluency, expressional fluency, word fluency, sensitivity to problems, originality/creativity, figural fluency, and figural flexibility, relevant for idea production, which he believes is a basic human characteristic. In the term 'idea production', the notion of 'idea' is to be taken in the broadest possible sense: it can be any verbal proposition, but may also be a gesture, a drawing, or a musical phrase; and 'production' is meant as a process distinct from recognition, identification, selection or comparison. Out of the nine first-order factors comprising it, eight are primarily concerned with the speed of idea production and are differentiated on the basis of the type of idea produced, whereas originality/creativity seems to determine the quality or level of the ability. Based on Carroll's findings, idea production is usually measured by tasks which prompt examinees to quickly think of a series of responses. While this is true for all the tasks used to measure the nine first-order factors, there is a special requirement when our aim is to measure originality/creativity. In that case the task itself needs to be difficult or challenging in order to urge respondents to go beyond the obvious and commonplace answers.

This factor-analytic investigation led to the formulation of Carroll's (1993) three-stratum theory of cognitive abilities, where the concept of idea production is labelled general retrieval ability, the ability which is "involved in any task or performance that requires the ready retrieval of concepts or items from long-term memory" (p. 625). Since it is a fundamental characteristic of factor-analysis that the input data determine the output, that is, the tests and tasks analysed and the scoring procedures employed necessarily influence and possibly constrain the outcome, the domain of general retrieval ability is in need of further research to clarify its structure. This could probably be achieved by devising more appropriate and highly reliable measurement procedures. It is also interesting that although Guilford's (1959) Structure of Intellect model is not compatible with the results of the exploratory factor analysis on which the three-stratum theory is founded, still the domain of general retrieval ability "is chiefly (but not entirely) concerned with Guilford's divergent production operation" (p. 638).

When trying to assess a person's creative potentials, usually two different approaches are taken. One option is measuring several non-cognitive aspects of

creativity, such as personality and motivation, in addition to intellectual processes and intellectual style as it was done by Sternberg and Lubart (1991), who tried to establish individual creativity this way. Although this approach is more in line with current constructs of creativity which state that creativity should be considered as a complex interplay of several cognitive, personality, motivational and social factors (Amabile, 1983, 1996; Sternberg & Lubart, 1991, 1996), it is not feasible in correlational research designs where creativity is only one variable to be measured. The other option, therefore, is to try to assess divergent thinking, the intellectual ability that is thought to be the most characteristic of the creative process (Guilford, 1967; Torrance, 1962). Although tests of divergent thinking have been criticised on many accounts (Jordan, 1975; Kogan, & Pankove, 1974), because of their reported validity, reliability (Cropley, 1972; Harrington, Block, & Block, 1983) and their relative ease of use, they are still widely applied as indicators of individual creativity in research on individual variables (Ghadirian, Gregoire & Kosmidis, 2000-2001; Jung, 2000-2001; Russ & Seja-Kaugars, 2000-2001). As McCrae (1987) pointed out, "although tests like Word Fluency certainly have limited face validity as measures of creativity, their ability to identify creative individuals is an empirical matter, and in fact they are reasonably successful in this" (p. 1258).

Having reviewed the relevant literature in connection with foreign language aptitude and creativity, we can see that there are possible points of interaction. When viewed from a cognitive perspective, there are several processes that are believed to be relevant both for aptitude and creativity, for example selective encoding, selective comparison, selective transfer, and selective combination (Grigorenko et al., 2000; Lubart, 1994). What appears to be problematic, however, is that the instruments available for measuring these individual variables in the Hungarian context do not originate from the cognitive tradition and reflect different theoretical backgrounds. Since measuring instruments tend to determine the construct measured by them to some extent, this might pose problems as far as the hypothesised relationship of the two variables is concerned.

Method

Research questions

The aim of the present study is to find answers to the research questions listed below in the framework of a quantitative, correlational research design.

- Is there a relationship between learners' level of proficiency and their language aptitude in the case of advanced learners?
- Is there a relationship between learners' level of proficiency and any facet of creativity in the case of advanced learners?
- Is there a relationship between language aptitude and creativity in the case of advanced learners?

Participants

The data presented in this paper originate from a correlational study investigating the relationship of various individual variables and language measures in the case of advanced learners. Consequently, the participants were 41 1st year English major students from ELTE who attended language practice or academic skills classes. The distribution of sexes reflects the general trend of humanities majors where females tend to be over-represented; there were 31 female participants and only 10 males in the sample. The age of the participants ranged from 19 to 24 at the time of data collection.

Instruments

The following instruments, which will be discussed in detail below, were used for the purpose of the study: a Hungarian language aptitude test (Ottó, 2002), a standardised test of creativity (Barkóczi & Zétényi, 1981), a paper-based Test of English as a Foreign Language (TOEFL-PBT) practice test (Phillips, 1989), and a standardised C-test (Dörnyei & Katona, 1992).

Language aptitude test (MENYÉT)

Ottó (2002) has recently developed and standardised (Ottó, & Nikolov, 2003) a language aptitude test for Hungarian learners; therefore, I used the Magyar Egységes Nyelvérzékmérő-teszt (MENYÉT) [Hungarian Language Aptitude Test] in order to measure the language aptitude of the participants. The test consists of four sections: in the section 'hidden sounds', respondents are required to identify sounds and connect these to their orthographic symbols. The 'language analysis' section is a translation task, while in the 'words in sentences' part respondents have to identify the grammatical function of certain words. In the last, 'vocabulary learning' section Swahili words and their Hungarian translation are to be memorised then tested.

Creativity test

The standardised creativity test (Barkóczi & Zétényi, 1981) available for Hungarian adults consists of five parts, but as the first task is only meant to serve as a warm-up, only the remaining four tasks are scored. There is a time limit set for each task, and the participants are not allowed to go back to previous tasks. The warm-up task is a sentence completion exercise, in which respondents are asked to finish sentences within three minutes. The first two evaluated tasks require verbal responses from the participants. In the task called 'unusual uses' respondents have to invent unusual uses of everyday objects such as a brick. In the 'distant associations' task (in a similar fashion to Mednick's (1962) Remote Associates Test), stu-

dents have to create associations on the basis of the common characteristics of two unrelated words (e.g., given the words *cannon* and *sky* think of a word related to both of them but in different ways: *thunder*). The last two tasks are drawing tasks (based on the Torrance Tests of Creative Thinking; Torrance, 1966). Respondents are asked to draw as many pictures as they can, starting out from the shape of a circle; and to finish abstract shapes in a creative manner. The four tasks last for five, six, eight and ten minutes respectively.

Proficiency tests (TOEFL-PBT, C-Test)

Two tests were used for measuring the language proficiency of the participants. One of them is a C-test standardised for Hungarian learners of English (Dörnyei & Katona, 1992) and the other one is a practice test for the paper-based version of the Test Of English as a Foreign Language (TOEFL-PBT) (Phillips, 1989). Despite the fact that it does not measure oral skills, the TOEFL-PBT is a comprehensive test consisting of three parts: listening comprehension, structure and written expression, and vocabulary and reading comprehension. The test format of TOEFL-PBT is multiple choice; therefore, it requires a different answering technique and can be expected to tap different aspects of language proficiency than the C-test.

Procedures

The participants filled in the creativity test (Barkóczi & Zétényi, 1981) in September 2001, the TOEFL-PBT test (Phillips, 1989) in November 2001, and the C-test (Dörnyei & Katona, 1992) and the language aptitude test (Ottó, 2002) in May 2002 in the presence of the researcher. The reason for this is that all of the above tests are timed, and should reflect the knowledge or skills of the individual; therefore, the presence of an invigilator is required.

Analysis

The following section contains the type of data gathered with the help of the aptitude, creativity and language proficiency tests. Data analysis was performed on these variables with the help of the statistical package SPSS for Windows.

Measures of language aptitude

Since the aptitude test (Ottó, 2002) used in this study is protected by copyright, the participants' answers, recorded on answer sheets, were scored by the author of the aptitude test. Scores for all of the four traditionally measured components of language aptitude (Carroll & Sapon, 1959) were calculated for each respondent: phonetic coding ability, inductive language learning ability, grammatical sensitivity, and rote learning ability.

Measures of creativity

The scoring of the standardized creativity test was carried out in accordance with the process specified by Barkóczi and Zétényi (1981). Each item of the test was scored for three out of the four measures of creativity as defined by Baer (1993) (the standardized Hungarian test of creativity does not measure elaboration), and the sub-scores were added up for the different tasks. Therefore, each of the four sub-sections of the test received three scores independently, a fluency score, a flexibility score, and an originality score.

The *fluency* score, which in this article is called *creative fluency* in order to differentiate it from the temporal variable also called fluency which is commonly used in SLA research, equals the number of responses given, while the *flexibility* score reflects the number of categories the subjects select their answers from (the categories were set up in the course of the standardisation procedure by Barkóczi & Zétényi, 1981). The *originality* score is assigned on the basis of a list containing an index calculated from the statistical frequency of the given response (set up in the course of the standardization procedure by Barkóczi & Zétényi, 1981). Originality scores of different items vary from 0.01 to 0.99 points, while flexibility and creative fluency scores of each response are worth 1 point.

The *total creativity* score is calculated by adding up all the creative fluency, flexibility and originality scores of the various sub-tasks. Apart from this total score, different sub-scores are also calculated: *total creative fluency*, the sum of the four creative fluency sub-scores; *total flexibility*, the sum of the four flexibility sub-scores; and *total originality*, the sum of the four originality sub-scores. Since the four tasks differ in their modality, we can also calculate *verbal creativity* and *figural creativity* scores, were verbal creativity scores are calculated by adding up the creative fluency, flexibility and originality scores of the verbal tasks (Unusual uses and Distant associations), while figural creativity scores are calculated in a similar fashion for the two drawing tasks (Circles and Unusual uses).

It is easy to demonstrate that in this scoring system the creative fluency score (more precisely the number of responses the subject produces) influences both the originality and the flexibility total scores significantly and this usually results in high inter-correlations between the three sub-scores of the test. For this reason, the establishment of creative fluency free scores is very important as these could provide information about other facets of the subjects' creativity, regardless of the number of responses they produced. In order to achieve this *relative flexibility* (the ratio of total flexibility and total creative fluency) and *average originality* (the ratio of total originality and total creative fluency) should also be calculated, in line with the procedure specified in the test (Barkóczi & Zétényi, 1981). This way the *total creative fluency* score can be used to measure creative fluency, the *relative flexibility* score to measure flexibility and the *average originality* score to measure originality as defined above. The measures of creativity used in this study are presented in Table 1.

Table 1 : Measures of creativity

Measures of creativity	Description
Total originality	the sum of originality scores, reflecting the statistical rarity of answers
Total creative fluency	the sum of responses given by the respondent on the four sub-tasks
Total flexibility	the sum of flexibility scores, reflecting the number of categories the answers originated from
Average originality	the ratio of total originality and total fluency scores
Relative flexibility	the ratio of total flexibility and total fluency scores
Total creativity score	the sum of total originality, total flexibility and total fluency scores
Verbal creativity	the sum of total originality, total flexibility and total fluency scores on the two verbal tasks (Unusual uses and Distant associations)
Figural creativity	the sum of total originality, total flexibility and total fluency scores on the two drawing tasks (Circles and Picture completion)

Measures of proficiency

The standardised C-test used in this study consists of three texts, which contain 63 gaps altogether. The scoring of the C-test involved adding up the number of correct answers. The TOEFL-PBT test was evaluated according to the procedures specified in the practice test booklet (Phillips, 1989). After adding up the number of correct solution, the raw scores for each of the three sections (listening comprehension, structure and written expression, and vocabulary and reading comprehension) were converted to the appropriate scales, thus enabling the computation of a composite TOEFL-PBT score.

Results and discussion

Before discussing the relationship of the two individual variables under examination and the way they are related to language proficiency independently and jointly, we should first turn our attention to the descriptive statistics of the different variables. Table 2 presents figures for the aptitude test MENYÉT (Ottó, 2002) for the population of my study, 1st year English majors at ELTE. In order to be able to interpret the figures, Tables 3 and 4 provide the same statistics for 1st year university students, and 1st year university students with at least one language examination, respectively (Ottó & Nikolov, 2003). When compared to the national sample of university students, it can be seen that 1st year English majors at ELTE

seem to have considerably higher language aptitude (m=64.60 for English majors, m=55.79 for university students), and their aptitude varies within a much more limited range (st.d=6.68 for English majors and st.d=11.61 for university students). As can be expected, 1st year university students with at least one language exam resemble the language major sample much more (m=61.03), although the variance displayed by their scores (st.d=10.24) indicates that their group is much more heterogeneous than that of language majors. It seems, therefore, that the group of 1st year English majors at ELTE can be characterised by high level of language aptitude, and this group is quite homogeneous with regard to this individual variable.

Table 2: Descriptive statistics of aptitude test MENYÉT (Hungarian Language Aptitude Test) for 1st year English majors at ELTE (N=41)

Variable	hidden sounds	language analysis	Words in sentences	vocabulary learning	language aptitude
Mean	15.87	16.92	14.78	17.02	64.60
Standard deviation	2.27	1.42	3.48	2.97	6.68

Table 3: Descriptive statistics of aptitude test MENYÉT (Hungarian Language Aptitude Test) for 1st year university students (N=130) (Ottó, & Nikolov, 2003)

Variable	hidden sounds	language analysis	words in sentences	vocabulary learning	language aptitude
Mean	13.53	16.28	10.95	15.04	55.79
standard deviation	3.89	3.70	4.40	4.31	11.61

Table 4: Descriptive statistics of aptitude test MENYÉT (Hungarian Language Aptitude Test) for 1st year university students with at least one language exam (N=65) (Ottó, & Nikolov, 2003)

Variable	hidden sounds	language analysis	words in sentences	vocabulary learning	language aptitude
Mean	14.98	17.54	12.48	16.03	61.03
Standard deviation	3.62	3.34	4.12	4.2	10.24

If we look at the measures of the other individual variable, creativity (see Table 5), slightly different conclusions can be drawn when comparing the figures to a representative national sample (see Table 6). When examining the corresponding means, figures suggest that the 1st year language major sample can probably be characterised by a higher level of creativity than the national standard for high school graduates (Barkóczi & Zétényi, 1981). This is not really surprising in the light of the fact that usually the best high school graduates become university students. Standard deviation figures, however, do not reflect a more homogeneous sample in the case of the university population, as they did for language aptitude. Therefore, we can conclude that although participants of the study were probably more creative on average than the Hungarian population having a high school

diploma, individuals with high and low creativity are both represented among them. The reason why scores achieved on the different subtests and not composite scores are used in Tables 5 and 6 is that only these figures were available for the national sample (Barkóczi &Zétényi, 1981).

Table 5: Descriptive statistics of the four subtests of the standardised creativity test for 1st year English majors at ELTE (N=41)

Title of subtest	Unusual uses		Distant associations		Circles		Picture completion	
Variable	mean	st.dev.	mean	st.dev.	mean	st.dev.	mean	st.dev.
Originality	6.60	3.25	5.64	3.34	9.36	3.71	4.79	1.66
Creative fluency	10.85	4.24	9.97	5.34	17.97	5.71	8.09	2.03
Flexibility	9.19	4.63	7.63	4.69	10.73	3.64	8.02	2.05
Average originality	0.57	0.14	0.56	0.14	0.52	0.13	0.57	0.17
Relative flexibility	0.78	0.20	0.69	0.24	0.61	0.17	0.96	0.15

Table 6: Descriptive statistics of the four subtests of the standardised creativity test for a representative sample of high school graduates (N=1,098) (Barkóczi & Zétényi, 1981) Hungarian national standard

Title of subtest	Unusual uses		Distant associations		Circles		Picture completion	
Variable	mean	st.dev.	mean	st.dev.	Mean	st.dev.	mean	st.dev.
Originality	3.58	2.18	3.37	2.08	5.69	3.25	3.68	1.44
Creative fluency	8.63	4.27	7.73	4.61	12.84	6.60	7.59	2.04
Flexibility	7.09	3.67	5.91	3.53	7.55	3.83	6.63	1.77
Average originality	0.39	0.15	0.42	0.13	0.43	0.14	0.49	0.16
Relative flexibility	0.78	0.24	0.73	0.27	0.61	0.24	0.87	0.16

Table 7 presents the participants' language proficiency measures: the scores they received on TOEFL-PBT, and their scores on a standardised C-test. In order to make the interpretation of TOEFL scores easier, it should be noted that most universities in the USA require points of 550 or above for admission to their undergraduate programs (McKeon, 2006). It can be seen that although the mean of our sample is slightly above this level (m=571); standard deviation figures (st.d.=46.61) suggest that there were some students whose proficiency was far below or above this level. It seems that the proficiency level of the sample ranged from intermediate to advanced.

Table 7: Descriptive statistics of language proficiency tests (TOEFL-PBT, standardised C-test) for 1st year English majors at ELTE (N=41)

Variable	TOEFL Listening comprehension	TOEFL Structure and written expression	TOEFL vocabulary and reading comprehension	TOEFL total score	C-test score
Mean	56.00	59.19	56.48	571.31	48.46
Standard deviation	5.12	5.65	4.73	46.61	6.05
Minimum	42	46	44	450	32
Maximum	64	68	63	643	60

Having looked at the descriptive statistics, we should now turn our attention to the relationship of each individual variable under examination with the proficiency measures in an attempt to answer our research questions. Based on the literature it is obvious that language aptitude is an individual variable that determines the rate of progress when learning a foreign language and not the ultimate level of language proficiency that can be attained by a learner. In this sense, without having any information about the time and intensity devoted to studying the foreign language, which is something that is very hard to measure or even to estimate in retrospect, it is not at all obvious whether we should expect to find any relationship between language aptitude and language proficiency or not. When examining the correlations between language aptitude scores and different measures of language proficiency (see Table 8), we can see that in general language aptitude does not seem to influence language proficiency at this level. This is not really surprising in the light of the relatively high mean and low standard deviation that characterised the language aptitude measures of the sample. The only exception is the language analysis subtest measuring inductive language learning ability, which positively correlates ($r=0.34$, $p<0.05$) with the vocabulary and reading comprehension subtest of TOEFL-PBT. While it is not really surprising that such a moderate relationship exists, the direction of the effect is not at all obvious. It can be argued that students characterised by good inductive language learning abilities, who are able to infer the rules governing a set of language materials can be expected to be better readers, or an equally convincing argument can be that good readers, who achieve high scores on reading tests and probably read a lot, eventually improve their inductive language learning ability.

Table 8: Correlations of aptitude test (MENYÉT) scores and scores on language proficiency tests (TOEFL-PBT and C-test) for 1st year English majors at ELTE (N= 41)

variable	TOEFL Listening comprehension	TOEFL Structure and written expression	TOEFL Vocabulary and reading comprehension	TOEFL total score	C-test score
Hidden sounds	0.045	0.171	0.173	0.139	0.064
Language analysis	0.141	0.217	0.340*	0.252	0.173
Words in sentences	-0.003	0.106	0.156	0.086	0.173
Vocabulary learning	-0.002	0.081	0.019	0.024	0.020
Language aptitude	0.043	0.196	0.221	0.156	0.140

* Correlation is significant at the 0.05 level

When looking at the correlations between various composite measures of creativity and language proficiency in Table 9, we can conclude that no statistically significant correlations can be found between creativity and language proficiency in this 1st year English-major sample. This result somewhat contradicts Ottó's (1998) earlier findings, who on a sample of high school students was able to demonstrate a relationship between creativity and end-of-year grades. One possible reason for such a discrepancy is that besides differences in the sample, Ottó used a different test of creativity in his study; moreover, the end-of-year grades that he used as indicators of language proficiency might be influenced by factors other than language proficiency, for example personal preferences of the teacher.

Although there seem to be no empirical studies examining the relationship between creativity and language proficiency, there is at least one new trend in aptitude research (see section 2.1. Grigorenko et al., 2000) which suggests that there are probably some cognitive processes such as selective encoding, selective comparison, selective transfer, and selective combination, which are relevant for both foreign language aptitude and creativity. Therefore, it would not have been very surprising to find positive correlations between these two measures despite the fact that the instruments used in the study have different theoretical backgrounds from the ones that suggest such a relationship.

Table 9: Correlations of composite creativity scores and language proficiency test scores (TOEFL-PBT and C-test) for 1st year English majors at ELTE (N=41)

variable	TOEFL Listening comprehension	TOEFL Structure and written expression	TOEFL Vocabulary and reading comprehension	TOEFL total score	C-test score
Originality	0.236	-0.081	0.124	0.070	0.137
Creative fluency	0.140	-0.147	-0.017	-0.038	0.032
Flexibility	0.193	-0.130	0.055	0.020	0.089
Average originality	0.243	0.027	0.237	0.166	0.243
Relative flexibility	0.154	-0.058	0.097	0.066	0.169
Total creativity	0.195	-0.122	0.056	0.018	0.088
Verbal creativity	0.267	-0.038	0.162	0.134	0.194
Figural creativity	0.043	-0.177	-0.087	-0.123	-0.065

In the light of the above, it is interesting to see that the only correlations that were found to be statistically significant between language aptitude and creativity are negative (see Table 10), and they appear in the case of one subtest of language aptitude only, the 'hidden sounds' subtest (it negatively correlates with creative fluency $r=-0.31$, $p<0.05$, and with flexibility $r=-0.31$, $p<0.05$). Since the 'hidden sounds' subtest supposedly measures phonetic coding ability, that is the coding and memorising of phonetic materials, it is quite difficult to interpret this finding. Two alternative explanations can be offered however: it is possible that the 'hidden sounds' subtest, besides phonetic coding ability, also measures something else, and it is this unknown variable that is negatively related to creative fluency and flexibility, that is to the number of responses given on a task and the number of categories these answers originate from. The other explanation is based on the different format of the aptitude and the creativity tests. While the creativity test used in this study is in fact a divergent thinking test which attempts to measure respondents' ability to generate a large number of solutions in response to a given task, the 'hidden sounds' subtest of the aptitude test is clearly convergent: respondents are required to memorise sounds and their associated orthographic symbols, and to provide the only correct solution on the test. While this explanation seems to be plausible, the question can be raised why no similar relationship was identified in the case of creativity and 'vocabulary learning', since the above described characteristics of the 'hidden sounds' task are in many ways similar to that of 'vocabulary learning'.

Table 10: Correlations of composite creativity test scores and language aptitude test (MENYÉT) scores for 1st year English majors at ELTE (N=41)

variable	Hidden sounds	Language analysis	Words in sentences	Vocabulary learning	Language aptitude
Originality	-0.217	-0.166	0.041	0.163	-0.015
Creative fluency	-0.312*	-0.292	0.039	0.063	-0.119
Flexibility	-0.311*	-0.189	0.056	0.103	-0.071
Average originality	0.038	0.032	0.082	0.173	0.139
Relative flexibility	-0.105	-0.036	0.112	0.053	0.039
Total creativity	-0.287	-0.221	0.047	0.112	-0.070
Verbal creativity	-0.307	-0.192	-0.042	-0.009	-0.163
Figural creativity	-0.163	-0.178	0.134	0.194	0.063

* Correlation is significant at the 0.05 level

Since bivariate correlations suggest that there is a weak negative relationship between the phonetic coding ability component of language aptitude and the creative fluency and flexibility components of creativity, we can hypothesise that their relationship might influence the correlations each of them has with other variables. For this reason, partial correlations were calculated for language aptitude as well as for creativity with measures of language proficiency, instead of the bivariate correlations done earlier. This allowed us to remove the effect of the variable partialled out from the relationship of two variables. Table 11 presents the partial correlations between various measures of the aptitude test and the language proficiency scores, while controlling for the effect of creative fluency and flexibility. We can see that having removed the effect of creative fluency and flexibility the only statistically significant correlation, which existed between scores on the language analysis subtest and the vocabulary and reading comprehension section of TOEFL-PBT (see Table 8), dropped from $r=0.34$ to $r=0.30$, which is not significant at this sample size any more. Although this might suggest that without the mediating effect of creativity, aptitude does not influence language proficiency at this level, there is another plausible explanation as well. It is also possible that on a larger sample the relationship would have remained statistically significant, and in that case it would be impossible to argue for a mediating effect of creativity. Consequently, the issue whether creativity affects the relationship of language aptitude and language proficiency or not cannot be settled.

Table 11: Partial correlations of aptitude test (MENYÉT) scores and scores on language proficiency tests (TOEFL-PBT and C-test) for 1st year English majors at ELTE (N=41), controlling for creative fluency and flexibility

variable	TOEFL Listening comprehension	TOEFL Structure and written expression	TOEFL Vocabulary and reading comprehension	TOEFL total score	C-test score
Hidden sounds	0.106	0.135	0.193	0.145	0.089
Language analysis	0.148	0.183	0.308	0.215	0.148
Words in sentences	-0.019	0.111	0.148	0.078	0.164
Vocabulary learning	-0.036	0.088	-0.010	0.003	-0.047
Language aptitude	0.037	0.179	0.198	0.133	0.124

Table 12: Partial correlations of composite creativity test scores and scores on language proficiency tests (TOEFL-PBT and C-test) for 1st year English majors at *ELTE (N=41), controlling for Hidden sounds (phonetic coding ability)*

Variable	TOEFL Listening comprehension	TOEFL Structure and written expression	TOEFL Vocabulary and reading comprehension	TOEFL total score	C-test score
Originality	0.252	-0.045	0.167	0.103	0.155
Creative fluency	0.162	-0.099	-0.039	0.005	0.054
Flexibility	0.218	-0.081	0.116	0.067	0.114
Average originality	0.241	0.021	0.234	0.162	0.241
Relative flexibility	0.160	-0.041	0.117	0.082	0.176
Total creativity	0.217	-0.077	0.111	0.061	0.111
Verbal creativity	0.295	0.015	0.229	0.187	0.224
Figural creativity	0.051	-0.153	-0.060	-0.102	-0.055

Table 12 presents the partial correlation figures for the composite creativity test scores and scores on language proficiency tests, while controlling for phonetic coding ability (the hidden sound subtest of MENYÉT). As can be seen, no statistically significant correlations were found between creativity and language proficiency in this case either; therefore, it seems that creativity is unrelated to global language proficiency on this sample.

Conclusion

Based on the present study, the following conclusions can be drawn with regard to the relationship of foreign language aptitude, creativity and language proficiency for a sample of 41 advanced learners. When examining the relationship of the two individual variables, it seems that some aspects of creativity (creative fluency and flexibility) and language aptitude (phonetic coding ability) are negatively correlated. Therefore creative learners tend to be at a disadvantage when a task calls for the memorization and retrieval of sound symbol associations. Although we might intuitively feel that such a relationship is easy to explain as good performance on the 'hidden sounds' subtest does not require any creativity and in fact creative learners might want to change the pairs instead of memorizing them, this finding is still surprising in the light of cognitive theories of aptitude and creativity, which suggest there are some processes that are relevant for both individual variables. However, since the instruments used in this study are not grounded in cognitive theory, this finding might purely be an artefact of the measuring instruments applied.

As far as the relationship of aptitude and proficiency is concerned, in the case of the advanced learners who participated in the study, only inductive language learning ability (measured by the language analysis task) seems to be related to language proficiency at the advanced level. In order to be able to analyse the relationship of aptitude and proficiency in greater detail, more data would be needed on the language learning history of the participants. It is also possible that if either the proficiency or the language aptitude of the respondents displayed more variance, it would have been possible to find some relationship between them.

Although a previous study (Ottó, 1998) suggested that there might be a direct relationship between creativity and language proficiency, the findings of the current investigation failed to lend support to this hypothesis. In this sample no statistically significant correlations were found between different facets of learner creativity and students' proficiency. Creativity turned out to be unrelated to the overall level of proficiency. In light of this, it might be interesting to look at specific language tasks instead of general measures of language proficiency to see if creativity manifests itself there or not. It can be hypothesised that even if creativity does not directly affect overall language learning success, as its influence might be counterbalanced by other factors, other individual difference variables in the case of such a global measure as language proficiency, it is possible that creativity affects some aspects of task performance.

References

Albert, Á., & Kormos, J. (2004). Creativity and narrative task performance: An exploratory study. *Language Learning, 54,* 277-310.

Amabile, T. M. (1983). *The social psychology of creativity.* New York: Springer Verlag.

Amabile, T. M. (1996). *Creativity in context.* Boulder, Colorado: Westview Press.

Baer, J. (1993). Creativity and divergent thinking: A task specific approach. New Jersey: Lawrence Erlbaum.

Barkóczi, I., & Zétényi, T. (1981). *A kreativitás vizsgálata.* Budapest: Országos Pedagógiai Intézet.

Brown, R. T. (1989). Creativity: What are we to measure? In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 1-32). New York: Plenum Press.

Carroll, J. B. (1981). Twenty-five years of research in foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83-118). Rowley, MA: Newbury House.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies.* Cambridge: Cambridge University Press.

Carroll, J. B., & Sapon, S. (1959). *The Modern Language Aptitude Test.* San Antonio, TX: Psychological Corporation.

Cropley, A. J. (1972). A five-year longitudinal study of the validity of creativity tests. *Developmental Psychology, 6,* 119-124.

Csikszentmihalyi, M. (1988). Society, culture, and person: A systems view of creativity. In R. J. Sternberg (Ed.), *The nature of creativity: Contemporary psychological perspectives* (pp. 325-339). New York: Cambridge University Press.

Dewaele, J.-M., & Furnham, A. (1999). Extraversion: The unloved variable in applied linguistic research. *Language Learning, 49,* 509-544.

Dewaele, J.-M., & Furnham, A. (2000). Personality and speech production: A pilot study of second language learners. *Personality and Individual Differences, 28,* 355-365.

Dewaele, J-M. (2005). Investigating the psychological and emotional dimensions in instructed language learning: Obstacles and possibilities. *The Modern Language Journal, 89,* 367-380.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Dörnyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing, 9,* 187-206.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty, & M. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Malden, MA: Blackwell Publishing.

Freud, S. (1959). Creative writers and daydreaming. In J. Strachey (Ed.), *Standard edition of the complete psychological works of Sigmund Freud.* (Vol. 9, pp. 143-153). London: Hogarth Press. (Original work published 1908)

Gathercole, S. E., & Thorn, A. S. C. (1998). Phonological short-term memory and foreign language learning. In A. F. Healy & L. E. Bourne (Eds.), *Foreign lan-*

guage learning: Psycholinguisitc studies on training and retention (pp. 141-158). Mahwah, NJ: Lawrence Erlbaum Associates.

Ghadirian, A M., Gregoire, P., & Kosmidis, H. (2000-2001). Creativity and the evolution of psychopathologies. *Creativity Research Journal, 13*, 145-148.

Grigorenko, E. L., Sternberg, R. J., & Ehrman, M. E. (2000). A theory based approach to the measurement of foreign language learning ability: The Canal- F theory and test. *The Modern Language Journal, 84,* 390-405.

Guilford, J. P. (1950). Creativity. *American Psychologist, 5, 444-454.*

Guilford, J. P. (1959). Three faces of intellect. *American Psychologist, 14*, 469-479.

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw Hill.

Harrington, D. M., Block, J., & Block, J. H. (1983). Predicting creativity in preadolescence from divergent thinking in early childhood. *Journal of Personality and Social Psychology, 45*, 609-623.

Jordan, L. A. (1975). Use of canonical analysis in Cropley's "A five year longitudinal study of the validity of creativity tests". *Developmental Psychology, 11*, 1-3.

Jung, D. I. (2000-2001). Transformational and transactional leadership and their effects on creativity in groups. *Creativity Research Journal, 13*, 185-195.

Kogan, N., & Pankove, E. (1974). Long term predictive validity of divergent thinking tests: Some negative evidence. *Journal of Educational Psychology, 66*, 802-810.

Kris, E. (1952). *Psychoanalytic exploration in art*. New York: International Universities Press.

Lubart, T. I. (1994). Creativity. In J. R. Sternberg (Ed.), *Thinking and problem solving* (pp. 289-332). San Diego: Academic Press.

Maslow, A. (1968). *Toward a psychology of being*. New York: Van Nostrand.

McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology, 52,* 1258-1265.

McKeon, D. W. (2006). TOEFL score interpretation. Retrieved April 26, 2006, from Virginia Tech, Graduate School. Online: http://www.grads.vt.edu/igss/faculty_staff/toefl_interp.html

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review, 69,* 220-232.

Miyake, A., & Friedman, D. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguisitc studies on training and retention* (pp. 339-364). Mahwah, NJ: Lawrence Erlbaum Associates.

Ottó, I. (1998). The relationship between individual differences in learner creativity and language learning success. *TESOL Quarterly, 32*, 763-773.

Ottó, I. (2002). *Magyar Egységes Nyelvérzékmérő-Teszt.* Kaposvár: Mottó-Logic.

Ottó, I., & Nikolov, M. (2003). Magyar felsőoktatási intézmények elsőéves hallgatóinak nyelvérzéke. *Iskolakultúra, 13* (6-7), 34-44.

Parry, T. S., & Child, J. R. (1990). Preliminary investigation of the relationship between VORD, MLAT and language proficiency. In T. S. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 30-66). Englewood Cliffs, NJ: Prentice-Hall Regents.

Petersen, C., & Al-Haik, A. (1976). The development of the Defense Language Aptitude Battery. *Educational and Psychological Measurement, 36*, 369-380.

Phillips, D. (1989). *Longman practice tests for TOEFL*. New York: Longman.

Pimsleur, P. (1966). *The Pimsleur Language Aptitude Battery*. New York: Harcourt, Brace, Jovanovic.

Rogers, C. R. (1954). Toward a theory of creativity. *ETC: A Review of General Semantics, 11*, 249-260.

Russ, S. W., & Seja-Kaugars, A. (2000-2001). Emotion in children's play and creative problem solving. *Creativity Research Journal, 13*, 211-219.

Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (Ed.), *Cognition and second language acquisition* (pp. 319-353). New York: Cambridge University Press.

Sparks, R., & Ganschow, L. (2001). Aptitude for learning a foreign language. *Annual Review of Applied Linguistics, 21*, 90-111.

Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. *Journal of Personality and Social Psychology, 49*, 607-627.

Sternberg, R. J., & Lubart, T. I. (1991). An investment theory of creativity and its development. *Human Development, 34*, 1-31.

Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. *American Psychologist, 51*, 677-688.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds*.), Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Torrance, E. P. (1962). *Guiding creative talent*. Englewood Cliffs, NJ: Prentice Ha-ll.

The Role of Working Memory in Intensive Language Learning

Judit Kormos and Anna Sáfár

Department of English Applied Linguistics, ELTE, Budapest, Hungary

kormos.j@chello.hu, annasafar@gmail.com

Introduction

In this paper we discuss the role of verbal working memory capacity in the success of the acquisition of various language skills and competencies during a one-year-long intensive English language training program. Our study addressed the research question what the relationship is between verbal working memory capacity and performance in an end of year reading, writing, listening, speaking and use of English test. The participants of the study were 67 secondary-school students aged 15-16 in the first year of intensive language learning of a bilingual education program. Twenty-one of the students were beginners at the beginning of the school year, while 46 students started from a pre-intermediate level. The participants performed a non-word repetition test and took a Cambridge First Certificate Exam. The non-word repetition test scores were correlated with the results on the various sub-components of the language proficiency test. Our study indicates that verbal working memory capacity plays a different role in the case of beginners and pre-intermediate students in intensive language learning. Whereas in the case of beginners there was a significant correlation between non-word repetition scores and oral test results only, for the pre-intermediate students we found a highly significant relationship between scores on the use of English and writing test as well as the overall proficiency test.

One of the basic questions in second language acquisition (SLA) research is what accounts for students' differential success in language learning. The individual factors that influence language learning have been widely researched in the past 30 years (for a recent overview see Dörnyei, 2005). The variables along which language learners differ are generally sub-divided into affective, cognitive and personality-related individual differences (Gardner, 1985). With some overlaps, motivation, language learning anxiety and self-confidence are generally listed among affective factors, whereas personality-related differences comprise traits such as openness to experience, conscientiousness, extraversion, agreeableness and emotional stability (Costa & McCrae, 1992). The cognitive factors that are held to be important predictors of success in language learning are intelligence, aptitude and working memory capacity.

Research on the role of working memory in SLA started approximately 15 years ago, shortly after the importance of verbal memory capacity in child first language acquisition was discovered (Gathercole & Baddeley, 1989). In their review of the research evidence supporting the crucial role of working memory capacity in learning words,

Baddeley, Gathercole and Papagno (1998) call the phonological loop, i.e. the verbal component of working memory, a "language learning device" (p. 158). Miyake and Friedman (1998) even suggested that working memory is the central component of language aptitude. The role of verbal working memory capacity in the acquisition of second language (L2) words has long been acknowledged (Papagno & Vallar, 1995; Service & Kohonen, 1995). Working memory capacity has also been found to influence the learning of grammatical rules (Ellis & Sinclair, 1996).

Most studies on the role of verbal working memory capacity in L2 learning concentrated on various aspects of vocabulary learning, and they usually used a relatively small number of participants (e.g., Papagno & Vallar, 1995; Service & Kohonen, 1995, Speciale, Ellis & Bywater, 2004). The groups of participants were either highly motivated university students with good cognitive capacities (e.g., Papagno & Vallar, 1995; Speciale et al., 2004) or children (e.g., Service & Kohonen, 1995). Very few research projects investigated the general effect of working memory capacity on the acquisition of various L2 skills such as listening, reading, writing and speaking, and if they did, such as Harrington and Sawyer's (1992) and Fortkamp's (1999) study, they used the measure of reading span, the validity of which as a measure of verbal working memory capacity is questionable (Caplan & Waters, 1999; Waters & Caplan, 1996). Our research aimed to fill the gap of studies on the relationship of verbal working memory capacity and the acquisition of the major L2 skills. We also intended to investigate an under-represented age-group in this field, namely young adolescent learners, who constitute a major proportion of the language learning population in foreign language environments. Since in Hungary, a large number of secondary school students aged between 14 and 15 have been recently participating in intensive language learning programs, we were interested in to what extent the level of proficiency in the various L2 skills attained by the end of a one-year intensive language program correlates with performance on a test of verbal working memory capacity.

In this paper we first describe the working memory model and discuss previous research investigating the role of working memory in language learning. After the description of the methods used in our research, we present the results obtained, and discuss how they support a relationship between working memory and foreign language learning.

Review of literature

The working memory model

The most widely accepted conceptualization of short-term memory today is the working memory model developed by Baddeley and Hitch (1974; Baddeley, 1986). While previous theories of memory systems focused on the storage function of memory, the new model, as its name suggests, adapts a more dynamic approach. This conceptualization of working memory combines storage with the processing and manipulation of information, thus in this view working memory plays a far greater role than previously as-

sumed in cognitive activities such as comprehension, reasoning and learning (Baddeley, 2003).

The working memory model comprises a multi-component memory system consisting of the central executive, which coordinates two modality-specific subsystems, the phonological loop and the visuo-spatial sketchpad. Later, a fourth component was added to the model: the episodic buffer, which uses multi-dimensional coding, integrates information to form episodes, and is in communication with long-term memory (Baddeley, 2000). The visuo-spatial sketchpad works with visual and spatial information, while the phonological loop is specialized for the manipulation and retention of speech. The central executive, "the most important but least understood component of working memory" (Baddeley, 2003, p. 835), has several functions, including attentional control, directing the flow of information through the system and planning (Gathercole, 1999).

The most widely researched component of working memory is the phonological loop. This subsystem consists of a phonological store, which holds information for a few seconds, and an articulatory rehearsal process, which refreshes decaying information amongst other functions. The rehearsal process is analogous to subvocal speech and takes place in real-time, resulting in a limited span of immediate memory (after a certain number of items, the first one will fade before it can be rehearsed). Phonological loop capacity is often measured by tasks involving immediate serial recall of numbers (digit span) or words (Baddeley, 2003). One of the most widely used tests of phonological short-term memory capacity is the non-word repetition test, where participants have to repeat non-words of various lengths. Non-words are words that do not exist in the given language but conform to its phonotactic rules. Participants' short-term memory capacity may then be expressed in terms of the non-word span, which is the highest number of syllables the participant could repeat in at least 50% of the cases.

Working memory and language learning

A great number of studies investigated the relationship between short-term memory capacity and first language acquisition. Young children show considerable variation in both phonological loop capacity and vocabulary knowledge, and these two variables are closely related. The relationship is particularly strong between non-word repetition and native vocabulary knowledge, with correlation coefficients ranging between 0.4 and 0.6 and is independent of nonverbal intelligence (Gathercole & Thorn, 1998).

Gathercole, Service, Hitch, Adams and Martin (1999) found moderately strong correlations between phonological loop capacity and vocabulary knowledge in children aged 4, 5 and 13 years. This shows that the close link between verbal short-term memory and vocabulary knowledge persists throughout the childhood. They also managed to prove that this relationship is not due to individual differences in articulatory skills as some critics suggested. According to this claim, measures of verbal short-term memory require a high degree of articulatory accuracy (as participants have to repeat non-words and a single phoneme deviation is scored as an error), and thus may underestimate the phonological memory capacity of children who have phonological output problems.

However, although Gathercole and her colleagues confirmed that the non-word repetition task is limited by children's speech production skills, they found that there is a strong relationship between vocabulary knowledge and a non-word recognition task, where articulatory skills were not involved. Thus the link between verbal short-term memory and vocabulary knowledge remains even when articulation skills are not involved. Furthermore, the correlations between conventional short-term memory measures and vocabulary knowledge persisted even after the effect of articulation rate was partialled out.

Baddeley (1986) argued that the phonological loop plays a crucial role in the learning of new words by storing unfamiliar sound patterns while long-term representations are built, which supposes a direct link between short-term memory and long-term learning. This link between short-term memory and long-term knowledge, however, is not unidirectional. Research evidence shows that words are easier to recall than non-words (Hulme, Maughan & Brown, 1991), and non-words that conform to the phonotactic rules of the participants' first language are easier to recall than non-words which are less "wordlike" (Gathercole, 1995). (For a recent study on the effect of phonotactic irregularities on short-term memory performance, see Kovács & Racsmány, in preparation). Further evidence for the role of long-term knowledge in phonological short-term memory comes from a study by Gathercole, Hitch, Service and Martin (1997). The authors investigated the links between new word learning, phonological short-term memory and existing lexical knowledge in the case of five-year-old children. They found that the learning of new words was influenced not only by phonological loop capacity, but also by the level of children's vocabulary knowledge at the time. According to the authors, this proves a reciprocal relationship between phonological short-term memory skills and long-term lexical knowledge. "With greater amounts of lexical knowledge, the degree to which knowledge about the structure of the language can be used to supplement and enhance fragile traces of new words in the phonological loop will also be increased" (Gathercole et al., 1997, p. 977).

The link between phonological short-term memory and new word learning was extended to the learning of foreign languages by Service and her colleagues (Service, 1992; Service & Kohonen, 1995). They found that the ability to repeat English-sounding pseudowords was a good predictor of English language learning success (as expressed by children's grades in English) among Finnish primary school pupils during the first three years of language learning. Later, the link between non-word repetition and children's grades was found to be mediated by L2 vocabulary knowledge.

Cheung (1996) investigated the relationship between experimental learning of second-language vocabulary and L2 non-word repetition among 12-year-old high school students whose mother tongue was Chinese. He found that non-word span was the best predictor of second language vocabulary learning among those participants whose vocabulary level was below the group average, while in the case of the subgroup with a wide range of vocabulary no such relationship was detected between the two variables. According to Cheung, this suggests that there is an interaction between phonological short-term memory and long-term phonological knowledge about the L2, which explains that in the case of high-vocabulary participants their long-term knowledge supported the learning of new words.

Masoura and Gathercole (1999) tested children's non-word repetition performance and vocabulary knowledge in both L1 (Greek) and L2 (English). They found significant links between phonological memory skills and both native and foreign vocabulary knowledge, which were independent of general factors (nonverbal ability, age). L2 vocabulary knowledge was associated with non-word repetition independently of native vocabulary knowledge; however, the opposite was not true: after L2 vocabulary scores were partialled out, native vocabulary knowledge did not correlate with non-word repetition. The results may indicate that phonological short-term memory plays an even more important role in foreign language learning than in L1 acquisition. It also confirms previous evidence showing language specificity in memory for non-words.

Most of the studies discussed above involved children, but Papagno and Vallar (1995) showed that short-term memory and word-learning abilities are related among adults as well. They compared Italian university students who spoke at least two foreign languages and students who were not polyglots. The two groups were similar in terms of other characteristics (academic achievement, nonverbal intelligence, etc.). As expected, the polyglot group performed better on tests of phonological short-term memory and in the learning of Russian-Italian word pairs. The two groups showed similar performance in learning Italian word-pairs.

Speciale, Ellis and Bywater (2004) explored the relationships between phonological short-term memory capacity, phonological sequence learning ability and foreign language learning among undergraduate university students in two experiments. They found that both phonological sequence learning and phonological short-term memory capacity contribute to vocabulary learning. In the beginning of learning a language, these two variables are separable and contribute to vocabulary learning independently. As learning progresses, students begin to recognize the phonological regularities of the language, and vocabulary knowledge contributes to increasing the efficiency of short-term phonological storage as well as the learning of further sequences. These results confirm the relationship between short-term memory capacity and foreign language vocabulary learning, while indicating that the ability to learn phonological sequences also contributes to vocabulary learning, and that the combination the two variables has more explanatory power than that of phonological short-term memory capacity alone.

Some researchers claim that working memory plays a more general role in second language acquisition than just supporting vocabulary acquisition. Ellis (1996) argues that language learning is mostly sequence learning, and even abstract grammatical knowledge is a product of the analysis of sequences. As short-term memory is responsible for remembering sequential information, its role in language learning is far greater than previously supposed. Ellis suggests that the acquisition of syntax is also related to short-term memory capacity. Ellis and Sinclair (1996) present experimental evidence that rehearsing foreign language material has beneficial effects on both comprehending and learning foreign language material, metalinguistic knowledge of grammar, accuracy in pronunciation and productive grammatical fluency and accuracy.

As we have seen, there is convincing research evidence for a relationship between phonological working memory capacity and certain aspects of foreign language learning. Our study investigates the role of verbal working memory capacity in the case of two groups of secondary- school students taking part in an intensive language learning

program: those who were beginners at the beginning of the school-year and those whose level of proficiency was judged to be pre-intermediate in September. Our research questions were the following:

1. Is there a difference in the role of verbal working memory capacity between lower and higher level learners?

2. How does verbal working memory capacity influence the acquisition of vocabulary, grammar, reading, writing, speaking and listening skills?

Method

Participants

Our research was conducted in a Hungarian-English dual language secondary school in Budapest. The participants were in a year of intensive language learning program from September to June, in order to achieve a level of proficiency in English which enables them to participate in an immersion-type programme in English in several subjects during their four years of secondary-school studies. These students attended a five-year secondary programme and their age was between 15 and 16 years.

Out of the 72 students enrolled in the program, 67 students participated in the study, 29 participants were male and 38 students were female. 21 students in the program had learned English before coming to the school, and they began learning from the pre-intermediate level in September. 46 students had either not learned English in primary school, or the level of proficiency they reached was judged to be elementary, and therefore they started from beginner level.

The students learn English in six small groups (4 beginners and 2 pre-intermediate) and they are taught by five teachers who collaborate very closely. The teachers in the programme claim that the teaching method used is predominantly communicative.

Procedures

We used a non-word span test to measure students' short-term memory capacity. The test was administered at the end of the academic year in May. As their end-term exam, the students completed a Cambridge First Certificate language exam. We used the results of this test as the measure of second language proficiency. Results on the two tests were correlated. For the statistical analyses we used the SPSS 13.0 for Windows software.

Non-word span test

The Hungarian version of the non-word span test was developed by Racsmány, Lukács, Németh, and Pléh (2005). It consists of 36 non-words which conform to the phonotactic rules of Hungarian. The non-words are presented in a specific order, and participants have to repeat them one by one. Each non-word consists of one to nine syllables, and a set of four non-words corresponds to each number of syllables (the sets are not presented together). We recorded the test material with a digital voice recorder and presented it using portable CD players. Each student was tested separately in a quiet room by the authors of the article.

On the basis of the participants' performance, their non-word span was calculated, which is the highest number of syllables the participant can repeat in 50% of the cases (at least two times). This is the most widely accepted indicator of verbal short-term memory capacity. However, as we found that this indicator was sometimes contradictory (for instance, when a participant repeated two seven-syllable words but only one six-syllable word), we also used the weighted average of the number of repeated syllables (which differentiates between two participants with a seven-syllable non-word span if one repeated all the words below the eight-syllable level, and the other failed to repeat some of the words with a lower number of syllable).

We selected the non-word span test because of its attested validity in assessing verbal working memory capacity (for an overview of the research on the non-word span test see Racsmány et al., 2005). Although the digit-span test is also a valid measure of capacity of the phonological loop, we decided against using it because previous studies showed similar or lower correlations between digit span and vocabulary measures than in the case of non-word span (Gathercole et al., 1997; for a review see Baddeley et al. 1998). Because the investigation was conducted in the students' school, it was important to minimize interference with teaching in the institution. Therefore, we had to limit the selection of the test procedures to the one that produced the highest correlations in previous research. The reading span test is also a possible measure of working memory capacity, but both its reliability (Caplan & Waters, 1999; Waters & Caplan, 1996) and its validity as a measure of phonological short-tem memory capacity have been questioned (Towse & Hitch, 1995). Our decision to use an L1 based non-word repetition test was motivated by the fact that L2 non-word repetition ability is influenced by the knowledge of the phonological regularities of the L2 (Cheung, 1996; Masoura & Gathercole, 1999), therefore any correlation found between L2 non-word repetition ability and proficiency test scores might be the result of the fact that L2 non-word repetition ability itself is also an indirect measure of L2 competence.

Cambridge First Certificate Exam

The students' end-year exam was a Cambridge First Certificate exam, which was conducted by the teachers of the school. The written paper consists of reading and listening comprehension, composition, and Use of English test. The reading and listening comprehension sections contain three texts each, accompanied by multiple-choice items and

questions requiring short answers. In the composition task students had to write in three different genres, which were evaluated based on their content and accuracy. The oral exam consisted of an interview, a picture description task, and a problem-solving task in pairs. The Use of English test assessed students' level of vocabulary and grammar.

Results

An overview of the participants' general characteristics

As far as the verbal working memory capacity of the students is concerned, the results show that the students' non-word span was relatively high. The participants were able to repeat 6.94 syllables, which is higher than the national average of this age group as found by Racsmány and his co-workers (2005), whose participants aged between 12 and 16 scored 5.17 syllables. Our students' non-word span is even higher than that of the adult Hungarian national average (6.85 syllables – Racsmány et al., 2005). In our sample, 50% of students scored 7 syllables and above, whereas in the corresponding age group in the national sample only the highest 10 percentile of the population was able to repeat 7 syllables. The standard deviation in our population is similar to that of the same age group in Racsmány et al.'s study. Therefore, we can conclude that despite scoring higher, our sample is just as heterogeneous as the population tested in the national survey (see Table 1).

Table 1: The descriptive statistics of the verbal working memory test results

Variable	Non-word span	National average non-word span (12-16 years)	National average non-word span (adults)	Non-word average
Mean	6.94	5.17	6.85	2.55
SD	1.01	1.12	0.67	0.63
Min	4	3	5	1.14
Max	9	5	8	3.94

The distribution of the scores closely approximates the normal curve, although it is slightly negatively skewed. This suggests that we can regard the distribution of non-word span in our sample as normal. Among our participants, two students reached the maximum score of nine syllables, and six learners were able to repeat eight syllables (see Table 2). Most students (N=37) fell into the seven-syllable range. 54% of the participants scored in the average range, while 20% of the learners were below and 26% above the average in terms of their non-word span. One student could only repeat words consisting of four syllables, while eight students had a five-syllable working-memory span. This shows that there is enough variation in the non-word span of the participants to carry out meaningful analyses of the relationship of verbal working memory capacity

and second language proficiency. If we compare the students who were judged to be beginners in September and those who were identified as pre-intermediate learners, we find no significant differences in their verbal working memory scores (see Table 4). No significant difference in the non-word test scores of male and female students could be detected either (t = 0.79; p = 0.43)

Table 2: The distribution of the non-word span scores

Non-word span	Frequency	Percent
4	1	1.4
5	8	11.6
6	4	5.8
7	37	53.6
8	15	21.7
9	2	2.9

As regards the language test results, the students performed well in the proficiency exam: on average they scored 80.65% on the test. Their performance was best in the oral component, whereas their scores were the lowest in the Use of English test (see Table 3).

Table 3: The descriptive statistics of the proficiency test scores

Variable	Reading (max = 56)	Writing (max = 40)	Use of English (max= 75)	Listening (max = 30)	Speaking (max = 20)	Total score (max = 221)
Mean	48.26	32.74	54.62	22.26	17.00	174.15
SD	4.17	4.36	10.42	3.40	2.09	19.79
Min	37	20	28	13	12	125
Max	56	38	70	29	20	208

We compared the students who were considered to be beginners in September and those whose level of proficiency was already pre-intermediate at the beginning of the school-year in terms of their language test results and working memory scores by means of the Mann-Whitney U-test, which is the non-parametric counterpart of the t-test. Our selection of a non-parametric test was motivated by the fact that the number of students in the two groups was very different (21 learners in the pre-intermediate and 46 students in the beginner group). In order to reduce the probability that significant differences are found between groups due to the repeated use of the same statistical procedure, the method of Bonferroni correction was applied when establishing the significance levels. The results show that the pre-intermediate group achieved significantly

more points in every component of the exam except for the reading and writing paper (see Table 4).

Table 4: The comparison of beginner and pre-intermediate learners' proficiency and working memory test scores

Variable	df	Beginners (N=46) (SD)	Intermediate (N=21) (SD)	Z	p
Non-word span	64	6.82 (1.01)	7.19 (0.98)	1.44	0.15
Non-word average	64	89.47 (23.02)	97.38 (21.53)	1.52	0.12
Reading	64	47.65 (4.03)	49.50 (4.33)	1.97	0.05
Writing	64	31.73 (4.81)	34.75 (2.21)	2.25	0.02
Use of English	64	51.00 (10.03)	61.88 (6.86)	3.93	0.001
Listening	64	21.29 (3.46)	24.21 (2.32)	3.09	0.001
Speaking	64	16.48 (1.97)	18.04 (1.94)	3.01	0.003
Total proficiency score	64	168.14 (19.58)	188.37 (11.97)	3.88	0.001

The students who were beginners when the intensive language learning program started scored 20 points less out of the maximum number of possible points (221 points) than their more advanced peers. As the standard deviation figures suggest, the group of beginning learners was more heterogeneous in terms of proficiency than the pre-intermediate group. We have to note, however, that the difference between the two groups does not exceed 10% of the maximum number of possible points in any of the test components. Thus, even though the two groups of students differ statistically in their scores, this rate of difference does not seem to be considerable. Nevertheless, because the two groups followed different patterns of development: one from beginner to near upper-intermediate level, and the other from pre-intermediate to high upper-intermediate level, we treated the two groups as separate when analyzing the relationship of working memory and language proficiency test scores.

The relationship of working memory and proficiency test scores

If we perform correlational analyses to establish the relationship of working memory test scores and results in the various components of proficiency test in the case of the beginner and pre-intermediate group, the following findings emerge. As regards the students who were beginners at the start of the school-year, it can be seen that working memory span does not seem to be an important factor in explaining success at the end of the term test. Only the points awarded in the oral component of the exam are correlated with the average non-word score (see Table 5), and even this correlation is very weak. If we examine the relationship of non-word repetition scores and sub-scores in the oral component of the exam, we find that it is the points awarded for accuracy that correlate with average non-word repetition scores ($r = .34$, $p = 0.02$).

Table 5: The correlation of working memory and proficiency test scores in the case of the beginner group (N=46)

	Non-word span	Non-word average
Reading	0.05	0.04
Writing	0.05	0.10
Use of English	0.14	0.20
Listening	0.02	0.16
Speaking	0.19	0.29*
Total proficiency score	0.12	0.19

* $p < 0.05$

In the case of the learners who were at the pre-intermediate level in September, the average non-word score is moderately correlated with students' performance in the Writing and Use of English paper and with the total number of points students achieved in the test. Among the components of the oral test, average non-word score was found to be significantly correlated with the number of points students received for their fluency ($r = .46$, $p = 0.03$) and range of vocabulary ($r = .57$, $p = 0.01$).

Table 6: The correlation of working memory and proficiency test scores in the case of the pre-intermediate group (N=21)

	Non-word span	Non-word average
Reading	0.34	0.26
Writing	0.27	0.48*
Use of English	0.39	0.49*
Listening	-0.18	0.01
Speaking	0.20	0.34
Total proficiency score	0.43	0.47*

* $p < 0.05$

Discussion

In order to explain the different patterns of relationship between verbal working memory test scores and language proficiency, we have to refer back to what non-word repetition tests measure and what role verbal working memory capacity has been found to play in language learning. The non-word repetition test is claimed to be an indicator of the capacity of the phonological short-term memory system, which is responsible for the temporary storage and manipulation of phonological input (Baddeley, 2003). As shown in the review of literature, it has also been pointed out in a number of studies in the past ten years that due to the interaction of phonological knowledge stored in long-term memory and the functioning of the phonological short-term memory system, performance in this type of test is affected by the phonological characteristics of the non-words to be reproduced (Gathercole, 1995; Kovács & Racsmány, in preparation). Therefore, with the help of the Hungarian non-word repetition test we can gain an insight into how a particular learner is able to handle verbal input that resembles the phonotactic rules of his mother tongue. In other words, we can measure his verbal working memory capacity in L1. As our review of literature suggests, earlier studies in the field of L2 learning found a significant relationship between non-word repetition test scores in L1 and vocabulary learning in L2 (Masoura & Gathercole, 1999; Speciale, Ellis & Bywater, 2004) as well as the acquisition of L2 syntax (Ellis & Sinclair, 1996).

The results of our study indicate that in the case of the students who were beginners in September, verbal working memory capacity is related to only one constituent of language proficiency, namely speaking skills. If we consider how speaking differs from the other components of L2 competence, it can be concluded that this task requires on-line usage of one's knowledge of the L2, that is, in order to succeed in this component of the exam, students need to apply rules of the L2 automatically and should be able to retrieve words and memorized phrases within the time-constraints of the communicative situation. Because it is the accuracy score that seems to be related to non-word repetition test results, it is most likely that verbal memory capacity influences to what extent students can transform their declarative knowledge of grammatical rules of the L2 (i.e. knowing the rule) into procedural knowledge (i.e. being able to apply the rule) and use this knowledge automatically in the course of the one-year intensive program. Thus, this finding indirectly suggests that up to a certain stage of language proficiency, the automatization of syntactic and morphological rules seems to be affected by the ability to handle verbal material in working memory. This result is similar to the conclusion reached by Ellis and Sinclair (1996), who argued that working memory capacity influences the acquisition of L2 syntactic structures.

In the case of students who were already at a pre-intermediate level when the school-year started, we can see that significant correlations with non-word repetition scores emerge in the composition task and in the Use of English component, as well as in the case of the whole language proficiency test. Both in the case of the composition score and the achievement in the Use of English paper, knowledge of vocabulary and grammatical accuracy are the main components that are assessed. Whereas among the oral test components, significant correlations appear only in the case of the scores awarded for the range of vocabulary and fluency. Unfortunately, in the case of the Use

of English test most tasks assess both lexical and grammatical knowledge in an integrated manner, and in accordance with the rating guidelines of the exam, compositions were judged holistically. Therefore, we have no information on to what extent range of word knowledge and syntactic and morphological accuracy contributed to the final points in these two tasks. Nevertheless, we might speculate that at the level of competence the formerly pre-intermediate group achieved by the end of the school-year, which is high upper-intermediate if not advanced, most students display highly accurate and automatized grammatical processing. Therefore, what differentiates among them is most probably the knowledge of vocabulary and the ability to retrieve words quickly and efficiently.

This assumption is supported by the fact that unlike in the case of the beginner group, no significant correlations emerge between verbal working memory capacity and accuracy score in the oral exam. The effect of verbal working memory capacity on the on-line application of grammatical knowledge might have disappeared in this group because most formerly pre-intermediate students' grammatical encoding processes became automatized by the end of the year. The hypothesis that the source of differences between students in the more advanced group lies in the range of words and expressions they succeeded in acquiring during the school-year is supported by the strong correlation between scores awarded for the range of vocabulary and verbal memory test results. It seems that for higher-level learners the ability to handle verbal material in working memory plays an important role in the ability to acquire a wide repertoire of words and expressions. This finding is in accordance with earlier studies which suggested that achievement in L1 non-word repetition test is a good predictor of the success of L2 vocabulary acquisition (Masoura & Gathercole, 1999; Speciale, Ellis & Bywater, 2004) It is also quite likely that not only the encoding of vocabulary items in memory is affected by verbal working memory capacity, but the ability to access words easily and quickly, which is reflected in the correlation of the non-word repetition test and fluency scores. Fluency also presupposes the knowledge of a sufficient number of ready-made lexical units or formulae (Pawley & Syder, 1983), the acquisition of which constitutes an important part of language learning especially at higher proficiency levels (for a recent review see Wray, 2002). The ability to form larger units from smaller constituents, that is, to chunk has been supposed to be affected by working memory capacity (Zhang & Simon, 1985). Therefore it seems quite logical to suppose that the effect of verbal working memory capacity that manifests itself in the oral fluency score of the participants is also due to student variability in the ability to form linguistic chunks.

Finally, we might also presume that the relatively strong correlation between the total proficiency score and non-word repetition scores in the group of formerly pre-intermediate learners is caused by the relationship of verbal working memory capacity and vocabulary learning. As we argued above, it is reasonable to assume that what differentiates the two groups of participants in our study is mainly the repertoire of vocabulary and formulaic language, the acquisition of which has been found to be affected by verbal working memory. Therefore, we might conclude that above the intermediate level, verbal working memory seems to be an important predictor of L2 learning success by means of its influence on the acquisition of words and formulaic sequences.

Since in our study we used a Hungarian non-word repetition test, our findings also highlight the important role L1 verbal capacities might play in L2 acquisition. Sparks and Ganschow conducted a series of studies which indicated that L1 literacy skills serve as the foundation for L2 learning (e.g., Sparks & Ganschow, 1991, 1999, 2001). Dyslexic learners were also found to have limited verbal working memory capacity in addition to deficits in phonological/orthographic processing (Jeffries & Everatt, 2004). In their case, these impaired skills and capacities lead to increased difficulties in language learning (for a recent overview see Sarkadi, in press). Our study indicates that even small variations in verbal working memory capacity in a healthy group of learners might contribute to differential success in L2 learning.

Conclusion

In this paper we explored the role of verbal working memory capacity in the success of the acquisition of various language skills and competencies during a one-year-long intensive English language program. Our study addressed the question of what the relationship is between verbal working memory capacity and performance in an end of year reading, writing, listening, speaking and use of English test. The participants of the study were 67 secondary school students aged 15-16 in their year of intensive language learning of a dual language program. 46 of the students were beginners at the beginning of the school year, while 21 students started from a pre-intermediate level. The participants performed a non-word repetition test and took a Cambridge First Certificate Exam. Our study indicates that working memory capacity plays a different role in the case of beginners and pre-intermediate students in intensive language learning. Whereas in the case of beginners there was a significant correlation between non-word repetition scores and oral test results only, for the pre-intermediate students we found a highly significant relationship between the Use of English, the writing test and the overall proficiency test results.

From these results, we concluded that verbal working memory capacity influences the ability to apply linguistic rules automatically under time-constraints (i.e. during speaking) in the case of learners, whose L2 development took place from the beginner to the low upper-intermediate level. For the students, whose development proceeded from pre-intermediate to high upper-intermediate level, it was the success of the acquisition of words and formulaic sequences and the ability to retrieve and apply this knowledge efficiently that was affected by verbal working memory capacity. Because the non-word repetition test was conducted with Hungarian sounding non-words, the results of our study show that L1 based memory capacities play a role not only in the acquisition of L1 skills but also in several L2 learning processes.

Our study has several limitations, one of which is its small sample size. The other short-coming of our research is that relatively few students fell into the low working memory range, which might result in us not detecting all the significant correlations that might exist between verbal working memory capacity and L2 learning. The correlational design allows us only to draw conclusions concerning relationships, but it is impossible

to claim that verbal memory capacity is the cause of the differences in students' performance. Therefore, we plan to verify our claims made in this study by repeating the same study with similar students in the following year. By pooling the two sets of data and performing more complex statistical analyses, we can get a more complex insight into the issue of how verbal working memory capacity can affect the process of L2 acquisition.

References

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A. D. (2000).The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829-839.

Baddeley, A.D., Gathercole, S. E. , & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158-173.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation, Vol 8* (pp. 47–90). New York: Academic Press.

Caplan, D., & Waters, G. S. (1999) Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*, 77-126.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology, 3,* 867–873.

Costa, P. T., & McCrae, R. R. (1992). *NEO-PI-R. Professional manual.* Odessa, FL: Psychological Assessment Resources.

Dörnyei, Z. (2005) *The psychology of the language learner: Individual differences in second language acquisition.* Mahwah, NJ: Lawrence Erlbaum.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18,* 91–126.

Fortkamp, M. B. M. (1999). Working memory capacity and aspects of L2 speech production. *Communication and Cognition, 32,* 259-296.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation.* London: Edward Arnold.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition, 23,* 83–94.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3* (11), 410-419.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28,* 200-213.

Gathercole, S. E., & Thorn, A. S. C. (1998). Phonological short-term memory and foreign language learning. In A. F. Healy, & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, NJ: Lawrence Erlbaum.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33,* 966–979.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology, 13,* 65–77.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skills. *Studies in Second Language Acquisition, 14,* 25-38.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30,* 685–701.

Jeffries, S., & Everatt, J. (2004). Working memory: Its role in dyslexia and other specific learning difficulties. *Dyslexia, 10,* 196-214.

Kovács, G., & Racsmány, M. (in preparation). Handling L2 input in phonological STM: The effect of non-L1 phonetic segments and non-L1 phonotactics on nonword repetition.

Masoura, V. M., & Gathercole, S. E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology, 34 (5-6),* 383–388.

Miyake, A., & Friedman, D. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy, & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, NJ: Lawrence Erlbaum.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology, 48A (1)*, 98-107.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 317–331). London: Longman.

Racsmány, M., Lukács, Á., Németh, D., & Pléh, Cs. (2005). A verbális munkamemória magyar nyelvu vizsgálóeljárásai. *Pszichológiai Szemle, 60,* 479-506.

Sarkadi, Á. (in press). A diszlexiás nyelvtanulók angol tanításának elméleti és gyakorlati kérdései. *Iskolakultúra.*

Service, E. (1992). Phonology, working memory and foreign language learning. *Quarterly Journal of Experimental Psychology, 45A,* 21–50.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16,* 155-172.

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics 25*, 293–321.

Sparks, R. L., & Ganschow, E. (1991). Foreign language learning differences: Affective or native language aptitude differences? *Modern Language Journal, 75,* 3-16.

Towse, J. N., & Hitch, G. J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity? *Quarterly Journal of Experimental Psychology, 48A*, 108-124.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: critique of Just and Carpenter (1992). *Psychological Review, 4*, 761-772.

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.

Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and the acoustic loop hypothesis. *Memory and Cognition, 13*, 93-201.

Mapping the Mental Lexicon of Pre-Intermediate Learners:

Word Associations in a Depth of Word Knowledge Elicitation Task

Brigitta Dóczi

Department of English Applied Linguistics, Eötvös Loránd University, Budapest, Hungary

doczibri@eo.hu

Introduction

This paper reports on a research study intended to explore the organization of pre-intermediate EFL learners' mental lexicon through word associations in a depth of word knowledge elicitation task, as part of a larger research project aimed at observing the changes in the depth of students' word knowledge over a year. Fifteen participants' L1 and L2 word associations for 21 prompt words were examined. For data analysis on the word associations a mixed methods design was used: statistical procedures on presently available association categories (paradigmatic, syntagmatic, phonological 'clang' and analytic responses) were complemented with the analysis of language choice in the responses. The results illustrate that the students provided more paradigmatic associations in English for nouns than for verbs and adjectives, and they gave more translations for nouns and concrete words than for verbs, adjectives and abstract words. As regards the organization of the mental lexicon, three conclusions could be drawn from this set of data, confirming the findings of previous research studies. First, the better-known an L2 word is, the more central position it occupies in the learners' mental lexicon. Second, word pairs of concrete words and nouns tend to be stored together in memory. Third, nouns have a significant organizing role in pre-intermediate learners' bilingual lexicon.

Vocabulary is a widely researched area in second language research and a lot of attention has been given to the subject of vocabulary acquisition, especially to the size and development of lexicons, that is, breadth of word knowledge (Read, 2004). However, there is still a lack of research on depth of word knowledge, that is, how well single words are known; and even less is known about how particular words are acquired over a longer period of time. The problem is all the more complex since it can be supposed that individuals differ in the acquisition of words; furthermore, the notion of knowing a word is also problematic. The reason for this is that words are learnt incrementally, not "in a not acquired/acquired manner" (Schmitt, 1998a, p. 283), thus there are different levels of knowing a word, and meaning is just one element of this.

It is important to examine what it means to know a word through individual behaviour for two reasons. On the one hand, observing the incremental nature of

vocabulary acquisition could provide a deeper understanding of the processes of vocabulary learning. On the other hand, through examining depth of word knowledge we might gain an insight into how words are organized in the bilingual memory. Research into vocabulary acquisition has indicated that L1 and L2 words are stored in a common lexicon and recent studies have also aimed at investigating the connections between words through word associations. However, as most of the research to date has either compared native and non-native associations or used advanced L2 speakers, little is known about the organization of beginner or pre-intermediate learners' mental lexicons.

The present paper describes the first phase of a larger research project, which aims at exploring the depth of word knowledge through examining all the word knowledge types (spoken form, written form, grammatical and collocational behaviour, frequency, register, meaning and associations) with a special emphasis on word associations. Following the line of research started by Schmitt and Meara (1997), a further aim is to investigate the relationships between the different word knowledge components. To achieve this, the participants of the longitudinal study will be exposed to the same target words in English at least twice over the period of one year, and changes in the depth of their word knowledge are to be observed. As so far only phase one of the data collection has been completed, this articles analyzes the word associations the participants provided at the beginning of the school year 2005/2006 in order to investigate the types of connections that exist between L1 and L2 words in pre-intermediate learners' bilingual memory.

Naturally, the fact that in the present study the participants are at pre-intermediate level, which is much lower than in previous studies, necessitates the use of more frequent target words and implies that their associations cannot and should not be compared to those of native speakers. As a result of this, besides quantifying the responses given by the students, there is also a need for a qualitative analysis of the associations to see what new associative phenomena emerge at this stage of language proficiency. The aim of the current study is to investigate the associations of fifteen pre-intermediate learners by assigning their responses to existing association categories based on the literature as well as setting up new categories emerging in the process of data analysis, which are to be complemented with descriptive and inferential statistical procedures. A mixed methodology design was adopted for two reasons: the low level of the learners and the fact that they were allowed to associate for the prompt words with L1 and L2 words alike.

The first part of the article explores the theoretical and empirical findings with regard to the bilingual mental lexicon, which is followed by the methods of data collection and analysis. The discussion of the results centres on the comparison of the present findings with those of previous research in an attempt to distinguish the mental lexicon of pre-intermediate learners from that of advanced L2 speakers. Finally, further research possibilities and limitations are discussed in the conclusion.

Review of literature

Theoretical issues

Although the significance of vocabulary acquisition in a second language is unquestionable, relatively little is known about the actual processes that take place during the course of activation and actual production; and much attention as the organization of the lexicon in a second language has gained recently, there is still a lot of confusion concerning terminology as well as theory (Kormos, 2006). Therefore, in order to gain a deeper understanding of the issues in question, in the first part of the literature review theoretical issues are to be addressed, followed by the findings of empirical research.

The conceptual organization of the mental lexicon

As for the organization of concepts in the lexicon, several models have been put forward. The *hierarchical network model* (Collins & Quillian, 1969, 1970, 1972, as cited in Reeves, Hirsh-Pasek & Golinkoff, 1998, p. 197) claims that concepts are organized as "pyramids" with superordinate ones (e.g., plant) at the top, more specific ones in the middle (flower) and subordinate ones (e.g., rose) at the bottom; and each word is only linked to the closest concept. Some typical features of the model include cognitive economy (the more typical a semantic feature is, the higher the level it is stored) and category size effect (the larger the category, the longer the time the search takes). Another feature-oriented model is the *feature comparison model* (Smith, Shoben & Rips, 1974, as cited in Reeves, Hirsh-Pasek & Golinkoff, 1998, p. 199) which assumes that two different types of characteristics are stored: defining features (which are indispensable for a category to be included) and characteristic ones (which are typical but not necessary) For example, a defining feature is that all birds have feathers, while a typical one is that they fly (which is not true for all the birds).

In contrast to these two models, the *spreading activation model* (Collins, 1969, 1970 as cited in Reeves, Hirsh-Pasek & Golinkoff, 1998, p. 200) sees concepts as connected nodes with differing lengths of line between different concepts, based on the degree of their association. For example, the degree of association is higher between concepts such as *parrot* and *speaks* than between *parrot* and *skin*, which means that in the former case the connection between the concepts is stronger and shorter. Out of the three theories the *spreading activation model* is the most well-founded, for two reasons. Firstly, although it is a network of associations, its structure does not imply a rigid hierarchy, but allows for words to be related to several others. Secondly, there is no need to distinguish between defining and characteristic features; because the stronger the association, the closer the connection between the two concepts. Aitchison's *"cobweb" theory* (2003) is very similar to this: in her view, words are connected to each other in a way that one lexical item might be directly linked with several other lexical items on the basis of phonology, orthography, syntax or semantics.

With regard to the structure and organization of the mental lexicon, it is important to discuss what information is contained in it: concepts (the ideas to be expressed), word forms (lemmas) or word meaning (semantic knowledge) as well. On the one hand, psychological and neurological research suggests that a lemma contains only syntactic information and the word meaning is stored at the conceptual level (Levelt, Roelofs & Meyer, 1999). On the other hand, there is a strong disagreement regarding the definitions of conceptual and semantic knowledge, and although in L1 production the two are considered inseparable, as regards the L2 lexicon, there is still an ongoing debate whether the two can and should be differentiated at all (for an overview see Pavlenko, 1999).

The mental lexicon of bilinguals

Although the models used for the organization of the L1 lexicon have also been claimed to be valid for bilinguals, they have been slightly revised for lexical storage in L2. These theories maintain that L2 lexical items are not directly linked to corresponding concepts (as stated in the *hierarchical model* of Potter, So, von Eckhardt & Feldman, 1984) or if they are, these connections are weak (as stated in the *revised hierarchical model* of Kroll & Stewart, 1990, 1994). As a result, beginners in a second language can only gain access to L2 words through their L1 translations, but as their level of L2 proficiency increases, the links between the concepts are formed and there is no need for the L1 equivalent of the word any more. Contrary to this view, de Groot's *conceptual feature model* (1992) assumes that individual words are connected to concepts and the same word might be represented differently or similarly in L1 and L2. That is, some words (for example, concrete words such as *chair*) might have the same conceptual representation in the two languages, others may overlap partially (such as the conceptual representation of *winter* in English and Russian), while for certain words the representations in the two languages might differ completely.

One of the most hotly debated issues is whether L1 and L2 lexical items are stored in two different mental lexicons or in a common one where concepts are interrelated. Research on lexical encoding has verified that there is a shared bilingual lexicon, and the lexical items of both languages compete for word selection. Meara's studies (1982, 1984 as cited in Wolter, 2001, p. 42) indicated that connections in the L2 lexicon are less stable than those of native speakers, phonology plays a more significant role in the organization of the L2 lexicon and semantic links are systematically different from those of native speakers. Wolter's (2001) findings are more mixed: for well-known words native and non-native mental lexicons were found to be structurally different, whereas for less well-known words they were found to be similar. This is also supported by Wilks and Meara (2002), who claimed that there are higher numbers of connections at the core of the lexicon than at the periphery, and postulated that the network structure of L1 and L2 lexicon might differ because L1 lexical items are more connected than L2 ones. According to Wolter (2001), it is the depth of word knowledge that might play a pivotal role in establishing to what extent individual words are integrated into the

structure of the L1 and L2 lexicons and the basic technique for exploring the lexical network is through word association tasks (Read, 2004).

The type of information stored in the lexicon – depth of word knowledge

In order to learn a new word, it is important to clarify what kind of information needs to be acquired. The most complete and useful description of word knowledge is given by Nation (1990), consisting of eight word categories: spoken form, written form, grammatical behaviour, collocational behaviour, frequency, stylistic register constraints, meaning, and associations of a given word. These categories include both receptive and productive skills.

Earlier Anderson and Freebody (1981, as cited in Read, 2004, p. 211) defined depth of word knowledge as "the quality of understanding of a word", which means that a word can be considered known by a learner if all the distinctions that would be understood by an adult native speaker in normal conditions are clear to them. This definition refers only to precision of meaning and disregards the fact that several high-frequency words might have numerous meaning senses or are vague (especially when seen out of context, as it happens in many test situations). Furthermore, the term 'adult native speaker' is also problematic as it is hard to define who counts as an average native speaker (see Davis, 2003).

The third way of describing depth of word knowledge was suggested by Henriksen (1999) as *network knowledge*. In his view, the greater a learner's vocabulary size, the more there is a need for new words to be incorporated into an already existing network of words, which therefore needs to be restructured. Thus, depth of word knowledge is the ability to relate to semantically linked words.

In conclusion, it can be said that Nation's list of components of word knowledge is the most well-defined and comprehensive. However, for purposes of a deeper analysis it is important to look into network knowledge, as this enables us to shed light on the sophistication of the semantic network and not just individual words.

Word associations

Perhaps due to the relative ease of measuring network knowledge through quick and easy to administer word association tasks, this component of word knowledge has gained distinguished attention in second language research. A word association task is defined as one where speakers of a language are given a set of stimulus words one by one and they are instructed to give the first word that comes to their mind (Read, 2004).

By tradition, three categories of word associations have been identified: paradigmatic, syntagmatic, and phonological or 'clang' responses (Wolter, 2001). *Paradigmatic* responses have the same grammatical function as the prompt word and can be of four types: coordinates, superordinates, subordinates, and synonyms. *Syntagmatic* responses have a collocational or sequential relationship with the

prompt word, and are not from the same word class. *Phonological* or '*clang*' associations are semantically unrelated but similar-sounding words. Read (1993) added a fourth category: *analytic* responses, which could be a definition of characteristics, as if explained in a dictionary.

First language acquisition research has indicated that the older the children, the higher the proportion of paradigmatic responses. This has been referred to as the *syntagmatic-paradigmatic shift*, and can be explained by the cognitive and lexical development in the L1 mental lexicon. In addition, it was also found that unclassifiable and clang responses diminish with age (Ervin, 1961, as cited in Wolter, 2001, p. 43).

Empirical research

Research on depth of word knowledge

Researching depth of word knowledge is complex and time-consuming as it is extremely difficult to investigate all the components. Furthermore, there is still a lack of appropriate measures for assessing the various kinds of word knowledge and research needs to be longitudinal in order to show vocabulary development. However, even a longitudinal test design can at best aim to show declarative knowledge and the more implicit procedural knowledge will remain hidden.

It is not surprising then that few studies have ventured to investigate how related the various types of vocabulary knowledge are (Schmitt, 1998a; Schmitt & Meara, 1997). The findings of the studies that aimed to explore these interrelationships indicated that even though word association knowledge and grammatical knowledge correlated with each other, there was considerable individual variation and no significant development (Schmitt & Meara, 1997) and there was no evidence for developmental hierarchy between word knowledge types (Schmitt, 1998a).

The use of word associations in research

Although associations were originally used to assess the cognitive development of children in their mother tongue, in the last two decades they have been adopted by second language researchers to determine word knowledge in L2. By comparing the associations of native and non-native speakers, they have attempted to find out to what extent associations given by non-native speakers are similar to those of native speakers' (Piper & Leicester, 1980; as cited in Wolter, 2001, p. 44; Schmitt, 1998b).

Schmitt (1998b) claimed that there were several advantages to using word association tasks. On the one hand, it is a simple and quick procedure: after being given a stimulus word, respondents are asked to produce the first response that comes to their mind, while on the other, it yields much richer data about the respondent than any other traditional vocabulary test. Nevertheless, he also listed

several pitfalls, which had characterized previous research studies and aimed to overcome them by proposing a procedure for analyzing and quantifying the word associations of non-native speakers. One such drawback, according to Schmitt, is that in the norming process the differences between native speakers' responses cannot be fully taken into consideration. Also, the fact that most research studies had relied on one single response might distort the data gained by native- as well as by non-native speakers. Therefore, he recommended using multiple responses, which could then be weighted according to order and frequency. Moreover, he argued that there is a need to set a threshold from which associations can be considered native-like and went on to propose a four-level native-likeness scale. Wolter (2001) listed other problematic features of studies comparing native and non-native speakers' word associations: either they applied a limited number of fairly common prompt words whose responses can be easily predicted or, if lower frequency words were used, then the responses became varied and at places "child-like".

As far as categories of word associations are concerned, when native speakers were compared to advanced and beginner learners in Piper and Leicester's study (1980, as cited in Wolter, 2001, p. 44), it was found that for the verbs and adjectives native speakers produced the highest number of paradigmatic responses, followed by advanced EFL learners and the lowest proportion was produced by the beginner group. However, with respect to nouns, little difference was found between the groups supporting the idea that EFL learners show development in response type for nouns at an earlier stage, similarly to native speaker children.

One of the shortcomings of previous studies is that few investigations have been made into the organization of the bilingual lexicon and there is a mismatch between theory and research. An exception is van Hell and de Groot's study (1998), which found that Dutch bilinguals, who were asked to associate once in the language of the stimulus word and once in their mother tongue, used more translations for concrete words, cognates (similar words in L1 and L2) and nouns than for abstract words, noncognates and verbs, which might indicate that different word types might be stored in different places in the bilingual memory. This assumption has been verified by Wolter (2001), who, through investigating how depth of word knowledge affects the connections an L2 word has with other lexical items in the mental lexicon, found that the better known a particular word was, the more central the position it occupied in the lexicon. Furthermore, paradigmatic connections were formed in the centre, syntagmatic associations were typical further outside and phonological responses were found on the periphery, indicating that connection to other words becomes looser as the speaker's knowledge of the depth of a word decreases.

In summary, the results of the empirical studies on depth of word knowledge and word associations corroborate the existence of a bilingual lexicon, where various L1 and L2 lexical items are connected to each other to a varying degree and different word types might be stored at different places. Nevertheless, as most research has been concerned with native speakers or advanced L2 learners (except for Piper & Leicester, 1980), it would be interesting to gain more insight into the mental lexicon of less advanced learners of the language. As regards data analysis,

as there are still no appropriate measures for the analysis of responses (except to compare them with those of native speakers), perhaps involving a more qualitative approach, where new phenomena or categories might emerge, would prove to be useful.

Research questions and hypotheses

- How are words organized and connected in pre-intermediate Hungarian EFL students' memory?
- How do the two languages (English and Hungarian) interrelate?

My hypotheses based on the findings of previous research studies are the following:

- The ratio of syntagmatic responses will be higher than that of paradigmatic or analytic ones for verbs and adjectives, but not for nouns.
- There will be several instances of unclassifiable and clang responses.
- Phonology is going to play a significant role in the organization of L2 lexicon.
- More translations will be given for concrete words and nouns than for adjectives and verbs.

Method

Participants

The participants of the present research are a group of first year students (N=15), aged fifteen, in the non-bilingual section of a dual language secondary school in Budapest, which means that they take part in standard Hungarian secondary education from the age of fourteen to eighteen. At the beginning of the longitudinal research the participants were at pre-intermediate level, and they had five English lessons a week with the same teacher. The group consists of fifteen students.

Among the participants there are four boys and eleven girls. Five students attended primary school in Kispest, five in Pestszentlőrinc, three in other districts close to the secondary school, one in the first district and in France, and one student in the countryside (Martfű). The average number of years they spent studying English is seven, with the number of lessons increasing with age: the average number is three in the lower grades (1-4), and 4.5 in the higher ones (5-8). None of the students has ever spent considerable time in an English-speaking environment.

Instrument

Selection of the frequency list

Twenty-one target words have been carefully selected from the second thousand words of the Brown frequency list. The rationale behind opting for the Brown list is three-fold: on the one hand, it is the most recent one, and other lists have been criticized for using frequency studies from the first half of the twentieth century (e.g., West's list, 1953) or only representing the first one thousand words properly and lower frequency words are in need of reconsideration, according to topic and genre divisions (the General Service list. Furthermore, the Brown list takes all the different dialects of the English language into consideration.

Word selection

In order to prepare for the research design and select the suitable vocabulary items to be tested in September, 2005, it was indispensable to establish the expected level of the students earlier than that, even if the final list could only be drawn up shortly before the experiment. A difficulty that arose was that the students in question, just starting their secondary education, would only first enter the secondary school in September. Therefore, I tried to estimate the English knowledge of Hungarian primary school leavers by administering a vocabulary levels test to two groups of fourteen-year-old pupils (altogether 29), in two different settings in Budapest: a six-grade secondary school in the centre of Budapest (13 students), and a regular primary school (16 students). The test was an adopted pen-and-paper version assessing the knowledge of the first thousand words developed by Nation and Laufer (1999) (adapted for the internet by Cobb, available at http://www.lextutor.ca/levels/). The test is validated and measures controlled productive ability, and the task is for learners to decide if a sentence is true or not, through which they demonstrate their knowledge of a key word. The scoring of the test is relatively simple: the number of correct responses divided by the total number give a percentage figure, which, in order to demonstrate productive knowledge of the first thousand words, should be over 0.83. The analysis of the vocabulary levels test yielded the following results: the mean score was 0.85 for group 1, and 0.78 for group 2.

On the basis of the above-mentioned results, it could be concluded that the experiment would be most successful if words were selected from the second thousand words because even though the second group scored lower than 0.83, the difference is within the margin of error, and the students' knowledge of the first thousand words could be verified. Therefore, the rational decision was to use the next thousand words as this way the words are likely to be receptively known by the participants at the beginning of the year, but hopefully there is a lot of room for improvement towards productive knowledge during the year. However, applying less frequent words would have meant very limited or no initial knowledge on part of the participants resulting in few associative links.

Having limited the range of vocabulary to experiment with, the next step was to decide on the number of words to be assessed. Schmitt's (1998a) longitudinal depth of word knowledge research included eleven and my aim was to test more words, especially because learners in this case would be of lower level, and the interview would probably take less time. Therefore, the number was raised to approximately twenty, as this would also allow for some words to be excluded if necessary, and still enough data would remain. More than 20 would be too taxing for the students in one interview. It was also of great importance that words should be representative of nouns, verbs and adjectives. Previous research studies preferred verbs and nouns because they lend themselves to freer associations, so their higher ratio was kept; however, as more words are included in the present study, it was decided that some adjectives should be included as well. Finally, the most important criterion for selection was that words should have depth. Finally, twenty-one words were selected:

- 7 nouns: accident, advantage, advice, experience, interest, nature, reason;
- 10 verbs: allow, appear, concentrate, develop, decide, grow, laugh, mean, move, relax;
- 4 adjectives: able, foreign, successful, worth.

As can be seen, the nouns selected fulfil several depth of word knowledge categories: for example, *accident*, *experience*, *interest*, *nature* and *reason* all have more senses of meaning, their grammatical behaviour is not so simple either (countable vs. uncountable with *experience*, the use of the article with *nature*, the preposition that can follow *advantage*), they all have various word forms and collocations. The verbs were also chosen in a way that the knowledge of prepositions (*concentrate*, *laugh*), verb patterns (*allow, decide, develop, mean*), different meaning senses (*develop, grow, mean, move*), word forms and collocations could be demonstrated. Among the verbs there was one cognate (*concentrate-koncentrál*). The selection of the adjectives was different in that it appears more difficult to translate them from Hungarian and explain their grammatical behaviour (for example, *able* or *worth*). However, it is with these words that the depth of word knowledge development could be more spectacular, so they were also included. Furthermore, except for *successful*, their antonyms are not obvious either; therefore the associations could also yield interesting results.

The interview questions

Schmitt's (1998a) interview protocol was the starting point in the present study; however, due to the fact that he aimed to analyze only four aspects of word knowledge – spelling, associations, grammatical behaviour and meaning senses – other categories were to be incorporated as well. However, from the point of view of the present study, only question three is relevant: What five words come to your mind when you hear the word … (either in English or in Hungarian)?

The rationale behind choosing five as the number of responses to be elicited is that a single response would not suffice (especially because students could opt to provide a word either in English or in Hungarian), but too many responses might tire the students and "affect the reliability of the testing instrument" (Wolter, 2002, p. 319). Therefore five seemed like an ideal number and if students were unable to produce all five, it would also be a useful source of information.

In order not to influence students' responses with words given for the other categories, associations were invited right after the spoken form and the written form were consolidated. For eliciting word associations both English and Hungarian responses were allowed. The rationale behind this is partly the participants' low level of English, but partly the fact that the relationship and ratio of words in English and Hungarian were meant to be observed to provide an insight into learners' semantic network connections in the mental lexicon.

Procedure

The fifteen students were interviewed about the twenty-one target words in September, 2005. A strictly structured interview protocol (see Appendix A) was used, and four interlocutors conducted the interviews on two school days at the same time so that students would not share the words with each other. Each interview took approximately fifty minutes and was recorded. The questions were asked in Hungarian to make students feel at ease (and also because their level of English is not high enough yet). The training of the interlocutors took approximately an hour, and the following points were discussed: the positive atmosphere of the interview, practical questions (e.g., no help was to be given with the words and the wording and the order of the words and the questions should be the same; if students did not know the word, then the interviewer should tell them the correct form both in speaking as well as in writing) and technical questions (e.g., arrangement of the furniture, making note of the names of the students and putting them on the cassettes, folding the paper after each word).

Analysis of word associations

As shown in the review of literature section, all research carried out with associations applied a purely quantitative approach, relying on either statistical comparisons between native and non-native speakers or grouping responses into already existing categories (*paradigmatic*, *syntagmatic*, *phonological 'clang'*, and *analytic*). Thus, the present study aimed to complement the word counts (which language students opt for, how many associations they provide) using a qualitative approach in order to find other emerging patterns besides the existing word association categories. The use of two languages at the same time especially lends itself to a different kind of analysis.

According to Johnson and Onwuegbuzie (2004, p. 17), mixed methods research is the class of research where "quantitative and qualitative research techniques,

methods, approaches, concepts or language" are combined into a single study. In Creswell's view (1994) there are five main reasons for mixing methodologies within one study: to ensure triangulation, to allow different facets of a phenomenon to emerge, to strengthen one method by using the two sequentially, to allow new perspectives to come up, to expand one's study by giving scope and breadth to it.

Thus, besides using descriptive and inferential statistics, the associations were analyzed using the constant comparative method (Maykut & Morehouse, 1994). Categories – along with the existing ones – were established in each cycle of analysis (Creswell, 1998). Units of responses were counted for each student and for each word. Besides the four word association categories, the following new categories emerged in an attempt to analyze the mental lexicon of the learners:

- language use – participants' choice of the English and Hungarian languages for the associations;
- language change within responses – how the choice of the two languages changed for the associations of the same word;
- unclassifiable and unusual responses – this category attempts to describe these two kinds of associations and make a distinction between them.

Results and discussion

The elicitation process resulted in 1178 responses altogether, out of which 758 (64%) were in Hungarian and 420 (36%) were in English. The lowest number of responses elicited from a student for the 21 words is 35, while the highest is 146, almost four times as high, which might be due to the differences in the students' lexical knowledge.

Concerning typicality in the responses, three groups of words could be established. Firstly, the better known a particular word was in all the depth of word knowledge categories, the higher the number of identical associations in English (see Appendix B, Tables 1-3). For example, for the word *accident* six students associated with *hospital*, for the word *foreign* six students responded with the word *country*, while for *relax* five produced *sleep*. Secondly, in the case of lesser known words, students tended to associate similarly, but in Hungarian (e.g., *concentrate→ figyel [pay attention]* six instances, *able→ tud [know]* five instances, *advantage→ verseny [competition]* five instances), while for little-known words the number of similar associations was low in English as well as in Hungarian (e.g., *experience→ tapasztalt [experienced]*, two instances). Research comparing the associations of non-native speakers with those of native speakers (for more see Wolter, 2002) has also shown that the reason for this might be that associations tend to converge if a word is well-known by the students, whereas with lesser-known words, there is a divergence of associations. It could be the scope of another study to compare these associations with those of native speakers'.

Language use

Concerning the two languages, the ratio of the responses in English and the total number of responses does not seem to differ with regard to nouns, verbs and adjectives (the mean is 0.35 for nouns and verbs, and 0.38 for adjectives). Interestingly, more differences could be highlighted in the use of languages as regards paradigmatic and syntagmatic responses for the three word types.

With regard to language use, students can be divided into two distinct groups: ten students provided most of their associations (over 86%) in Hungarian, while the other five gave most responses (over 78%) in English. From their placement test, which aimed to decide whether students should join the beginner or the pre-intermediate group at the beginning of the year and consisted of a multiple choice grammar test and a short piece of writing marked by two teachers of the school, it could be concluded that the more advanced the students were, the more associations they could provide in English. These findings corroborate the conclusions of Wolter (2001) as well as Wilks and Meara (2002), who have also found that the better known a certain word is, the more central position it occupies in the mental lexicon and the more connections it has with other words. Also, as they found, the connections between L2 items become stronger as the words become better known by the learner. Through the typical responses, the 15 pre-intermediate learners of the present study demonstrated that if a word is well-known, it has stronger links with other L2 words.

Language change, that is, giving a response in one language then turning to the other and continuing giving responses in that language, was often triggered by translations. 46 instances were found (concerning 92 words, 8% of the total number of words) where it was the translation that lead to a change in language use. The change was two-directional: sometimes it went from English to Hungarian or the other way round. Out of the 46 cases, 24 were translations of nouns, 15 of verbs and 7 of adjectives. The words translated were concrete words on 39 occasions (85% of the cases). For example:

advantage → <u>üzlet, business</u>, project, businessman, travelling;
accident → <u>hospital, kórház</u>, mentőkocsi [ambulance], seb [wound], orvos [doctor], sebészet [surgery];
foreign → tourist, museum, <u>trip, túra</u>, látványosság [sight];
worth → olcsó, drága, cheap, expensive.
These word pairs mentioned by the students are consistent with the findings of van Hell and de Groot (1998), who, after showing that more translations were given for concrete words in an association task, suggested that concrete words might be stored together in the bilingual memory due to a shared conceptual representation. Despite the fact that the participants of the present study were of lower level than those in van Hell and de Groot's, it was observed that high-frequency concrete words could indeed be organized together in the bilingual mental lexicon.

Word association categories

As can be seen in Table 1, the mean number of words given in the two languages together for nouns, verbs and adjectives is nearly the same (approximately 56 associations per prompt word) and the ratio of words associated in English (responses in English) and in the two languages (all responses) does not show any significant differences either (approximately 21 associations per prompt word).

Table 1: Mean number of responses according to word types and word association categories

word class	all responses	responses in English	all paradigmatic responses	paradigmatic responses in English	all syntagmatic responses	syntagmatic responses in English
Noun	56.71	21	38.57	14.57	12	3
Verb	56.3	20.1	15	4.6	37.2	12.6
Adjective	55.25	21.5	12.75	3	36.75	15.25
Total	56.24	20.66	22.43	7.62	28.71	9.9

Paradigmatic responses

Interestingly, Table 1 also shows that the mean number of total paradigmatic responses (associations having the same grammatical function as the prompt word) for nouns is 38.57, whereas for verbs it is 15 and for adjectives it is 12.75. This indicates that paradigmatic responses were almost three times as frequent for nouns as for verbs and adjectives. As for the use of English, significant differences could also be established with regard to the three word types, thus participants used fewer paradigmatic responses for verbs and adjectives than they did for nouns. The mean number of paradigmatic responses is 14.57 for nouns, 4.6 for verbs and 3 for adjectives. Since sample size was small, and data were not equally distributed (as the words varied in difficulty for the students), Friedman's ANOVA was used and significant differences could be found between the three conditions (χ^2=10.889 p=0.001). The mean difference between nouns and verbs is 9.97. The independent t-test showed that such a result can be considered significant (t (df=6.4) =3.2313, p=0.05), and there is also a significant mean difference (11.57) between nouns and adjectives (t (df=6.8) =2.12, p=0.029). On the basis of these, it can be concluded that paradigmatic responses in English were used more often by the students for nouns than for adjectives and verbs. This is in line with previous research, showing that the paradigmatic shift occurs with native and non-native speakers as well, and it comes earlier for nouns than for other word classes (Piper & Leicester, as cited in Wolter, 2001, p. 44). The rationale behind this might be that L2 learners acquire more nouns at the beginning of the language learning process, which might lead to a more stable organization of nouns at the earlier stages of proficiency.

As for the types of paradigmatic responses, for well-known words students tended to produce coordinates in English (e.g., *accident → doctor, nurse, headache,*

toothache), while for lesser-known words several synonyms were provided in Hungarian (e.g., *laugh → kacag, kuncog, röhög*). This could indicate that coordinates are organized earlier in the L2 memory than synonyms but further research needs to be carried out to verify this.

The ratio of paradigmatic responses in English and the total number of paradigmatic responses is lower for adjectives (mean=0.21) than it is for nouns (mean=0.33) and verbs (mean=0.35). Due to low number of words and the fact that equal variances could not be assumed (as the words varied in difficulty for the students), the Kruskal-Wallis test was administered, but no significance between the three conditions were found (mean=0.334 p=0.5).

Syntagmatic responses

The mean number of syntagmatic responses (associations that have a collocational or sequential relationship with the prompt word) for nouns is 12, whereas for verbs and adjectives it is 37.2 and 36.75 (see Table 1). Regarding the use of English, the ratio of syntagmatic responses in English and the total number of syntagmatic responses is higher for adjectives (mean=0.4) than it is for verbs (mean=0.33), and it is the lowest for nouns (mean=0.27). Again, owing to the low number of words and the fact that equal variances could not be assumed (as the words varied in difficulty for the students), the Kruskal-Wallis statistical test was administered, but there was no significant difference (mean=0.489, p=0.2) between the three conditions. What transpires from this set of data is that learners tended to associate with nouns for verbs and adjectives as well.

The above-mentioned findings do not support those of Wolter's (2001), which showed a preference for syntagmatic responses in the case of non-native speakers. Wolter argued for a "syntagmatically dominated" (p. 61) L2 mental lexicon and called for a re-evaluation of the syntagmatic-paradigmatic shift in the case of non-native speakers. By contrast, the results of the present study do not indicate a syntagmatically dominant mental lexicon, especially because the mean number of paradigmatic responses (7.6) and syntagmatic responses (9.9) in English does not differ significantly (p=0.43). The reason for this might be the difference between the proficiency level of the participants of the current study and those of Wolter's (who could only participate in the research if they scored higher than 500 on the TOEFL test) as L2 learners must possess a deeper knowledge of the language to be able to rely on collocational and sequential associations.

Phonological 'clang' responses

Surprisingly, there were very few clang responses provided by the students, 64 altogether, but due to the fact that 58 of these were provided by one student (her performance would be an interesting case study), this number can be considered insignificant. This finding seems to contradict some previous research studies (Meara, 1982, 1984 as cited in Wolter, 2001, p. 42), which emphasized the impor-

tance phonology plays in the organization of the L2 lexicon. Nevertheless, the few instances found in the present study are consistent with the findings of Wolter (2001), according to which non-native speakers did not produce more phonological responses than natives. In his model of the mental lexicon he claimed that phonological links are located on the periphery, indicating that a word is not known well enough to have paradigmatic or syntagmatic connections with other words. Likewise, in the present study the few cases when clang responses were given, the word was always little-known by the student. However, the fact that in the present study students were also able to produce words in Hungarian might indicate that they could use this as an avoidance strategy once they could no longer think of a word in English, and rather than give a clang response, they preferred a word closer to the meaning in Hungarian. Thus, the lack of phonological responses is by no means a proof that pre-intermediate students would not use them more frequently if the use of L1 was not encouraged.

Analytic responses

There were only twelve instances of analytic responses, but except for one, all are in Hungarian, which might be attributed to students' low level of English. An interesting finding is that analytic responses were never one word, but rather an explanatory phrase, which here was regarded as one unit. For example, for the verb *move*, one analytic unit of response was '*it needs a lot of time*' or for *interest*, '*valami fontos dolog*' *[something important]* and '*nem hétköznapi*' *[not ordinary]* were given. The recordings indicated that analytic responses usually appeared when the students found it difficult to come up with a new association either in English or in Hungarian, and they solved this problem by explaining the word in their own words.

Unclassifiable responses vs. unusual responses

The most problematic category seems to have been the unclassifiable words because in previous research studies no mention was made about the distinction between unclassifiable and unusual. Although the participants of the present study did provide a few unusual responses, if the connection with the prompt word could be established, the response was accepted and grouped into one of the categories (e.g., *appear→ varázsló [magician], laugh → száj [mouth], fogak [teeth])*. Unclassifiable responses were the ones where the connection could not be understood by an outsider: either the response was vague or personal (e.g., *foreign → London, relax → medve [bear]*), or the respondent misunderstood the meaning of the original word and associated to a different one (*worth → megéri az öregkort [reach old age], idős [old]; mean → jelent a hetes [something like: report to the teacher]*). Altogether there were 32 instances of unclassifiable responses but this low number might again be attributed to the fact that Hungarian associations were also allowed.

Conclusion

The aim of the present study was to gain a better understanding of how beginner L2 learners' mental lexicon might be organized. Hypothesis 1 and 4 have been confirmed: students gave more paradigmatic associations in English for nouns than for verbs and adjectives, and they gave more translations for nouns and concrete words than for verbs, adjectives and abstract words. Although hypothesis 2 and 3 have not been supported in this study, namely the students did not provide many clang and unclassifiable responses, this is likely to have occurred because of the possibility of using Hungarian words besides English. However, allowing the students to provide associations in both languages also made it possible to note some important distinctions between analytic, unclassifiable and unusual responses so far disregarded in the literature. Concerning the organization of the mental lexicon, the following conclusions could be drawn from this set of data, confirming the findings of previous research studies:

- Better-known L2 words occupy a central position in the learners' memory, having more established links with other L2 words, and the associations of well-known words also tend to converge more.
- As these pre-intermediate learners gave more translations of concrete words and nouns, it might be assumed that these word pairs are the first to be stored together in the mental lexicon.
- The significantly higher ratio of paradigmatic associations for nouns than for verbs and adjectives indicates that nouns have a pivotal role in the organization of less advanced learners' mental lexicon.

The limitations of the present study are manifold. First of all, a deeper analysis can only be made at the end of the academic year, once all the data for all the depth of word knowledge categories have been collected and changes in the knowledge of words have been observed. Secondly, in the present analysis there was only one rater grouping the associations. Thirdly, the fact that learners were allowed to give their responses in English as well as in Hungarian – besides being a source of help for them and a source of information for the researcher – might have distorted some of the findings because they could avoid producing a word only in English. Contrary to all the limitations, the findings of this study imply several avenues for further research. Firstly, in order to discover a more precise nature and structure of the mental lexicon, measures for the analysis of association responses are still in need of refinement. Secondly, word associations need to be investigated in the light of depth of individual word knowledge and how this knowledge changes with increasing proficiency over longer periods of time.

References

Aitchison, J. (2003). *Words in the mind: An introduction to the mental lexicon*. Malden, Basil Blackwell.

Brown frequency list. Available at http://www.edict.com.hk/lexiconindex/frequencylists/words2000.htm).

Creswell, J. W. (1994). *Research design – Qualitative and quantitative approaches*. Thousand Oaks: Sage Publications.

Creswell, J. W. (1998). *Qualitative inquiry and research design*. Thousand Oaks: Sage Publications.

Davis, M. H. (2003). Connectionist modelling of lexical segmentation and vocabulary acquisition. In P. Quinlan (Ed.), *Connectionist models of development: Developmental processes in real and artificial neural networks*. Brighton: Psychology Press. Available at http://www.mrccbu.cam.ac.uk/~mattd/pubs/davis.quinlan_chapter.pdf

De Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 1001-1018.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*, 303-317.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed method research: A research paradigm whose time has come. *Educational Researcher, 33* (7), 14-26.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah: Lawrence Erlbaum.

Kroll, J. F., & Stewart, E. (1990, November). Concept mediation in bilingual translation. Paper presented at the 31st annual meeting of the Psychonomic Society. New Orleans, LA.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory & Language, 33*, 149-174.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Science, 22*, 1-38.

Maykut, P., & Morehouse, R. (1994). *Beginning qualitative research: A philosophic and practical guide*. London: The Falmer Press.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

Nation, P., & Laufer, B. (1999). A vocabulary size test of controlled productive ability. *Language Testing, 16* (1), 33-51.

Pavlenko, A. (1999). New approaches to concepts in bilingual memory. Bilingualism: *Language and Cognition, 2, 209-230.*

Potter, M. C., So, K. F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning and Verbal Behaviour, 23*, 23-38.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards, & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 209-227). Amsterdam: John Benjamins.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*, 355-371.

Reeves L. M., Hirsh-Pasek, K., & Golinkoff, R. (1998). Words and meaning: from primitives to complex organization. In J. Gleason., & N. Berstein-Ratner (Eds.), *Psycholinguistics* (2nd ed., pp. 157-226). Harcourt College Publishers.

Schmitt, N. (1998a). Tracking the incremental acquisition of second language vocabulary: a longitudinal study. *Language Learning, 48* (2), 281-317.

Schmitt, N. (1998b). Quantifying word association responses: What is native-like? *System, 26*, 389-401.

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition, 20*, 17-36.

Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Language and Cognition, 1* (3), 193-211.

Wilks, C., & Meara, P. (2002). Untangling word webs: Graph theory and the notion of density in second language word association networks. *Second Language Research, 18*, 303-324.

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition, 23*, 41-69.

Wolter, B. (2002). Assessing proficiency through word associations: Is there still hope? *System, 30*, 315-329.

Appendix A: Interview protocol

The following questions were asked about each word:

Do you know how to say the word *baleset* in English? (accident) – *to elicit the spoken form + translation*
Can you spell it? – *to elicit the written form*
What words come to your mind when you hear this word? (either in English or in Hungarian) – *to elicit associations*
Do you know what part of speech this word is? – *to elicit grammatical behaviour*
Do you know any other forms of this word? – *to elicit other forms of the same word*
Can you think of any other meanings of this word? – *to elicit other meaning senses*
Could you say a sentence with this word? – *to elicit productive knowledge*
Can you think of any collocations of this word? – *to elicit collocational behaviour*

Appendix B: Word associations according to word types and word association categories

Table 1: Nouns

prompt word	most typical response	number of responses	responses		paradigmatic responses		syntagmatic responses		clang responses		analytic responses*	unclass ifiable*
			E	H	E	H	E	H	E	H		
accident	hospital 6	72	34	38	31	35	2	2	0	1	0	1
nature	növény(ek) 6 flower 4	72	30	42	24	35	1	6	2	1	0	3
advantage	verseny 5	53	20	33	15	22	3	5	4	0	3	1
interest	érdekes 3	51	22	29	17	14	3	11	2	0	3	1
reason	okozat 3	41	8	33	3	23	2	8	1	1	1	3
advice	barát 4 friend 2	55	20	35	11	21	3	12	4	2	0	2
experience	tapasztalt 2	53	13	40	1	18	7	19	3	1	1	2
total		397	147	250	102	168	21	63	16	6	8	13

* There were so few instances of analytic and unclassifiable responses that there seemed no point in separating the two languages

Table 2: Verbs

prompt word	most typical response	number of responses	responses E	responses H	paradigmatic responses E	paradigmatic responses H	syntagmatic responses E	syntagmatic responses H	clang responses E	clang responses H	analytic responses*	unclass ifiable*
develop	grow 5	49	24	25	7	9	16	16	1	0	0	0
decide	elhatároz 4** question 3	56	18	38	4	14	9	22	3	0	1	3
grow	develop 5***	67	28	39	9	8	14	31	5	0	0	0
move	house 4 ház 3	68	26	42	2	7	21	33	3	2	0	0
relax	sleep 5	56	22	34	6	6	16	27	0	0	0	1
appear	eltűnik 3 disappear 3	49	17	32	5	15	8	16	4	0	0	1
concentrate	figyel 6	55	16	39	5	18	8	21	3	0	0	0
laugh	happy 4 boldogság 4	68	24	44	4	12	18	32	2	0	0	0
mean	szó 4 word 2	43	10	33	1	6	6	26	3	0	0	1
allow	szabad 3 parent 2	52	16	36	3	9	10	22	3	0	1	4
total		563	201	362	43	114	126	246	27	2	2	10

** The Hungarian word given at the beginning was 'eldönt'

*** The reason why students associated with *develop* for the word *grow* might be that *develop* was the first word in the interview. Perhaps it was not the best choice to include these two words from the point of view of associations but their other word knowledge categories (e.g. word form, other meaning, collocations) justify this choice.

Table 3: Adjectives

prompt word	most typical response	number of responses	responses E	responses H	paradigmatic responses E	paradigmatic responses H	syntagmatic responses E	syntagmatic responses H	clang responses E	clang responses H	analytic responses*	unclass ifiable*
foreign	country 6	64	30	34	3	6	26	27	1	1	0	0
able	tud 5 can 3	47	13	34	0	7	12	23	1	2	1	1
Successful	siker 2 sztár 2 actor 2	57	26	31	5	16	17	15	3	0	0	1
worth	expensive 2	53	17	36	4	10	6	21	4	0	1	7
total		221	86	135	12	39	61	86	9	3	2	9

Vocabulary as a Filter with First-Year English Majors

Magdolna Lehmann

University of Pécs, Department of English Applied Linguistics, Pécs, Hungary

lehmannm@btk.pte.hu

The beginnings

Lexis has been a major focus of research in applied linguistics in the last few decades and an abundance of materials has been developed to measure either the size or the depth of learner vocabulary, constituting the two main directions of approaching the field. It emerges from current second/foreign language (L2) vocabulary studies that language is more lexical than grammatical and that vocabulary is a good predictor of success on reading comprehension tests, and thus indirectly and in a wider context, in academic studies. This article discusses the major issues, results and the future of introducing vocabulary testing as a separate subtest of the long-existing proficiency test developed and administered at the Department of English Applied Linguistics, University of Pécs. In the first part of the paper when I introduce the early versions of the proficiency test I rely on the works of my colleagues, as I joined the testing team when it had been operating for eight years. The second part, after introducing the essential background to vocabulary research, presents the results and lessons of two trials of an acknowledged and widely used vocabulary test. Finally, conclusions are drawn and some possible future directions of vocabulary testing at the department are outlined.

The idea of testing the language production of first-year students at the English Department was first raised in 1993. Various aspects of language testing and curricular requirements had been carefully considered before deciding on the aims and specifications of the initial test battery which went through numerous modifications until it reached its present format in 2002. The rationale for testing language proficiency, explains Szabó (1996), who has been a member of the testing team from the beginnings, was on the one hand the top priority of language development in the first two years at the Department. On the other hand, the fact that the first overall measurement of language performance took place only in the fourth semester of English studies at the first complex exam ("szigorlat") necessitated getting information and giving feedback on the efforts and improvement of the students within this period and filtering students before this consequential examination.

The filter/proficiency test

The aim of the original test, which has not changed till the present day, was to select those students whose language level was below a carefully determined level, therefore the test devised was a filter test. As it aimed to measure overall language achievement, not performance based on or compared to any syllabus, it was a criterion referenced proficiency test at the same time, filtering those who were not "proficient enough" to pursue further studies (Szabó, 1996). In its original form the battery intended to test all four skills (reading, writing, listening and speaking) with an additional grammar component, however, in later stages of its development the idea of testing speaking was dropped due to its time consuming nature, the high number of candidates and the limited number of interlocutors resulting in unbridgeable reliability issues, as well as the fact that students' speaking skills were developed in the language practice courses where the instructors had a whole semester to continuously evaluate and incorporate speaking performance into the course grade, making the assessment of this skill more reliable (Szabó, 1996). Therefore, the speaking subtest was administered only once as a trial in the form of peer-to-peer interaction in an information gap task, further on it was omitted from all later versions of the proficiency test.

Testing reading

In the construction of the reading subtest special attention is to be paid to the selection of the texts. Szabó (1996) refers to the richness of literature on the decisive impact of the text on test results and reliability and explains that among test parameters authenticity, familiarity and subject interest were to be primarily borne in mind. Testing techniques have shown a great variety in the history of the test, including multiple choice items for objectivity, ease and speed of administration and assessment, as well as suitability for computerized item analysis; information transfer; short answer; and sequencing (for problem areas see Szabó, 1996) to measure both micro- and macro skills of reading.

Since I joined the team in 2000, two kinds of multiple matching tasks have been used, where "two sets of stimuli have to be matched against each other" (Alderson, 2000, p. 215), part one (with eight items) being an article where eight paragraphs have to be inserted to fit into the numbered gaps to measure macro skills of text comprehension. Similarly, in part two testees have the same task with missing parts of sentences to test micro skills of reading. It is important to note here the significance of providing more alternatives than the number of items required by the matching task to avoid the possibility that there is only one possible final choice at the end (Alderson, 2000).

Testing listening

Although the majority of the above techniques can be utilized in the design of the listening comprehension subtest, certain issues are more complex and thus require special attention on the part of the item writer. Apart from general text selection criteria Szabó (1996) emphasizes the somewhat different interpretation of authenticity in the case of a listening text, varying considerably in rate of delivery, accent, and clarity of recording. It was agreed that both American and British English native speakers were to be chosen, who speak "standard" English with "average" speed of delivery. Items involved multiple choice and multiple matching with careful spacing of information to avoid testing the memory of the candidates.

Testing writing

Writing is tested directly for the sake of validity and reliability. In the early versions of the subtest candidates were presented with a choice of topics to be selected from a three-column grid, where each column contained a part of the possible essay title to be chosen (for details see Horváth, 1996). More recent versions have remained student-friendly in this respect; however, the choice of topics has been restricted to two with four guiding points each, specifying what aspects of the topic to elaborate on in each paragraph of the 300-word essay. Multiple scoring based on analytic scales has been common practice right from the beginning for enhanced objectivity.

Grammar as a separate test component

Although grammar is not a skill, all four skills make use of it indirectly, therefore "it is legitimate to construct direct tests of grammar separately" (Szabó, 1996, p. 82). The basic principle of item writing with the grammar subtest was discrete point, rapid mass testing, he explains, with techniques and content selected on the basis of practicality, objectivity and suitability for statistical item analysis. The techniques and item types will be discussed in more detail later.

Assessment and data analysis

The early versions of the Proficiency test used to be pass/fail tests; however, to meet new curricular requirements in 2003, evaluation had to conform to a 1-5 scale of marking. By constant and detailed item analysis the battery has continuously been improved, employing the latest available approaches to language testing and statistics to become as objective as possible. In the analysis the (Szabó, 1996). Every semester two test dates are offered, thus two tests need to be devised. Thorough item analysis procedures, relying on the Rasch-model within the Item Response Theory of practical language testing, and the use of anchor items from previous

tests ensure maintaining the same level of difficulty and making the comparison of test results possible from time to time (for details see Szabó's paper in this volume). Pass marks have long been set on the basis of score distribution curves and the original idea of item banking has been developed to its pedagogical potential serving as a basis for recent test construction.

The present

Mention has been made of grammar not being one of the four skills, but it is a component of linguistic competence in the widely accepted communicative competence model (e. g., Hymes, 1971; Swain, 1985; Bachman, 1990). So is vocabulary. A sufficient knowledge of both is a prerequisite of successful text comprehension, written or aural, as well as using language productively in an essay writing task. A wealth of research support the prevalence of word knowledge over knowledge of syntax in text comprehension (Widdowson, 1978 cited in Boyd Zimmerman, 1997, p. 4; Laufer, 1997; Nation, 1993), while other studies (Morris & Cobb, 2003) describe vocabulary as a good predictor of academic success determining how students are able to cope with the reading load at university.

Background to vocabulary research

The main body of vocabulary research focuses on the question of how many words a language learner needs to know to be able to perform certain tasks at certain levels. The main difference between a native speaker and a language learner in this respect proves to be the gap in their vocabulary size, i.e. how many words they know. Goulden, Nation and Read (1990, p. 341) estimated that "well-educated adult native speakers of English have a receptive vocabulary of around 17,000 base words", excluding proper nouns, compound words and abbreviations. Based on similar studies by D'Anna, Zechmeister and Hall (1991) and Nusbaum, Pisoni and Davis (1984), Zechmeister, D'Anna, Hall, Paus and Smith (1993) concluded that the vocabulary size of an average native speaker university undergraduate is in the range of 14,000 – 17,000 words, thus the receptive knowledge of 14,000 words should be sufficient to pursue academic studies.

However, Hazenberg and Hulstijn (1996) warn that fourteen thousand words may not be a minimal or optimal figure for a non-native speaker, referring to studies on text coverage that claim for various languages, that the first most common 5,000 words in a language cover 90-95 percent of all text tokens in an average text. On the other hand, Hirsh and Nation (1992) assume that readers need to be familiar with 95 percent of the words in a text to understand the main points of unsimplified texts (i.e. to reach some degree of text comprehension). Thus, it follows that a vocabulary of 5,000 words would sufficiently equip a non-native speaker student for university studies, although Laufer (1987, 1992a, 1992b) argues that 5,000 is really a "bottom line" for reading English at an academic level. There is no consensus in research on the optimal figure, but some studies estimate it to

be much higher. With Dutch students Hazenberg and Hulstijn (1996) concluded that a minimal receptive vocabulary of 10,000 base words is necessary for successful university studies in Dutch. Although no two languages are the same, the similarities of context, and English and Dutch both being agglutinative languages may make their results worth considering for English as a foreign language.

Vocabulary as a separate test component

Based on reports by both students and staff of our department, as well as my own classroom experience I assumed that first-year students find it difficult to understand the academic reading materials required for their courses. I hypothesized that the lack of sufficient vocabulary for academic studies hindered reading comprehension. In a preliminary study, therefore, I attempted to define the size of the receptive/passive EFL vocabulary of 93 participating students (Lehmann, forthcoming) and the result of 7,000 words on average seemed to support the case. On the basis of these findings and the above discussed predictive value of vocabulary knowledge on later academic success, the testing team responsible for writing items for the Filter test decided to insert a separate vocabulary section into the original test battery in 2002.

The general guiding principle remained the same as the one applying to the rest of the test, i.e. to filter out those candidates who were below a certain level, in this case lacking in vocabulary knowledge. This, together with other criteria of practicality, objectivity, suitability for computer item analysis and ease of both scoring and administration lead us to the decision of using a discrete point, receptive vocabulary size test first developed by Nation (1990), later revised and validated by Schmitt (2000), the Vocabulary Levels Test (VLT). This widely used test of vocabulary breadth provides an estimate of how many words a candidate knows on five levels of English word knowledge: the first 2,000, 3,000, 5,000 and 10,000 most frequent words on the basis of large corpora of texts. The fifth level, however, is not based on frequency data, it comprises a representative sample of words from the Academic Word List (Coxhead, 1998), i.e. words that are most common in academic texts from a wide ranges of fields and subjects. Each level consists of ten clusters of six words that need to be matched with three synonyms or short definitions in order to reduce the need for reading and knowledge of syntax, and at the same time to minimize the chances of guessing correctly. Below is an example of a noun cluster from the academic word level of the VLT (with key):

1 area
2 contract   __2___ written agreement
3 definition   __5___ way of doing something
4 evidence   __4___ reason for believing
5 method   something is or is not true
6 role

Considering the place of such a long, 150-item test in the construct of our original test battery restricted by limitations of time allotment, scoring and item analysis procedures, the testing team agreed on using the academic level clusters of VLT only. It was decided to be included in the grammar and usage component of the Filter test as a third section, to follow a four-option, 30-item multiple choice test of grammar, and a 10-item multiple choice section of three-option corpus-based sentences, where only one of the options fill in all three sentences correctly. Thus, with the 30 items of the vocabulary test the possible maximum score on the grammar and usage section is seventy. To avoid over-emphasizing the importance of the vocabulary test, no pass marks are set for the separate sections, the three sections are treated as one unit.

Based on facility values and Rasch item difficulty figures Schmitt (2000) found that the words in the academic level of VLT fit between the 2,000 and the 10,000 word levels, thus it covers a broad range of word knowledge. He also recommends the use of the levels test not as a fixed battery, but as flexible on the basis of the specific demands in each testing situation, allowing for picking a certain level only to be administered, or mingling the two versions for a given level to create a longer test.

The academic word level test

Considering all these issues, we decided to apply the academic level only, as the knowledge of these words is essential in understanding authentic academic texts. The format of the clusters was tailored to suit the ITEMAN software better, i.e. the clusters were numbered 1-10 and the items within the clusters were assigned capital letters A-F. The first tests involving the vocabulary component were then administered in the Spring semester of 2002, 154 students taking Version 1, and 66 students taking Version 2 of the VLT academic level test. Data were analyzed with ITEMAN (TM) Version 3.50; the results shown below first represent the overall results gained from the grammar and usage component supplemented with the academic word level test of the VLT. It is important to note here that the examinees of the two test versions were not the same students.

The mean on the grammar and usage component was 46.39 and 41.14 on the two versions (st.dev 7.51 and 5.4, respectively; see Table 1). A difference in variance can be observed between the two versions indicating a wider scale of scores on version 1. Cronbach's alpha values indicate a high reliability of both instruments. Taking a closer look at the vocabulary sub-tests (Part 3 columns of Table 2), the relatively high means (26.14 and 25.1) and the narrow range of scores (min1= 12, min2= 17; max1= 27, max2= 30) suggest that the academic word level test proved to be manageable for most candidates on both versions.

Table 1: Results of the academic word level tests

GAU + Academic word level	Version 1	Version 2
N of Items	70	70
N of Examinees	154	66
Mean	46.39	41.14
Variance	56.37	29.15
Std. Dev	7.51	5.4
Skew	-0.25	-0.02
Kurtosis	0.84	-0.41
Minimum score	23	30
Maximum score	67	53
Median	47	41
Cronbach's α	0.83	0.65

Score distribution curves illustrate the above data in a more practical way. Figures 1 and 2 present score distributions for Version 1 and Version 2 (see Appendix for figures). Both curves are shifted downwards showing that the number of correct answers is high, especially on the second version of the test.

The first curve is more elongated which means that the range of scores is wider (4-30) than on version 2 displaying a stubbier curve (17-30). In case of the second version the distribution of scores is closer to normal distribution, although shifted downwards as no candidate scored lower than 17 out of 30.

Table 2: Results of the academic word level tests

GAU + academic word level	Version 1			Version 2		
Part	1	2	3	1	2	3
N of Items	30	10	30	30	10	30
N of Examinees	154	154	154	66	66	66
Mean	14.9	5.38	26.12	13.46	2.61	25.08
Variance	16.12	3.04	11.27	12.49	1.39	7.56
Std. Dev	4.02	1.74	3.36	3.53	1.18	2.75
Skew	0.39	0.02	-2.69	0.71	0.74	-0.6
Kurtosis	0.41	0.14	12.61	1.63	0.28	0.2
Minimum score	6	0	4	6	1	17
Maximum score	28	10	30	25	6	30
Median	14	5	27	13	2	25
Cronbach's α	0.67	0.43	0.78	0.57	-0.11	0.55

Item discrimination indexes range from .00 to .59 and .61 showing a high number of non-discriminating items and a very few highly differentiating ones on both sub-tests. However, the vocabulary sub-tests were found to correlate well with the other two sub-tests of the GAU component on both test occasions (Version1: r_{13}= .502, df= 152, p< 0.001; r_{23}= .400, df= 152, p< 0.001; Version 2: r_{13}= .33, df= 64, p< 0.01; r_{23}= .0103, df= 64, p>0.05) indicating that those who did well on the vocabulary test scored higher on the grammar test as well. This significant correlation might allow us to assume that vocabulary was a good predictor of performance on the grammar test with these two groups of examinees in this particular testing situation, however, further investigation is necessary to be able to arrive at more generalizable conclusions in this field, keeping in mind that correlation indicates relatedness but does not mean causality.

The 10,000 word level test

The four VLT levels were sampled on the basis of word frequency, thus it seems reasonable to expect that the higher the level, the better the test differentiates among candidates with larger and smaller vocabularies, and consequently among more and less proficient examinees. On the other hand, although there is no consensus on the minimal vocabulary size necessary for pursuing university studies (Hazenberg & Hulstijn, 1996; Laufer 1992b; Zechmeister et al., 1993), these findings are by no means to be ignored. Therefore, in 2003 we decided to administer the ten thousand word level of the Levels Test. Two clusters (i.e. six items) were kept to serve as anchors linking the test to the assessment instrument of the previous year for purposes of statistical analysis, which meant dropping two 10,000 level clusters of both Version 1 and Version 2 and substituting them with 2-2 clusters from the academic level. I hypothesized that these two clusters would prove to be less difficult and thus discriminate less than higher level clusters. The following academic level items were used as anchors:

Version 1:

5.
A colleague
B erosion   __F___ action against the law
C format   __B___ wearing away gradually
D inclination   __C___ shape or size of something
E panel
F violation

8.
A anticipate
B compile   __E___ control something skilfully
C convince   __A___ expect something will happen
D denote   __F___ produce books and newspapers
E manipulate
F publish

Version 2:

2.
A debate
B exposure   __E___ plan
C integration   __D___ choice
D option   __C___ joining something into a whole
E scheme
F stability

3.
A access
B gender   __B___ male or female
C implementation   __F___ study of the mind
D license   __A___ entrance or way in
E orientation
F psychology

The use of anchor items necessitated some practical modifications in the cluster number and sequence of the original Schmitt (2000) test battery. To stick to the original idea of inserting 30 vocabulary items (ten clusters) only into the grammar and usage component of our proficiency test, clusters 1 and 2, and 1 and 7 were dropped from the 10,000 word level (Version 1 and Version 2, respectively) due to limitations of space and time. In the new battery then, the above academic clusters were inserted to constitute the first and third clusters. Thus, it follows that in Version 1 items 1-3 and 7-9 (items 41-43 and 47-49 on the whole GAU test component) represent academic level vocabulary, similarly to items 1-3 and 7-9 in Version 2 (items 41-43 and 47-49 on the GAU test).

Version 1 was administered to 94 participants in May 2003, whereas Version 2 was taken by 117 students in June as part 3 of the Grammar and Usage component of the Proficiency test. Before discussing the results item by item, let us examine some general figures. Table 3 presents the results of all three component parts separately for the two versions, part 3 being the thirty-item vocabulary test. The means of the two vocabulary tests exceed the fifty percent of the maximum score (30). The scores are spread out in a wide range (0-23 and 6-26), as shown by the scale of variance values, and high variance is accompanied by high Cronbach alpha figures indicating a high reliability of the instrument. Correlations among the

three subtest were again examined and stronger, significant correlations were found here between the vocabulary and the grammar components (Version 1: r_{13}= .446, df= 92, p< 0.01; r_{23}= .448, df= 92, p< 0.01; Version 2: r_{13}= .341, df= 115, p< 0.001; r_{23}= .31, df= 115, p< 0.01).

Table 3: Results of the 10,000 word level tests

GAU + 10,000 word level	Version 1			Version 2		
Part	1	2	3	1	2	3
N of Items	30	10	30	30	10	30
N of Examinees	94	94	94	117	117	117
Mean	17.63	4.78	15.830	15.97	4.39	17.21
Variance	13.45	2.6	13.42	12.29	2.68	11.12
Std. Dev	3.67	1.61	3.66	3.51	1.64	3.34
Skew	-0.15	0.27	-0.87	-0.07	-0.12	-0.5
Kurtosis	-0.02	0.23	2.48	-0.41	-0.18	1.01
Minimum score	9	1	0	7	0	6
Maximum score	27	9	23	23	8	26
Median	18	5	16	16	4	17
Cronbach's α	0.57	0.34	0.59	0.54	0.29	0.57

The distribution of scores is illustrated for both versions. On Version 1 the curve resembles a platykurtic distribution (Figure 3), but instead of the plateau there are several small peaks at scores 13, 15, 17 and 19. The polygon is a bit negatively skewed, as the mean is a little lower than it would be in a normal distribution, indicating that some students performed much better than expected.

Running the same analysis on the scores of Version 2, we get a distinctly peaked, leptokurtic distribution, where skewness is somewhat higher, but still in the negative (-0.5) as shown in Figure 4. The mean (17.21) of this version was higher than that of version 1 (15.83), with 56% of the examinees scoring below this figure. The positive extreme scores of 23-26 out of thirty were achieved by four students only, shifting the mean upwards.

The discrimination index of an item shows the capacity of the item to differentiate among candidates with larger and smaller vocabularies. Comparing the discrimination indexes of academic level and 10,000 word level items within this test it can clearly be seen that the academic level items' figures all fall in the discrimination index range of the 10,000 level items (.03- .55 and .00- .52 for Version 1 and Version 2, respectively), therefore, it is not legitimate to assume that the academic level items were, in this case, too easy for every candidate and thus differentiate less well.

Conclusion and future directions

Comparing the discrimination indexes of the academic words and the 10,000 word level tests we can conclude that in the case of these four tests and four groups of candidates the 10,000 word level items differentiated among candidates better. Although the two trials brought similar results, the academic word level tests yielded higher means and lower item discrimination index values than the 10,000 word level tests, implying that these words may be considered 'easier' for this population. However, the fact that the candidates taking the four tests were not the same raises a question mark over the generalizability and the comparability of these findings. As an abundance of studies focus on exploring the relationship between vocabulary knowledge and language proficiency (e.g., Morris & Cobb, 2003; Muncie, 2002; Qian, 2002; Zareva, 2005), future plans include comparing performance on vocabulary tests to performance on reading comprehension tests and essay writing tasks, as well as exploring the relationships among these and general academic performance at the English departments.

Less frequent words are not likely to occur in every text and constitute a very low percent of the words in any text, thus do not require specialized teaching. Even more so, as reading is a lot more than vocabulary only: strategy training on how to deal with unknown low frequency words might be beneficial in the long-run. For the forthcoming years, therefore, I find it more plausible to test academic vocabulary as these words occur frequently in academic texts, and thus need special attention and training. At present effort is being devoted to compiling a corpus of representative samples of compulsory readings for English majors at the three English departments of the University of Pécs. The analysis of this corpus may then serve as a basis of developing new vocabulary tests adjusted to the special needs of English majors. Besides measuring how many and what words a language learner is to know for certain purposes, perhaps the most interesting area of vocabulary research today is the research into the depth of word knowledge and the greatest challenge for us would be to develop a test that could measure both the quantitative and the qualitative aspects of lexis.

References

Alderson, J. C. (2000). *Assessing reading.* Cambridge: Cambridge University Press.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Boyd Zimmerman, C. (1997). Historical trends in second language vocabulary instruction. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 5-19). Cambridge: Cambridge University Press.

D'Anna, C. L., Zeichmeister, E. B., & Hall, J. W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behaviour, 23,* 109-22.

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 11* (4), 341-363.

Hazenberg, S., & Hulstijn, J.H. (1996). Defining a minimal receptive second- language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics, 17* (2), 145-163.

Hirsch, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8 (*2), 689-96.

Horváth, J. (1996). The assessment of essay writing skills in the first-year proficiency test. In M. Nikolov, & J. Horváth (Eds.), *Learning lessons: Innovation in teacher education and assessment* (pp. 88-106). Pécs: Lingua Franca Csoport.

Hymes, D. (1971). *On communicative competence.* Philadelphia, P.A.: University of Pennsylvania Press.

Laufer, B. (1987). Teaching vocabulary: The lexical perspective of reading comprehension. *English Language Teaching Journal* (Israel), *35,* 58-67.

Laufer, B. (1991). How much lexis is necessary for reading comprehension? In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132)*.* London: Macmillan.

Laufer, B. (1992a). Reading in a foreign language: How does L2 lexical knowledge interact with the reader's general academic ability? *Journal of Research in Reading, 15* (2), 95-103.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 20-34). Cambridge: Cambridge University Press.

Lehmann, M. (forthcoming) The size of an English receptive vocabulary: An exploratory study of the EFL vocabulary size of English majors at the University of Pécs.

Morris, L., & Cobb, T. (2003). Vocabulary profiles as predictors of the academic performance of Teaching English as a Foreign Language trainees. *System, 32,* 75-87.

Muncie, J. (2002). Process writing and vocabulary development: comparing lexical frequency profiles across drafts. *System, 30,* 225-235.

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* Boston: Heinle and Heinle.

Nation, I. S. P. (1993). Vocabulary size, growth and use. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 115-134). Amsterdam: Benjamins.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception*, Progress Report No. 10. Bloomington, Indiana University.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, *52* (3), 513-536.

RANGE. http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx Retreived on September 1, 2005.

Schmitt, N. (2000). *Vocabulary in language teaching.* Cambridge: Cambridge University Press.

Szabó, G. (1996). Effective filter? Proficiency testing of first-year English majors. In M. Nikolov & J. Horváth (Eds.), *Learning lessons: Innovation in teacher education and assessment* (pp. 76-87). Pécs: Lingua Franca Csoport.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & M. Madden (Eds.) *Input in second language acquisition*. Rowley, Mass.: Newbury House.

Webster's third new international dictionary. (1963). Springfield, MA: G. & C. Merriam Co.

Zareva, A. (2005). Models of lexical knowledge assessment of second language learners of English at higher levels of language proficiency. *System*, *33*, 547-562.

Zechmeister, E. B., D'Anna, C., Hall, J. W., Paus, C. H., & Smith, J. A. (1993). Meta-cognitive and other knowledge about the mental lexicon: Do we know how many words we know? *Applied Linguistics*, *14* (2), 188-206.

Appendix

```
Number  Freq-  Cum
Correct uency  Freq PR PCT
------- ------ ------ ---- ----
No examinees below this score . . .
  3       0      0   1    0   |
  4       1      1   1    1   |#
  5       0      1   1    0   +
  6       0      1   1    0   |
  7       0      1   1    0   |
  8       0      1   1    0   |
  9       0      1   1    0   |
 10       0      1   1    0   +
 11       0      1   1    0   |
 12       1      2   1    1   |#
 13       0      2   1    0   |
 14       0      2   1    0   |
 15       0      2   1    0   +
 16       0      2   1    0   |
 17       1      3   2    1   |#
 18       1      4   3    1   |#
 19       1      5   3    1   |#
 20       3      8   5    2   +##
 21       2     10   6    1   |#
 22       8     18  12    5   |#####
 23       3     21  14    2   |##
 24       8     29  19    5   |#####
 25      21     50  32   14   +##############
 26      23     73  47   15   |###############
 27      22     95  62   14   |##############
 28      25    120  78   16   |################
 29      24    144  94   16   |################
 30      10    154  99    6   +######
                              | ----+----+----+----+----+
                                5   10  15  20  25
                              Percentage of Examinees
```

Figure 1: Academic level score distribution. Version 1

```
Number  Freq-   Cum
Correct uency   Freq PR PCT
 -------  -------  ------ ---- ----
 . . . No examinees below this score . . .
  16      0     0   1   0    |
  17      1     1   2   2    |##
  18      1     2   3   2    |##
  19      0     2   3   0    |
  20      1     3   5   2   +##
  21      5     8  12   8    |########
  22      5    13  20   8    |########
  23      4    17  26   6    |######
  24      5    22  33   8    |########
  25     11    33   5  17   +################
  26     14    47   7  21    |####################
  27      6    53  80   9    |#########
  28      7    60  91  11    |###########
  29      4    64  97   6    |######
  30      2    66  99   3   +###
                            |----+----+----+----+----+
                             5   10   15   20   25
                           Percentage of Examinees
```
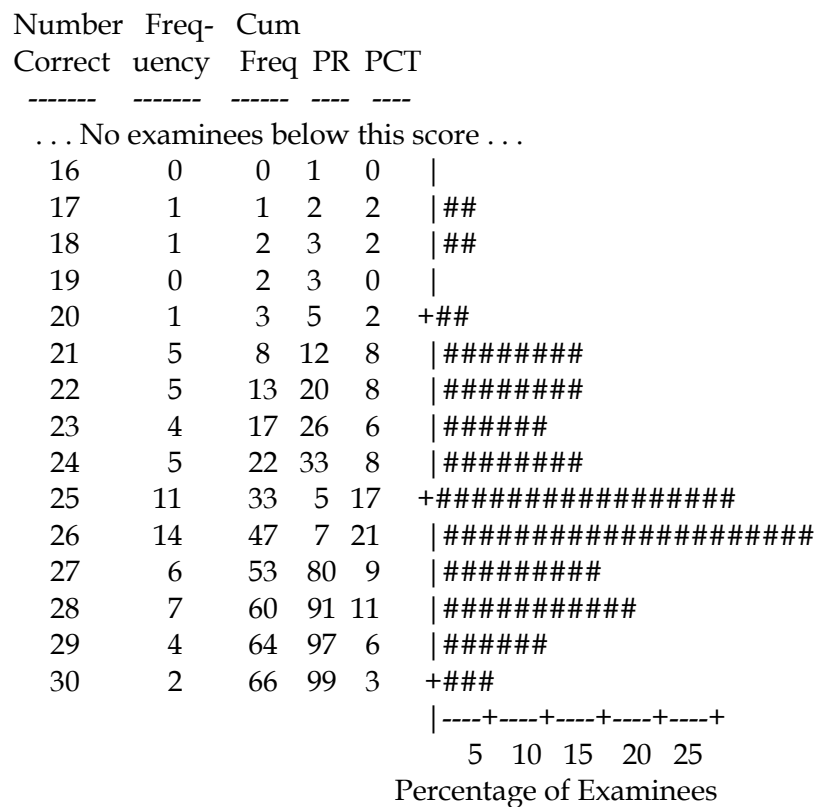
Figure 2: Academic level score distribution. Version 2

```
Number  Freq-  Cum
Correct uency  Freq PR PCT
 ------   ---    ---  ---  ----
   0      1      1    1    1    +#
   1      0      1    1    0    |
   2      0      1    1    0    |
   3      0      1    1    0    |
   4      0      1    1    0    |
   5      0      1    1    0    +
   6      0      1    1    0    |
   7      1      2    2    1    |#
   8      1      3    3    1    |#
   9      1      4    4    1    |#
  10      1      5    5    1    +#
  11      2      7    7    2    |##
  12      7     14   15    7    |#######
  13     10     24   26   11    |###########
  14      7     31   33    7    |#######
  15     13     44   47   14    +##############
  16      7     51   54    7    |#######
  17     12     63   67   13    |############
  18      6     69   73    6    |######
  19     12     81   86   13    |############
  20      6     87   93    6    +######
  21      3     90   96    3    |###
  22      2     92   98    2    |##
  23      2     94   99    2    |##
  24      0     94   99    0    |
  25      0     94   99    0    +
No examinees above this score .     . |
                                |----+----+----+----+----+
                                  5   10  15  20  25
                             Percentage of Examinees
```
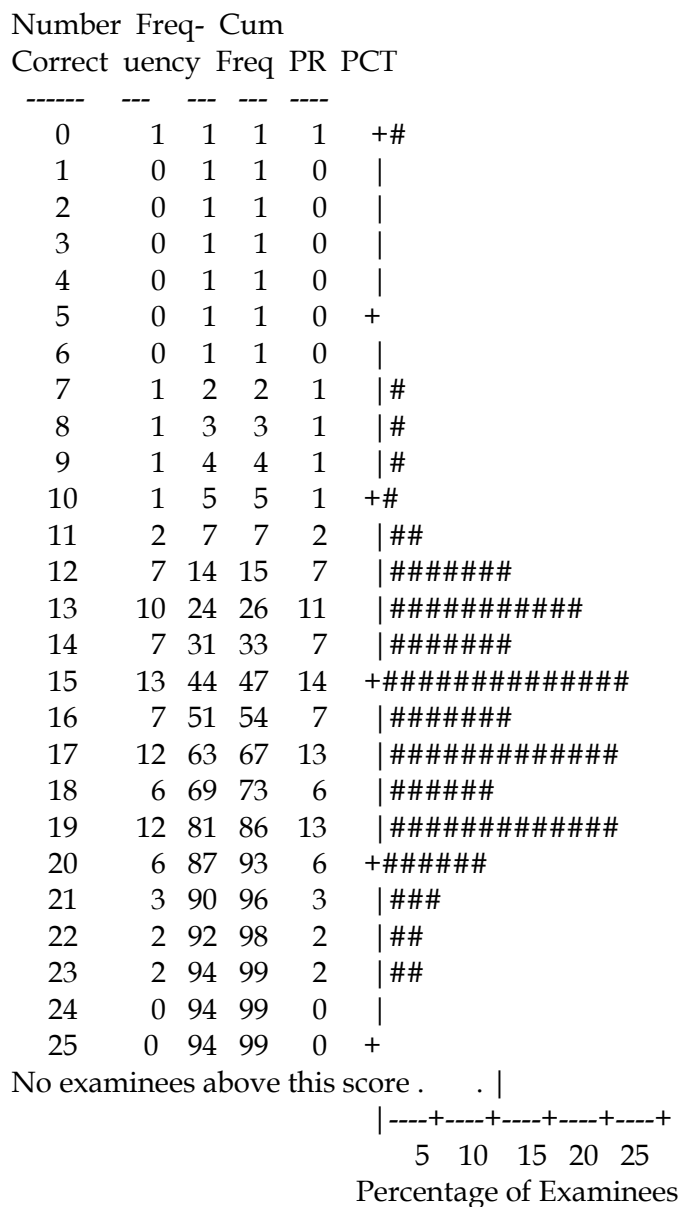
Figure 3: 10,000 level score distribution. Version 1

```
Number  Freq-  Cum
Correct  uency  Freq  PR  PCT
-------       ------- ------ ----   ----
  No examinees below this score . . .
    5     0    0    1   0  +
    6     1    1    1   1  |#
    7     1    2    2   1  |#
    8     0    2    2   0  |
    9     1    3    3   1  |#
   10     0    3    3   0  +
   11     4    7    6   3  |###
   12     3   10    9   3  |###
   13     4   14   12   3  |###
   14     6   20   17   5  |#####
   15     9   29   25   8  +#########
   16    13   42   36  11  |###########
   17    23   65   56  20  |####################
   18    10   75   64   9  |##########
   19    15   90   77  13  |#############
   20     9   99   85   8  +#########
   21     9  108   92   8  |#########
   22     5  113   97   4  |####
   23     1  114   97   1  |#
   24     2  116   99   2  |##
   25     0  116   99   0  +
   26     1  117   99   1  |#
   27     0  117   99   0  |
   28     0  117   99   0  |
  No examinees above this score . . .    |
                        |----+----+----+----+----+
                         5   10   15   20   25
                        Percentage of Examinees
```

Figure 4: 10,000 level score distribution. Version 2

The Effects of Hypertextual Input Modification on L2 Vocabulary Acquisition and Retention

Gyula Sankó

Department of English Language Learning and Teaching, University of Debrecen, Hungary

Sanko@delfin.unideb.hu

Introduction

Recent vocabulary learning research has been based on cognitive interactionist theories of psycholinguistics and second language acquisition (SLA), which emphasize the importance of input, interaction and output in the SLA process. It has also been proposed that learners can successfully integrate L2 input into their knowledge system, provided they apperceive and comprehend it. The first stage in this process is noticing input, which presupposes the allocation of attention to input, as unattended pieces of information are likely to go unnoticed, which in turn blocks the way to further stages of the language acquisition process. It has been suggested therefore that learners' attention should be directed to input by making it salient, which can be achieved through input enhancement and interactional modifications (i.e. the negotiating of input).

The interactive nature of computers, together with the potential of new hypermedia-platformed information technologies allow learners to engage in interaction with input made salient through consulting hypermedia annotations of various contents. The research in this area has manifested a great diversity. The majority of studies have investigated the potential advantage of multimodal presentation of the material in the computerised annotations comparing the effects of text, graphics, video, and sound (Al-Seghayer, 2001; Brett, 1997, 1998; Chanier & Shelva, 1998; Chun & Plass, 1996, 1997). There are also studies examining the effect of using annotations on reading comprehension as well as on incidental and intentional vocabulary learning conditions (e. g., Davis & Lyman-Hager, 1997; De La Fuente, 2002; DeRidder, 2002; Groot, 2000; Harrington & Park, 1997; Hulstijn, Hollander & Greidanus, 1996; Koren, 1999; Laufer & Hill, 2000; Liu & Reed, 1995; Lomicka, 1998; Lyman-Hager, Davis, Burnett & Chennault, 1993; Martinez-Lage, 1997; Rott & Williams, 2003; Son, 1998). These studies usually yielded positive results concerning the impact of computerised glosses (especially multimodal ones) on reading comprehension and the acquisition of L2 lexical elements, but they also pointed out that a lot more empirical research was needed on the practical benefit of using hypertext applications in language instruction. Similarly, there has been little research into what effect the (at least partial) computer imitation of

the form- and meaning-based associations in the mental lexicon might have on the acquisition and retention of new lexical units.

The present study has a double focus investigating both a) the effect of interactional input modification through form- and meaning-focused computerised hypertext annotations on L2 vocabulary acquisition and retention, and b) the effect of incidental and intentional learning on L2 vocabulary acquisition and retention when using form- and meaning-focused hypertext annotations. The study is predominantly text-based, as the author was interested in the effect of various linguistic tools on the success of the acquisition and retention of L2 lexical units, rather than in the influence of multimedia elements that the majority of the previous studies dealt with.

Previous research

Hypertext and vocabulary acquisition

Sengupta (1996) points out that hypertext has several advantages over teaching from a book: it enables quick reading of different linked texts; learners are offered a richly interactive environment where they can get immediate feedback; and they can progress at their own pace. The success of acquiring a second language is deemed to depend on expanding and activating learners' L2 mental lexicons. Comparing the mental lexicon to a computer McCarthy (1990, p. 35) offers the computer metaphor suggesting "split-second processing ability, complex storing with myriad cross-referencing, and virtually instant recall". Umamoto (1997) maintains that since computers were created imitating the human brain, it may be a useful tool for representing the mental lexicon and capable to expand the learners' own mental lexicons. Chanier and Shelva (1998) also state that the nonlinear structure offered by hypertext applications is much closer to the mental lexicon because they reflect the semantic associations between concepts. According to the above, developing easily interpretable lexical networks through the application of hypertext allows learners to have a much better understanding of the meaning of a word than the use of a bilingual alphabetical list, and therefore they also appear to better facilitate vocabulary acquisition.

Modified interaction in hypertext-based L2 vocabulary learning

The interactionist view of language learning asserts that the modified input learners are exposed to is a crucial element in the language acquisition process. Long's (1996) updated version of his interaction hypothesis states that negotiation of meaning promotes acquisition by productively connecting input, internal learner capabilities, (especially selective attention), and output. Negotiation means information restructuring and modification, which takes place with the aim of enabling learners to understand information that is beyond their level of competence.

One of the greatest strengths of computer-based language learning and reference materials is their "ability" to interact with learners working on their own. Brett (1998) and Chanier and Shelva (1998) state that by using interactive hypermedia materials, learners can engage in individualised negotiated clarification of meaning and linguistic phenomena by exploiting available computer resources, where and when their own comprehension breaks down. Simpson (1994) also emphasizes the advantages of interactivity claiming that the high level interactivity provided by hypermedia tools actively engages the brain through interacting and manipulating the environment is likely to result in better, enhanced learning quality. Chapelle (1998) suggests that besides making linguistic input salient, computerised learning environments also need to offer learners opportunities for linguistic input modifications in the form of repetition, simplification through restatements, elaboration through reference materials, and change of input mode, etc. In a computerised reading situation normal interaction means receiving input and requesting more input by scrolling down the page. Modified interaction occurs when the reader cannot understand or remember something, the normal interaction sequence is interrupted by the learner clicking on glosses for grammatical or semantic clarification or by scrolling back to reinforce or check memories of certain information (Chapelle, 2001).

Chun and Plass (1996), Hegelheimer (1998), Laufer and Hill (2000), Lomicka (1998), Lyman-Hager et al., (1993) all investigated the extent to which various forms of on-line 'vocabulary help facilitated L2 learners' reading comprehension and vocabulary retention. Taking everything into account, the findings suggest that learners do benefit from interactional modification in the form of hypermedia glosses. Therefore, it is worth investigating further how the quality and appropriate use of hypertext glosses influence the effectiveness of hypertext computer technologies in vocabulary acquisition

Input enhancement through electronic glossing

According to Gass's (1988) interactionist cognitive model of second language acquisition, learners receive input in the form of target language words, some parts of which are not noticed by them for various reasons. It is only the apperceived (i. e. noticed) part of the input that is going to be processed further. Following from this Chapelle (1998) suggests that instructional materials should be designed in a way that they include features that instigate learners to notice important aspects of the linguistic input. Attention is essential for learning, as attention is the necessary and sufficient condition for long-term memory storage to occur, and it is necessary for input to become available for further mental processing.

> The more a learner pays attention to a word's morphophonological, orthographic, prosodic, semantic, and pragmatic features and to intraword and interword relations, the more likely it is that the new lexical

information will be retained. (Hulstijn, 1990, as cited in Laufer & Hill, 2000, p. 59)

Anderson (1990, p. 183) notes that when information is committed to memory, elaboration with additional information may facilitate recall by adding additional retrieval paths. Depth of processing is best explained as the number of elaborations generated by the learner and consequently referred to as "elaborateness of processing". Chun and Plass (1996, 1997), Hulstijn, Hollander and Greidanus (1996), Knight (1994), Luppesku and Day (1993), Lyman-Hager and Davis (1996), Mondria (1993) have all found that L2 vocabulary acquisition depends on attention and the quality of information processing. Therefore, these researchers argue that tasks that direct learners' attention to the target words (e. g., the use of dictionaries or glosses) will favourably influence both acquisition and retention of unknown vocabulary. Highlighting hyperlinks and therefore making them very visible helps noticing, i.e. it guides learners' attention to the potentially problematic target word forms. However, it may also encourage learners to click excessively, which could lead to a more temporary retention as a result of processing the particular lexical units more shallowly (Roby, 1999). It is hypothesized that existing but invisible hyperglosses, on the other hand, may incite learners to use annotations more rationally and circumspectly, and rely on inferencing from the context more heavily. This enhances mental effort and allows for deeper processing, which is likely to result in firmer acquisition and longer retention (De Ridder, 2002).

The impact of glossing individual vocabulary items as a way of interactional input modification has recently been in focus. Researchers have been trying to find out as much as possible about the effect of applying a variety of glosses in various modalities. The gloss types that have been investigated include printed text, graphics, video, and sound. Davis and Lyman-Hager (1997) claim that computerized glosses in general are appealing, because the computer's capacity allows for a more complex, many-sided and interesting glossing than a printed format does. Rott and Williams (2003) also affirm that glosses, in general, have been shown to aid or at least not interfere with text comprehension. However, they also add that the use of glosses containing merely L1 meaning have resulted in mixed findings in vocabulary learning and retention research. Research conducted by Brett (1997, 1998), Chun and Plass (1996, 1997), Hulstijn, Hollander and Greidanus (1996), Lomicka (1998), Lyman Hager, Davis, Burnett and Chennault (1993) on the effectiveness of annotations on text comprehension and vocabulary learning found them to be beneficial to several aspects of language learning, including reading comprehension and vocabulary learning.

Computerized glossing has been found to have a favourable effect on both L2 vocabulary acquisition and reading comprehension due to the "availability of different types of information, the absence of interruptions during reading, the generation of causal inferences, and the construction of a situation model" (Al-Segayer, 2001, p. x).

The role of form and meaning in vocabulary acquisition

It has been shown that form and meaning compete (VanPatten, 1990). Whether attention paid to word form or word meaning is more important for successful L2 vocabulary acquisition has been a topic of ongoing debate for researchers. Should form-focused instruction be practised, which develops explicit knowledge, or we should aim at meaning-focused instruction that results in implicit knowledge? In studying new vocabulary is it the meaning- or the form-oriented learner-reader orientation that yields both better immediate acquisition and long-term retention of the new lexical items? Singleton (1999, p. 151) emphasizes the importance of formal attributes in the early stages of dealing with new lexical items claiming that the phonological factor is not peculiar in the work of the L2 mental lexicon "but is prominent in the early stages of dealing with particular lexical items in both L1 and L2". Singleton also highlights the significance of semantic associations in lexical processing, which manifest the apprehension of new lexical items and constitute a significant organizing element in the mental lexicon. He concludes that ignoring either form or meaning in lexical research and concentrating only one of them is a reductionist attitude through which no real picture of the lexical processing can be formed (p. 273.).

Harrington & Park (1997) also state that, due to the complex nature of the lexical internal links, a set of links is unlikely to function in isolation. Therefore, it is essential to examine how form-based links interact with other types of links. For this purpose they compare and contrast meaning-based links with the form-based links. Form-based processing seems to be relatively more important for the beginning learner, who spends more time doing bottom-up linguistic processing studying the orthographic and phonological aspects of vocabulary. For more advanced learners, semantic links can be used as a tool to explore and develop their own networks of lexical knowledge (Levy, 1997, as cited in Harrington & Park, 1997, p. x). It has yet to be demonstrated that systematically teaching these links will actually effect the development of vocabulary knowledge in the L2 learner.

Incidental vs. intentional vocabulary acquisition

Hulstijn (1993) puts forward the 'mental effort' hypothesis, which claims that the greater the mental effort required by inferring, the better information recall and retention can be achieved. Hulstijn (2001, p. 266) claims that

> (a) most vocabulary items are acquired incidentally, that is, as a by-product of the learner being engaged in a listening, reading, speaking or writing activity, and (b) that few words are acquired by an act of 'intentional' learning, as in the learning of a bilingual vocabulary list.

Groot (2000), on the other hand, expresses a view that due to the lack of sufficient time for exposure to new words this cannot be replicated in L2 vocabulary learning. Consequently, due to the more superficial exposure, the processing of the new

lexical units will be shallower. This, in turn, will hinder the creation of a sufficient number of associations and links with other words for solid storage and efficient retrieval.

A number of research results (Chun & Plass, 1996; Hulstijn, 1992, 1993; Hulstijn, Hollander, & Greidanus, 1996; Mondria & Wit-de Boer, 1991; Paribakht & Wesche, 1999; Watanabe, 1997) have been published recently on the role of reading and context in L2 vocabulary acquisition. In general it was found that getting students to notice unknown lexical items through turning their attention to them by dictionary use or the application of marginal glosses, incidental vocabulary acquisition during the reading process may be largely improved. Hulstijn (1993) and Laufer and Hulstijn (2001) claim that inferred meanings are more likely to be remembered than meanings provided by glosses. They also admit, however, that in order to be able to infer word meanings from context correctly, learners must be familiar with 95 % of the words in a given text

Purpose of study and hypotheses

The present research had two primary aims:

1. It aimed at finding out whether interactionally modified input made salient through computer-based hypertext annotations would yield better L2 vocabulary acquisition and retention results among intermediate EFL learners than the application of a traditional paper-based text with a paired associates vocabulary list to assist learners.

2. The second main aim of this study was to determine whether there was a difference between the achievement of the aforementioned intermediate students when they acquired new L2 lexical units incidentally and intentionally.

A sub-problem in connection with the first main aim was to shed light on whether the application of form- or meaning-focused gloss contents would give more support to intermediate learners of English involved in the experiment. A sub-problem related to the second main aim was to examine how the potential impact of incidental or intentional learning condition on the acquisition and retention of L2 vocabulary was modified through the application of either form- or meaning focused hypertext annotations.

On the basis of previous research findings the following hypotheses (H) were formulated:

H1: Participants exposed to input during negotiated interaction through link-based hypertext presentation will attain higher levels of L2 vocabulary acquisition and retention than learners subjected to non-negotiated input.

H2: Participants subjected to meaning-focused hypertext input enhancement will outperform learners exposed to form-focused hypertext enhancement in L2 vocabulary acquisition and retention.

H3: Students learning target vocabulary intentionally will attain higher levels of L2 vocabulary acquisition and retention than students acquiring the same vocabulary incidentally.

H4: The application of form- and meaning-focused hypertext glosses will lessen the difference assumed in H3. It is hypothesized that the difference between the attainment of the participants in the control group in the intentional condition and that of the participants in the treatment groups (either form- or meaning-focused) in the incidental condition will not be significant.

Method

Design

A seven-week cross-sectional study was set up. The research had a balanced design with forty subjects in each group studied. Considering the complex nature of the study a mixed research design was applied. A between-groups experimental design was used to measure the potential effects of form- and meaning-focused hypertext annotations compared to the application of traditional paper-based printed text format. To measure the difference in outcome between incidental and intentional vocabulary learning a within-groups design was applied, i. e. the same three groups involved in the experimental situation described above were studied both in incidental and intentional learning conditions. This was organised by giving these groups two different texts with as similar linguistic characteristics as possible, each text on a separate occasion. Besides studying the potential differences between the form- and meaning-focused enhancement and the no-enhancement conditions, on the first occasion (and with text 1) the effects of incidental learning and on the second occasion (and with text 2) the effects of intentional learning were also measured. In order to measure the effects of the above conditions on the acquisition and retention of L2 lexical units both the between-groups and the within-groups designs were mixed with a repeated-measures testing battery including an immediate post-test right at the end of the study session, and a delayed post-test three weeks later.

For a comprehensive description of the experiments see Table 1.

Table 1: The comprehensive research design box for the study

	Form-focused Enhancement		Meaning-focused enhancement		Zero enhancement	
Incidental vocabulary learning	Immediate post-test	Delayed post-test	Immediate post-test	Delayed post-test	Immediate post-test	Delayed post-test
Intentional vocabulary Learning	Immediate post-test	Delayed post-test	Immediate post-test	Delayed post-test	Immediate post-test	Delayed post-test

In order to make sure that the participants really consulted the hypertext annotations, and therefore their vocabulary learning achievement could be potentially attributed to the use of hypertext annotations, a tracking device was attached to

the computer program, which followed and logged the students' moves in the program.

Participants

A total of 120 Hungarian 14-17-year old secondary school students of English participated in the study. Due to feasibility reasons students participating in the actual experiment belonged to 10 intact classes of intermediate English. The students in these classes were then randomly assigned to one of three groups, each using a different treatment method (form-focused hypertext, meaning-focused hypertext or no-enhancement control). All the students to be exposed to computer-based treatment reported that they were computer-literate and knew how to use the World Wide Web.

Materials

Computer program

To enhance L2 vocabulary acquisition by providing readers with the necessary annotational input enhancement of target words via hypermedia links a small-scale computer program was designed in the form of hypertextually-annotated web pages, which was then uploaded to a university server. Four versions of the program were prepared: one for the incidental and another for the intentional learning condition, each of which had a form-centered and a meaning-centered counterpart. In all four cases the program functioned basically in the same way, the difference being only in the texts and naturally in the contents of the belonging hypertext annotations. In each text thirteen words were annotated and all annotated words were printed in a colour different from the rest of the text. The remaining words in the text were not explained in any way. If they happened to be unknown, students were encouraged to infer their meanings from the context of the whole text provided. Students could look up a gloss by clicking on the word in the text located in the upper frame of the screen, which caused the glosses to appear in the bottom frame of the screen. It was hoped that by being able to study both the text (i. e. the context) and the gloss content of a particular word simultaneously, learners would comprehend the text more easily and thereby learn the new target words more effectively.

Hypertext/hypermedia glosses

An essential feature of the study was the application of hypertextually linked annotations instead of the traditionally applied linear text. On the basis of previous studies (e.g., Al-Seghayer, 2001; Chun & Plass, 1996; Liu & Reed, 1995) it was believed that through the application of hypertext and hypermedia glosses learners could have easy access to a wealth of related information.

In form-focused hypermedia annotations the following types of information were presented:

- the basic morphological characteristics of the word including such features as word class, countability, plural form, past tense and past participle forms;
- the basic Hungarian equivalent(s)
- The IPA phonemic transcription of the word
- an audio file with the word pronounced.

The compromise of including the audio component was decided upon, because psycholinguistic research on the L2 mental lexicon (see Singleton, 1999, pp. 148-151) shows that L2 learners in the earlier stages of learning tend to heavily rely on phonological phenomena.

In meaning-focused hypermedia annotations the following types of information were presented:

- The basic meaning(s) of the word presented in an English monolingual dictionary definition(s)
- Various sense relations of the word (synonyms, antonyms, etc)
- The Hungarian equivalent(s)
- Some contextualized examples (collocations, sentences) of word use.

The rich textual (and audio) information found in the annotations was hoped to speed up the learning as students would not have to interrupt the reading and thus the vocabulary study process by needing to consult monolingual or/and bilingual paper dictionaries. It was also hoped that the ultimate achievement would be better than under traditional vocabulary learning circumstances due to cognitive scaffolding, i. e. the opportunity to activate prior knowledge and link new information to it more easily.

Tracking device and log files

A tracking device was constructed to complement the program with the aim of following the users' vocabulary learning behaviour. It was assumed that if input was made salient through the use of hypertext glosses, it was assumed that learners would attend to such input. Therefore, it was necessary to make sure that all

the participants really consulted all (or at least the overwhelming majority of) the hypertext glosses created to promote their learning in order to be able to conclude that their achievement was probably due the facilitative effect of these annotations. Each time a participant chose to consult a gloss by clicking on the hyperlinked word, the tracker device registered the click in a log file.

Texts and stimulus words

In order to carry out the study it was necessary to find two texts that the participants could learn new words from. The texts had to be of the level: not too easy or not too difficult with challenging new vocabulary provided as roughly tuned input for intermediate learners of English. In order to validate the suitability of the texts, they were analyzed with a concordance and lexical profiling software. The length of both texts was approximately identical, with 275 words in the first text and 282 words in the second. The two selected texts were also analysed for syntactic complexity, readability and lexical density. Concerning readability, the Gunning–Fog index of text 1 was 10.68, and that of text 2 was 9.46. The two texts were very similar as far as their ease of reading is concerned, and that the lexical density of text 1 was 62.18 % and that of text 2 was 58.15 %, which again shows that the two texts are quite similar. However, it also shows that both texts were relatively dense lexically, which may have adversely affected the vocabulary learning process.

After the two texts were piloted for unknown words by 61 intermediate students of English who did not participate in the experiment, 26 words (13 words per text) were eventually chosen. Thus, the number of new (glossed) target words per text equaled approximately 4.6 % of the total number of words in each text, which figure was low enough in order not to disturb the global comprehension level or the learning of vocabulary. The selected target words were also analyzed for various word features (frequency, visual complexity, grammatical category, semantic features), as these characteristics may influence the memorization of target vocabulary (see Laufer, 1997).

Testing instruments

Two types of vocabulary tests were administered in order to measure the levels of previous (i.e., pre-treatment) word knowledge, as well as the immediate and delayed effects of the hypertext treatment conditions. In all three tests learners were asked to perform recognition tasks.

Pre-test

In order to make sure which words the participants were not familiar with the target words, a self-report pre-test was conducted. The participants were asked to mark the words they knew on a list of 50 words (preselected by piloters) plus 12 non-words.

Immediate and delayed vocabulary acquisition post-tests

A so called "banked gap filling" vocabulary testing technique was used to measure participants' receptive vocabulary learning achievement. The vocabulary tests were carried out in a traditional paper-and-pen format in both the experimental and the control groups. In order to potentially enhance students' vocabulary acquisition achievement, it was decided that the original context (i. e. the texts used in the study sessions) would be used. Each testing session lasted 15 minutes. To enhance reliability, all post-tests were piloted by 61 students not participating in the study, and they were scored by the researcher only.

Procedure of data collection

The study was conducted over a period of seven weeks. Data collection took place during regular class time in three different secondary schools of Debrecen. The experiment was composed of three basic stages: 1) a pre-test 2) a study and immediate post-test session, and 3) a delayed post-test, which was the exact replica of the immediate post-test and a questionnaire. The learning sessions were computer-based in the form-focused and the meaning-focused experimental groups, whereas the control group got the same texts in traditional paper-based hard copies. All testing sessions were done with paper and pen. The exact schedule of data collection during the experiment can be seen in Table 2 below.

Table 2: Schedule of data collection

Time	Action Taken
Week 1	Self-report pre-test
Week 2	Tutorial session 1 – incidental learning condition
	Immediate post-test 1 (at the end of the study session)
Week 3	Tutorial session 2 – intentional learning condition
	Immediate post-test 2 (at the end of the study session)
Weeks 4-6	Break
Week 7	Delayed post-tests (the exact replicas of immediate post-tests 1 and 2)

During week one participants were asked to take a paper-and-pen self-report pre-test, so that a final set of target words unknown to all students could be selected. After analysing the results of the self-report pre-test, the form of the materials to

be used in the study was finalised, and the experiment was launched in three days.

During week two the first treatment session took place under incidental learning conditions. Before the study sessions students in the ten intact classes were randomly assigned to the form-focused, meaning-focused or the control learning conditions. At the beginning of the incidental vocabulary learning session participants were asked to study the text and the additional information provided (i. e., the glossary of paired equivalents in the control group, and the gloss content in the experimental groups) carefully, and they were forewarned that after studying the passage for 30 minutes they would be given a test checking comprehension. In this way it was hoped that learners would be concentrating on understanding the text as much as possible, rather than on learning the target words. In other words, it was hoped that in this way incidental vocabulary learning would take place (see Hulstijn, Hollander & Greidanus, 1996, p. 327). The students in the two experimental groups were working with computers. They received input in a visually-enhanced form, and they were provided help in the form of hypertext/hypermedia annotations to be able to understand the target vocabulary. The members of the control group received the same text without any enhancement. They were given help to understand the stimulus words in the form of a bilingual word list which, besides the Hungarian equivalents, also included the phonemic transcription of the target words. After 30 minutes students were unexpectedly given a 15-minute vocabulary test on the target words to measure their receptive incidental word acquisition achievement.

In week three the same procedure was repeated as in week two. Besides the fact that the same participants were given a different text, the only difference was in the introduction to the session. At the beginning of these sessions participants were asked to learn as many new words from the text as possible. They were forewarned that at the end of the session they would be given a test measuring the acquisition of the new lexical items. In order to avoid their learning only the enhanced words or only the words provided on the bilingual list (as a result of their test-taking experience from the previous session), they were also forewarned that any lexical item occuring in the text could be tested. Beyond such "guidance" participants were given no more help with the target texts whatsoever. Instead, they were encouraged to try and infer the meanings if they should come across unknown words other than the ones explained in the glosses or translated on the bilingual list. So as to measure immediate word gain, students were given a 13-item banked close test with the target words omitted in a text. To avoid the variation-due-to-task effect, it was decided that the texts used in the tutorial sessions would be used for the tests without any alterations at all.

In week seven, three weeks after the last study session, participants were unexpectedly given the same two vocabulary tests as they had been given in week two and three respectively. As they were asked to take two tests they were allowed to work for 30 minutes. The aim of these tests was to measure potential long-term retention of the words acquired in the previous two sessions. In order to prevent students from consulting the web documents during the "three-week rest" period, the web sites were removed from the server immediately after the

last session of week 3. The teachers of the students involved in the experiment also confirmed that during this three-week period students had not learned any of the target vocabulary institutionally.

Data analysis

Both a within-subjects and a between-subjects repeated-measures multivariate analysis of variance (MANOVA) was applied to try and find significant differences in the vocabulary test achievements between the different methods and learning conditions. As the test results gained did not represent normally distributed data, nonparametric statistical methods were used. The Kruskal-Wallis nonparametric test procedure followed by paired Mann-Whitney U-tests provided analysis of variance for the effects of the three different and independent input types (viz. form-focused, meaning-focused and control), and the Friedman test repeated-measures procedure complemented by the Wilcoxon Signed Rank Test provided analysis of variance for making the same measurement twice with related data, once in incidental and once in intentional learning condition with the same research population. The hypotheses were tested at a $p < .05$ level of significance. The results gained through the use of the testing instrument were checked for validity. In order to verify that all students consulted all, or the prevailing majority of the glosses in the text, the average number of glosses consulted without repeated clicks was calculated.

Results and discussion

Descriptive analysis of test results

The results observed in Table 3 below suggest various implications. For one thing, it can be clearly seen that the range of results is rather wide in almost all groups, which implies that the tasks were difficult enough for the target population. The widest range can be observed in the delayed post tests of the groups, especially under incidental condition where the range was 100% of the total score in all three groups (i.e., in those subjected to texts with form- and meaning-focused enhancement and also the control group that received input without such enhancement).

The narrowest range can be observed in the immediate post test results of groups in the intentional learning condition. While the range of scores of the participant groups generally varies between 11 and 13, which means 84.6% to 100 % of the total score, in the case of the immediate post test under intentional experimental condition the range is 8 (61.5% of the total score) in the group exposed to texts with form-focused enhancement, and it is only 5 (38.4% of the total score) in the group using the meaning-focused annotations.

Table 3: Descriptive analysis of the mean test results

| | Incidental condition | | | | | | Intentional condition | | | | | |
| | Immediate post test | | | Delayed Post test | | | Immediate post test | | | Delayed Post test | | |
	F	M	C	F	M	C	F	M	C	F	M	C
Mean	9.45	10.05	8.98	6.75	8.78	7.18	11.3	11.93	10.00	7.73	8.40	7.50
Std. Dev.	4.24	3.68	3.31	5.08	4.38	4.30	2.53	1.65	3.17	4.33	4.14	4.10
Min.	0	1	2	0	0	0	5	8	2	0	1	0
Max.	13	13	13	13	13	13	13	13	13	13	13	13
Range	13	12	11	13	13	13	8	5	11	13	12	13

F = input provided with form-focused enhancement
M = input provided with meaning-focused enhancement
C = (control) input provided with no enhancement

The first main aim of the research was to examine whether intermediate level students exposed to texts with interactionally enhanced input modifications in the form of form- and meaning-focused computer-based hypertext/hypermedia annotations would outperform the control group students, who received input in traditional paper format, without such enhancement. The results of the Kruskal-Wallis test suggested that statistically significant differences (p=0.011) exist between the acquisition achievements of the form-focused, meaning-focused and the control groups only in the intentional learning condition and on the immediate post-test.

Hypothesis 1

The Mann-Whitney U test analyses showed that members of the group using form-focused input enhancement significantly (p= 0.020) outperformed the control group on the immediate post-test in the intentional learning The difference between the performance of the students in the group using meaning-focused annotations and that of the control group proved to be even more robust (p= 0.004) in favour of the former.

The results did not provide satisfactory evidence to fully confirm this hypothesis. Even though there was a general tendency observable according to which participants using hypertext annotations had better achievements on both immediate and delayed post-tests in incidental and intentional vocabulary learning conditions, this tendency was not significant statistically, therefore it must be considered as conjectural evidence. As can be seen above, participants using form-focused annotations and those using meaning-focused annotations both significantly outperformed students using the traditional paper-and-pen bilingual word list technique on the immediate post-test in the intentional learning condition. The same statistically significant (p < 0.05) difference was not to be found on the delayed post tests or in the incidental learning condition. Therefore, in sum, it must be stated that the results of the current research do not fully confirm the claim formed in hypothesis one.

Hypothesis 2

When comparing the immediate post-test results of the meaning-focused group with those of the form-focused group using the Mann-Whitney U test the mean ranks as well as the sum of ranks clearly demonstrated that the students using meaning-focused input enhancement outperformed those using form-focused input enhancement. The results of the analysis also show, however, that no significant difference can be spotted between the test results of these two groups. The strongest significance found in the intentional condition on the immediate post-test was only p= 0.438. Therefore Hypothesis 2, which predicted that learners subjected to meaning-focused input during negotiated interaction would outperform those subjected to form-focused enhancement condition in L2 receptive vocabulary acquisition and retention under both incidental and intentional learning conditions has to be rejected.

Hypothesis 3

The second main aim of the study was to identify whether the students acquiring new L2 words incidentally would be outperformed by students learning new vocabulary intentionally.

The rank values of the Friedman test suggest that the two levels (i.e. the incidental and intentional independent variables) produce different outcomes: (p=0.000) calculated level of significance of the within-groups repeated-measures statistical test. The Wilcoxon Signed Ranks tests applied as a follow-up, reporting on the evidence of the results of the immediate post tests, show that the participants learning the new words intentionally significantly (p= 0.000) outperformed those acquiring the same words incidentally. The same cannot be said for the delayed post-tests though. In the case of delayed post- tests, the learners under the intentional learning condition also performed better than those under the incidental learning condition This difference, however, was statistically not significant (p= 0.184). This means that hypothesis 3 stating that students learning target vocabulary intentionally will attain higher levels of L2 vocabulary acquisition and retention than students acquiring the same vocabulary incidentally was only proved for acquisition, but disproved in the case of the delayed post-tests measuring retention.

Hypothesis 4

In the case of Hypothesis 4, the paired Mann-Whitney statistical analyses showed that participants using meaning-focused input enhancement in the incidental learning condition outperformed students using no enhancement in the intentional learning condition. However, the difference in favour of the students using meaning-focused enhancement was not significant (p=0.652). Students using form-focused input enhancement in the incidental learning condition were slightly outperformed by particpants using no-enhancement (control) group in the intentional

learning condition, but the difference was also insignificant (p=0.909). As the above data show, even tough intentional vocabulary learning proved to be more effective than incidental learning, thanks to the beneficial effect of interactional input modification, there are no significant differences between the two pairs of groups compared. This also means that hypothesis 4 stating that the difference between the attainment of the students in the control group in the intentional condition and that of their peers in the treatment groups (either form- or meaning-focused) will not be significant has been proved.

Verifying test results

The results of the above four tests for all 120 students participating in the research were subjected to statistical item analysis in order to draw conclusions about the reliability, facility value, homogeneity and discriminating power of the tests. Table 4 presents the descriptive statistics from the item analysis of each of the above tests.

Table 4: Item analysis statistics of the post-test results for all 120 participants

| | Incidental Learning Condition | | Intentional Learning Condition | |
	Immediate Post-test	Delayed Post-test	Immediate Post-test	Delayed Post-test
N of Items	13	13	13	13
N of Examinees	120	120	120	120
Mean	8.175	7.833	11.283	9.750
Variance	17.078	22.206	6.320	14.087
Std. Deviation	4.133	4.712	2.514	3.753
Skew	-0.208	-0.338	-1.502	-0.919
Kurtosis	-1.346	-1.339	1.559	-0.375
Minimum	0.000	0.000	2.000	0.000
Maximum	13.000	13.000	13.000	13.000
Median	8.000	8.000	13.000	11.000
Alpha	0.903	0.936	0.844	0.903
SEM	1.285	1.188	0.996	1.168
Mean P	0.629	0.603	0.868	0.750
Mean Item-Tot.	0.675	0.753	0.576	0.680
Mean Biserial	0.881	0.962	0.936	0.937

Table 4 shows that in the incidental learning condition the mean immediate post-test result is 8.175, and the mean delayed post-test result is 7.833 only, which natural decrease can be put down to forgetting during the three weeks between the immediate and the delayed post-tests. It is apparent from the table that the particpants acquiring new lexical units in the intentional learning condition outperformed their peers in the incidental learning condition both on the immediate and the delayed post-tests. The medians also demonstrate better achievement under the intentional learning condition.

The average proportion of the results in the four tests are 0.629 (62.9 %); 0.603 (60 %); 0.868 (86%); and 0.750 (75%). These figures demonstrate a slightly higher

than the ideal (especially in the case of the immediate post-test in the intentional learning condition) facility value. Despite some of the comparatively high facility values the mean item totals indicating the discrimination indices of the four tests (in the order indicated above) are 67.5 %; 75.3 %; 57.6 %; and 68.0 %. Even though the number of test items in each test was rather limited (13), the above figures show that the mean item totals (i. e., the discrimination index of the tests) were between 0.573 and 0.753. This proves that the level of the banked cloze tests applied discriminated between the student-participants quite well as the discrimination index is well above 30%. The figures also show considerable inter-item consistency, as the Cronbach alpha index in the four tests were 0.903; 0.936; 0.844; and 0.903, that is they ranged between 0.844 and 0.936. As can be clearly seen, the Cronbach alpha values range between 0.844 and 0.936, all fairly close to 1, which indicates high test reliability.

Log results

The data in the log provided convincing evidence that the great majority of glosses were indeed consulted by the students. Analysing the log files, which reflected learners' study behaviour through their clicking manner, it was found that the average number of target words consulted in text 1 (measuring incidental learning) was 11.75 (90.38%) in the case of the version with form-focused enhancement, and 12.15 (93.46%) in its meaning-focused counterpart. In text 2 (measuring intentional vocabulary learning), the average number of such stimulus words consulted was 12.52 (96.30%) in the text version with form-focused annotations and 12.52 (96.30%) in the meaning-focused annotated version. Considering that the total number of stimulus words in each text was 13, the average results gained are quite convincing.

Conclusions

The present research findings have proved that hypertextual input enhancement favourably affects vocabulary learning, and they have also shown that intentional vocabulary learning appears to be more effective than its incidental counterpart. The outcome of the study, however, is only partly convincing, as positive findings are mainly confined to acquisition results. Hypotheses related to retention results measured three weeks after learning have all been disproved. This could be attributed to the fact that the study measured the retention of new L2 words after only one exposure whereas 5-8 exposures (plus regular recycling) would be necessary for gaining solid vocabulary knowledge. Due to retention loss, the scores of the delayed post-tests were significantly lower in the case of both the incidental and the intentional conditions. The assumption that intermediate learners of English would benefit more from meaning-focused glossed texts was not proved. Even though the students using meaning-focused annotations outperformed all other groups in all experimental conditions, these differences could hardly ever be

supported by statistically significant results, which leaves these findings speculative and conjectural.

There were some limitations to this study that might lessen the external validity of the research. Firstly, the results gained are not representative. Secondly, assessment of the receptive learning outcome was measured only with banked cloze tests. By using other assessment techniques, a more complex picture reflecting various aspects (e.g., the productive aspect) of vocabulary knowledge could be provided. Thirdly, the number of test items per text was rather low (13), which may have negatively influenced the explanatory value of results gained through statistical measures. Fourthly, although hypertext programs offer the user relative flexiblility as far as learning styles and strategies are concerned, the software program in the study did not take particular language learners' individual characteristics, memory span, language aptitude, learning styles preferences into consideration, which might also have affected the final outcome of the study. Pre testing participants' exact proficiency levels would have allowed for a more refined analysis.

Even though this study has revealed that significantly better results can be achieved through intentional vocabulary learning, by no means does it mean that the role and significance of incidental vocabulary learning should be neglected. In fact, we are much more frequently involved in incidental vocabulary learning experiences than in intentional ones. We need to raise learners' awareness of the novel study strategies that match new technologies best, and create intelligent software programs that keep track of individual learner differences and preferences, and assist the learner in a „tailor-made" way throughout the learning process.

As was mentioned among the limitations of the study, hypertext programs, although accommodating a range of learning styles, are still rather rigid by nature with a fixed material content, fixed links, and additional information available in the glosses. As not all students are typical, their individual characteristics, learning style preferences, and many more individual features and circumstances should be taken into account when designing the structure of hypertextual language learning environments. Further research is needed to design such systems and investigate their effect on vocabulary learning and second language learning in general. By making hypertext/hypermedia-based language learning environments more adaptive, the learning process could be further individualised, allowing the L2 student to follow a learning path that suits their short- and long-term needs, interests or mood. To achieve this Lomicka (1998) suggests that all users should fill in an online questionnaire on learning style preferences when starting to work with the program, and on the basis of the results gained the software would channel the user accordingly in later stages of vocabulary learning. De Ridder (2002) also emphasizes the importance of research on cognitive flexibility and cognitive profiles considering up-to-date technological developments such as Dynamic HTML, which could adapt the design of the screen to the learning and reading profile of the user. A more sophisticated and flexible model is offered by Eklund (1995), who proposes adaptive hypertext. Such hypertext system would continually monitor the language learner's moves through their clicking behaviour, map

their learning history, remember what words have been consulted, how many times, what information type or help facility has been asked for, and gently assist and guide them along the program (which can also be a relatively open system as large as the Internet) on the basis of the detected, thus potentially providing considerably higher language learning effectiveness.

The present research yielded only certain tendencies that indicate the beneficial effect of such hypertext annotations. The limited framework of the current research, however, does not allow broad generalisations of the results. The current study could be followed up with a much larger sample of participants taken from a wider, more varied population of Hungarian or non-Hungarian learners of English. Also further research is needed to shed light on the most effective way of using form- and meaning-focused hypertext/hypermedia glosses in second language vocabulary acquisition.

References

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning and Technology, 5* (1), 202-232.

Anderson, J. R. (1990). *Cognitive psychology and its implications* (3rd ed.). San Francisco: W. H. Freeman & Company.

Brett, P. A (1997) A comparative study of the effects of the use of multimedia on listening comprehension. *System, 25* (1*)* 39-54.

Brett, P. (1998). The design, implementation and evaluation of a multimedia application for second language listening comprehension Unpublished doctoral thesis. University of Wolverhampton Retrieved July 03, 2004, from http://pers-www.wlv.ac.uk/~le1969/

Cambridge advanced learner's dictionary online – http://dictionary.cambridge.org/

Chanier, T., & Shelva, T. (1998). The Alexia system: The use of visual representations to enhance vocabulary learning. Laboratoire d'Informatique de Besancon, Universite de Franche-Comte (Online). Retrieved March 14, 2003, from http://comte.univ-fcomte.fr/RECHERCHE/P7/pub/ call98/ call98.htm

Chapelle, C. A. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning and Technology, 2* (1), 22-34.

Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research.* Cambridge: Cambridge University Press.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80* (2), 183-198.

Chun, D. M., & Plass, J. L. (1997). Research on text comprehension in multimedia environments. *Language Learning and Technology, 1* (1), 60-81.

Davis, J. N., & Lyman-Hager, M. (1997). Computers and L2 reading: Student performance, student attitudes. *Foreign Language Annals, 30,* 58-72.

De la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary: The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition, 24* (1), 81-112.

De Ridder, I. (2002). Visible or invisible links: Does the highlighting of hyperlinks affect incidental vocabulary learning, text comprehension, and the reading process? *Language Learning and Technology, 6* (1), 123-146.

Eklund, J. (1995). Cognitive models for structuring hypermedia and implications for learning from the world-wide web. In Debreceny RS and A. E. Ellis, Ed. (1995). *Innovation and Diversity - The World Wide Web in Australia. AusWeb95 - Proceedings of the First Australian World Wide Web Conference.* Lismore, NSW, Norsearch Publishing. p.111-116. [Online]. Retrieved on September 09. 2004 from http://ausweb.scu.edu.au/aw95/hypertext/eklund/index.html

Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition, 21* (2), 285-301.

Gass, S. M. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics, 9*, 198-217.

Groot, P. J. M. (2000). Computer assisted second language vocabulary acquisition. *Language Learning and Technology, 4* (1), 60-81.

Harrington, M. & Park, S. (1997). A link-based approach to lexical CALL. *ON-CALL, 11* (3). (Online). Retrieved February 2, 2002, from http://www.cltr.uq.edu.au/oncall/mwharr113.html

Hegelheimer, V. (1998). Effects of textual glosses and sentence-level audio glosses on online reading comprehension and vocabulary recall. Unpublished doctoral dissertation. Department of Educational Psychology, College of Education, University of Illinois, Urbana, IL.

Hulstijn, J.H. (1993) When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *Modern Language Journal, 77* (2), 139-147.

Hulstijn, J. H. (2000).The use of computer technology in experimental studies of second language acquisition: A survey of some techniques and some ongoing studies. *Language Learning & Technology, 3* (2), 32-43.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: a reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal, 80* (3), 327-339.

Knight, S. (1994). Dictionary: The tool of last resort in foreign language reading? A new perspective. *Modern Language Journal, 78,* 285-299.

Koren, S. (1999). Vocabulary instruction through hypertext: Are there advantages over conventional methods of teaching? *TESL-EJ, 4* (1). Retrieved February 26, 2002, from http://www-writing.berkeley.edu/TESL-EJ/ej13/a2.html

Laufer, B. (1997). What's in a word that makes it hard or easy: some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140-155). Cambridge: Cambridge University Press.

Laufer, B., & Hill, M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning and Technology, 3* (2), 58-76.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22* (1), 1-26.

Lightbown, P. M., & Spada, N. (1993). *How languages are learned.* Oxford: Oxford University Press.

Lomicka, L. L. (1998). "To gloss or not to gloss?': An investigation of reading comprehension online. *Language Learning and Technology, 1* (2), 41-50.

Long, M. H. (1996). The role of linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego, CA: Academic Press.

Liu, M., & Reed, W. M. (1995). The effect of hypermedia assisted instruction on second language learning through a semantic-network-based approach. *Journal of Educational Computing Research, 12* (2) 159-175.

Luppesku, S., & Day, R. R. (1993). Reading, dictionaries, and vocabulary learning. *Language Learning, 43* (2), 263-287.

Lyman-Hager, M., Davis, J. N., Burnett, J., & Chennault, R. (1993). Une Vie de Boy: Interactive reading in French. In F. L. Borchardt & E. M. T. Johnson (Eds.), *Proceedings of the CALICO 1993 annual symposium on "Assessment"* (pp. 93-97). Durham, NC. Duke University.

Martinez-Lage, A. (1997). Hypermedia technology for teaching reading. In M. Bush & R. Terry (Eds.), *Technology enhanced language learning* (pp. 121-163). Lincolnwood, IL: National Textbook Company.

McCarthy, M. (1990). *Vocabulary.* Oxford: Oxford University Press.

Mondria, J-A. (1993). The effects of different types of context and different types of learning activity on the retention of foreign language words. Paper presented at the 10th AILA World Congress of Applied Linguistics, Amsterdam.

Mondria, J-A., & Wit-de Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics, 12* (3), 249-267.

Paribakht, T., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition, 21* (2), 195-224.

Roby, W. B. (1999). "What's in a gloss?": A commentary on Lara L. Lomnicka's "To gloss or not to gloss?': An investigation of reading comprehension online. *Language Learning and Technology, 2* (2), 94-101.

Rott, S. & Williams, J. (2003, April). Making form-meaning connections while reading: A qualitative analysis of word processing. *Reading in a Foreign Language,* 15 (1) [Online]. Retrieved September 11, 2004 from http://nflrc.hawaii.edu/rfl/April2003/rott/rott.html

Sengupta, S. (1996). Creating a hypertext database to help Hong Kong English language teachers teach writing. *System, 24*, 187-198. In A. Shah (n.d.). *Text and hypertext as different mediums of creative writings.* [Online]. Retrieved March 4, 2002,

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition, 15* (2), 165-179.

Simpson, M. S. (1994). Neurophysiological considerations related to interactive multimedia. *Educational Technology Research and Development, 42* (1), 75-81.

Singleton, D. (1999). *Exploring the second language mental lexicon.* Cambridge: Cambridge University Press.

Son, J.-B. (1998). Understanding hypertext: A discussion for TEFL. Retrieved June 14, 2002, from http://www.usq.edu.au/opacs/cllt/sonjb/papers/kate98.htm

Son, J.-B. (2001). CALL and vocabulary learning: A review. *Journal of the English Linguistic Science Association of Korea, 7*, 27-35.

Umamoto, T. (1997). Hypertext and the mental lexicon: Using the homepage for learning vocabulary. Hijiyama University [Online]. Retrieved September 11, 2004 from http://www.hiroshima-pu.ac.jp/~umamoto/r22.htm

VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition, 12* (3), 287-301.

Resetting the Null Subject Parameter by Hungarian Learners of English

Ildikó Furkóné Banka

Department of English Language Learning and Teaching, University of Debrecen, Debrecen, Hungary

bankai@delfin.unideb.hu

Introduction

The null subject parameter has been one of the most intensively studied phenomena in the field of theoretical generative linguistics and Second Language Acquisition (SLA) research. In the past decades, a great number of SLA studies have investigated the acquisition of syntactic properties subsumed under the null subject parameter by second language (L2) learners. As the increasing number of research studies indicates, this structural area is rich in properties worth investigating. Also, via the investigation of this parameter insights can be gained into SLA issues such as interlanguage development and parameter resetting.

This paper addresses the issue of L2 learners' developing grammars in the context of parameter resetting and examines the role of transfer in L2 acquisition. More specifically, the present empirical study investigates the acquisition of syntactic properties associated with the null subject parameter by Hungarian learners of English. Subjects were tested on these properties by a grammaticality judgment task (GJT) and a translation task. Particular focus was put on structural differences between Hungarian and English.

The structure of the paper is as follows. First, a short theoretical overview of the notion of the null subject parameter will be given. Then, the issue of parameter resetting in L2 acquisition will be briefly discussed. Third, the method and procedure of the empirical research will be described. Finally, the research results will be presented and interpreted.

Theoretical background

Linguistic focus of the research: the null subject parameter

In the Principles and Parameters framework (Chomsky 1981, 1986), it is assumed that that there is an innate language acquisition device (LAD), which helps children acquire their mother tongue. This language faculty includes Universal Grammar (UG), which consists of two types of components: "(i) rigidly fixed prin-

ciples, which are common to all human languages, and (ii) parameters, choices that have to be set for each language" (Haegeman & Guéron, 1999, p. 584). While principles are the same in all human languages, parameters may take one of two or more values, which are predetermined by UG.

The present paper focuses on the L2 acquisition of one of the parameters, namely the null subject parameter (Chomsky 1981), which is also known as the pro-drop parameter. The term *pro* refers to the empty category of phonologically null subjects. The pro-drop/null subject parameter was one of the first parameters proposed by Chomsky (1981) to account for cross-linguistic variation among languages. It captures the phenomenon that some languages may omit subject pronouns, while others require overt subject pronouns in finite sentences.

The null subject parameter is of binary nature, thus, there are two possible values assigned to it: [+null subject] and [-null subject]. Therefore, languages differ as to whether or not they allow null subjects in finite sentences. In languages like Hungarian, known as [+null subject] languages, null subject pronouns are allowed, taking the form of an empty category, *pro*. In contrast, [-null subject] languages, such as English, require subject pronouns to be overtly expressed.

Null subjects are possible in Hungarian because it has a rich inflectional system. As shown by the following examples, verbs are inflected for person, number, tense and mood.

(én) beszélek	I speak
(te) beszélsz	you speak
(ő) beszél	he/she speaks
(mi) beszélünk	we speak
(ti) beszéltek	you speak
(ők) beszélnek	they speak

As can be seen, Hungarian is a [+null subject] language, with rich inflectional morphology, therefore null subjects can be licensed and identified by agreement on the verb. Of course, it does not follow that [+null subject] languages require all subject pronouns to be null, since overt pronouns are also possible. However, they do not occur in the very same contexts, and there are restrictions on their distribution (for details see White, 2003).

Subject pronouns in Hungarian are rarely used overtly because the verb form, that is, the suffix of the verb, clearly indicates the person and number of the subject. However, if an overt subject pronoun is inserted into the sentence, it is usually used for contrastive purposes. In contrast, English is a [-null subject] language, with poor verbal morphology, thus it requires overt subject pronouns and expletives in finite sentences.

A central claim of the Principles and Parameters framework is that a particular parameter setting is associated with a cluster of syntactic properties. In addition to the possibility of null subjects, the null subject parameter is assumed to be related to two other syntactic properties, namely, subject-verb inversion in declarative

sentences and *that*-trace effect, which refers to *wh*-extraction of subjects across an overt complementizer *that*.

According to Haegeman and Guéron (1999, p. 598), [+null subject] languages have the following four properties in common. The structural differences between [+null subject] and [-null subject] languages are illustrated by Hungarian and English sentences.

Property 1: non-overt pronoun subjects
a, **Szeretem** a fagyit.
b, ***Like** ice cream.

Property 2: post-verbal definite subjects
a, Telefonált **a feleséged**.
b, *Has telephoned **your wife**.

Property 3: wh-extraction of subjects across an overt complementizer
a, **Kit** gondolsz, **hogy** telefonált? (Mit gondolsz, hogy ki telefonált?)
b, ***Who** do you think **that** has telephoned?

Property 4: no overt expletives
a, **Esik**.
b, ***Is raining**.

c, **Világos**, hogy János hazajön.
d, ***Is clear** that John will return home.

The example in (3a) may not seem to be grammatical to every native speaker of Hungarian, but this structure is used in spoken Hungarian. The more appropriate version of the sentence is given in brackets.

As Haegeman and Guéron (1999) claim these four properties are not independent but consequences of the setting of the null subject parameter, hence, they form a cluster. One of the main issues in SLA research is to find out whether or not there is clustering of these properties in L2 acquisition. In other words, are these properties acquired simultaneously or separately by L2 learners?

However, it should be noted that there is no general agreement on the clustering properties. On the one hand, Hilles (1986) and Hyams (1986) propose that subject-verb inversion and *that*-trace can be regarded as parameters of their own. On the other hand, Cook and Newson (1996) claim that only null subjects and subject-verb inversion can be attributed to the null subject parameter.

Parameter resetting in second language acquisition

In SLA research using the generative framework, the achievement of L2 learners is often measured as to what extent they are capable of acquiring the structural properties associated with a parameter in the L2. A related issue is whether L2 learners are capable of resetting a parameter appropriately in L2 acquisition in the case of differing parametric values. If they are able to use all the structural properties within a parameter correctly, they are assumed to have reset the value for the given parameter.

A great number of studies have investigated the issue of parameter resetting, as it is essential in determining the role of UG in L2 acquisition. However, they provide no clear evidence of parameter resetting. Nevertheless, there are four views on parameters in interlanguage grammars as summarized by White (2003, p. 102):

1. global impairment: implies no parameters at all;
2. local impairment, or breakdown: in the case of some parameters;
3. no parameter resetting: assumes that only L1 settings are available;
4. parameter resetting: assumes that it is possible to acquire parameter settings different from those of the L1.

Within the last perspective, two hypotheses are put forth:

1. Full Transfer Full Access Hypothesis: Parameters will initially be set at their L1 values; however, they can be reset to the L2 values. In other words, interlanguage grammars will be restructured in response to L2 input.

2. Full Access without Transfer: L1 parameter setting will not be carried over into the L2, but the appropriate L2 settings will be used immediately.

Since the mid-1980s, a great number of studies have been conducted on the L2 acquisition of the null subject parameter within the UG-based framework, examining the issue of parameter resetting in both directions from [-null subject] to [+null subject] and vice versa, and the effect of the L1 upon the acquisition of the L2 (see Al-Kasey & Pérez- Leroux 1998; Cook, 1996; Hilles, 1991; Liceras, 1989; Phinney, 1987; Tsimpli & Roussou, 1991; White, 1985, 1986;). However, due to the different theoretical frameworks and methodologies, the research studies have often produced inconclusive or controversial results.

Considering the L2 acquisition of the null subject parameter, the following parameter setting alternatives are available for L2 learners (Cook & Newson, 1996):

L1 [+null subject]→ L2 [+null subject]
L1 [+null subject]→ L2 [-null subject]
L1 [-null subject]→ L2 [-null subject]
L1 [-null subject]→ L2 [+null subject]

As Cook and Newson (1996) point out, only if the parameter settings of the L1 and L2 differ, and L2 learners start with the unmarked [+null subject] option, could it be demonstrated that the acquisition of an L2 is the same as that of an L1. Consequently, if speakers of a [+null subject] L1 use null subjects in a [-null subject] L2, this could be due to transfer of the L1 parameter setting to the L2, or they simply start with the unmarked [+null subject] setting like children in L1 acquisition. In contrast, if speakers of a [-null subject] language start with the unmarked [+ null subject] setting in L2 acquisition, this could be evidence for access to UG and the similarity of L2 learning and L1 acquisition.

A major finding of relevant SLA studies is that certain structural properties associated with the null subject parameter are likely to be transferred from the L1 to the L2 if the values for a given parameter are set differently in the two languages. More specifically, transfers tend to appear when L2 learners with an unmarked [+null subject] L1 acquire a marked [-null subject] L2. As Towell and Hawkins (1994, p. 116) claim,

> adolescent/adult native speakers of [+pro-drop] languages such as Spanish, Italian and Greek learning a [-pro-drop] language such as English take time to acquire obligatory subjects, rapidly recognise that subject-verb inversion is not possible, but have great difficulties learning the constraint on *wh*-subject extraction, perhaps never acquiring it.

The study

Since Hungarian and English differ in their parameter settings, it can be assumed that Hungarian learners may have difficulties while acquiring English. The present study investigates the acquisition of syntactic properties subsumed under the null subject parameter from the parameter resetting perspective. In other words, it will be examined if Hungarian L2 learners of English are capable of resetting the parameter from the L1 [+null subject] value to the L2 [-null subject] value while the original setting of the parameter is maintained for the L1. From the learnability perspective, having to reset the parameter may cause problems in the process of L2 acquisition, which may result in transfer errors.

Research questions

In order to investigate the issue of parameter resetting and find answers to the research questions, two groups of Hungarian L2 learners of English were tested on different aspects of the null subject parameter. A further aim of the study was to examine if Hungarian learners would transfer their L1 structures associated with this parameter into their L2 English.

The research questions were the following:

1. Do the properties associated with the null subject parameter cluster in L2 acquisition?
2. If the L1 and L2 differ in parameter values, are Hungarian learners able to reset the parametric value of the L1 into the L2?
3. What structural properties are transferred from the L1 to the L2?
4. What is the relationship between learners' proficiency and their performance on various syntactic properties?

Method

Participants

The participants of this study were 33 Hungarian learners of English (27 females and 6 males with an average age of 19) who had learnt English primarily in a formal setting. All of them were first- and second-year university students of English at the University of Debrecen, enrolled in various classes.

The subjects were divided into two groups on the basis of their scores on the Oxford Placement Test (OPT): there were 14 intermediate and 19 advanced students.

Instruments and procedure

The tasks were performed in the order in which they are presented here. Task 1 was a proficiency test, more specifically Oxford Placement Test (OPT). The experimental tasks (Task 2 and Task 3) tested the following syntactic properties associated with the null subject parameter: (null) subject pronouns (both referential and expletive); subject-verb inversion in declarative sentences; and *wh*-subject extraction for *that*-trace effect. All of the test sentences were controlled for length and simplicity of vocabulary.

The proficiency test and the experimental tasks were administered on three occasions in the fall semester of the 2005-2006 academic year. There were two weeks between each test in order to eliminate cross-task effects.

Oxford placement test

To measure students' level of English proficiency, Oxford Placement Test 2 (Allan, 1992) was administered. The OPT was chosen because it is objective, reliable as well as easy to administer. The test consisted of two parts: a 100-item Listening Test, which focuses on authentic situations where mishearing may take place, and a 100-item multiple-choice Grammar Test, which consists of several parts, testing different aspects of grammar. Subjects were asked to do both parts of the test.

As stated above, the aim of the OPT was to measure L2 learners' level of proficiency. Marking of the test was done according to the User's Guide. The total score was 200 points. The proficiency levels of subjects were identified with the Levels Chart. Subjects were divided into two groups on the basis of their proficiency levels: intermediate and advanced.

Task 2: Grammaticality judgment task

In the present study a 50-item Grammaticality Judgment Task (GJT) was administered in order to tap L2 learners' linguistic competence. As these properties had been tested in other research studies, the test sentences were designed on the basis of earlier SLA studies with slight modification in vocabulary. They focused on the following four syntactic properties: 1 null (referential and expletive) subjects, 2 subject-verb inversion, 3 *wh*-subject extraction/*that*-trace violations, 4 sensitivity to subject/object asymmetry. Students were asked to read and judge individual sentences according to the following categories: *definitely grammatical, probably grammatical, definitely ungrammatical, probably ungrammatical*. They were also asked to go by first impression and were not allowed to go back and change their initial decisions. The aim if this task was to indicate the degree of certainty of the subject with respect to their grammaticality judgment. There were 19 grammatical, 31 ungrammatical sentences in the test.

All the ungrammatical sentences would have been grammatical in Hungarian, but were ungrammatical in English. The ungrammatical sentences can be divided into the following categories, based on the sub-properties tested:

1. Sentences with null subjects:
a, missing referential subjects:
(5) Wakes up every morning at 7.
b, missing expletives (it, there)
(6) Was important for him to pass the test.
(7) Were five people sitting around the table.
2. Sentences with subject-verb inversion:
(8) Went the girl to the cinema last night.

3. Sentences with *that*-trace effects:
(9) What does the man think that caused the accident?
(10) Who do you think that has phoned?

As for the evaluation of the task, the GJT was scored according to the following scoring system (adopted from Kiss-Gulyás, 2004): subjects were awarded three points per correct answer, namely, three points for each correctly accepted sentence (grammatical sentence judged as *definitely grammatical*) and three points for each correctly rejected sentence (ungrammatical sentence judged as *definitely ungrammatical*). Participants were given 2 points if they judged a '*definitely grammatical*' sentence as '*probably grammatical*' and conversely, a '*definitely ungrammatical*' sentence as '*probably ungrammatical*'. They were given no points if they judged a 'definitely grammatical sentences' as 'definitely/probably ungrammatical', and vice versa.

Applying the above scoring system, the total score would be 150. The following features of the data were analyzed and compared: individual achievement in percentage and mean overall judgment on each sentence in percentage.

Task 3: Translation from Hungarian to English

The use of translation tasks in SLA research is a much debated issue because in this type of task it is difficult to make a distinction between transfer errors and developmental errors. Moreover, during translation, transfer occurs naturally. Nevertheless, this type of task was chosen to find out areas of difficulty for Hungarian L2 learners of English, which could be the result of the different parameter settings between the two languages.

The aim of this production task was to examine the use of various syntactic structures in English; moreover, to provide insights into transfer effects, that is the interaction of the participants' L1 and L2. Students were given a list of 22 sentences which were perfectly grammatical in Hungarian, but they would have been absolutely ungrammatical in English, if they had been translated word-for-word. The Hungarian test sentences focused on the following properties: null subjects (both referential and expletives), subject-verb inversion and *wh*-subject extraction across *that*.

Students were asked to translate the sentences one by one in writing, relying on their first impressions, and not to change their answers. The translation task was evaluated in the following way: 1 point was given for structurally correct sentences. No point was given if they applied 'mirror translation' of Hungarian sentence, or if they did not produce the relevant structure. No point was subtracted for using a wrong lexical item. In the data analysis, the total number of points was divided by the total number of sentences to obtain an overall accuracy percentage score.

Results

Results by properties

Assuming that all the investigated properties are related to the null subject parameter, no specific clustering order is expected since all these properties are subsumed under the very same parameter. In other words, syntactic features should simply be acquired at about the same time. Moreover, there should not be any significant differences between syntactic properties. Table 1 presents the means obtained on the properties of the null subject parameter in the GJT. Ungrammatical structures are referred to as (UnG), whereas grammatical structures are indicated as (G)

Table 1: Grammaticality judgment task (Accuracy scores in %)

Sentence type	Intermediate group	Advanced group
Null subjects (UnG)	52.8	72.5
a, referential	47.62	66.08
b, expletive	55.63	76.08
Overt subjects (G)	87.57	82.46
a, referential	92.86	90.18
b, expletive	80.95	72.81
S-V inversion (UnG)	70.24	80.12
S-V inversion (G)	59.52	67.54
that-trace violations (UnG)	49.74	51.46
S-O asymmetry (G)	34.76	43.16
a, subject extraction	31.75	41.52
b, object extraction	39.29	45.61

As can be seen, L2 learners' performance seems to vary from property to property. The intermediate group was more likely to accept ungrammatical null subject sentences in English compared to the advanced group, which may indicate that they transferred the L1 value of the parameter to the L2. It was easier for the participants to judge sentences with missing expletives as ungrammatical than sentences with missing referential pronouns. It is, however, surprising that the advanced group achieved lower accuracy scores in the rejection of sentences with null referential subjects. They appeared to be more tolerant towards subjectless sentences, probably because in spoken English, they may often hear such sentences. As for overt subjects, most subjects accepted the sentences with overt subjects. It was also found that it was easier for students to accept grammatically correct sentences than to reject ungrammatical sentences with missing subjects.

Participants showed the best performance on rejecting S-V inversion in declarative sentences. They were able to recognize easily that post-verbal subjects are ungrammatical in English. They even tended to reject grammatical sentences with

S-V inversion, which may be the result of overgeneralization of the target language rule. Both groups performed poorly on *wh*-subject extraction. Of the three properties, *wh*-extractions proved to be the most difficult for both groups of the participants. They seemed to be quite insensitive to subject-object asymmetry in English, though they were slightly better at accepting grammatical sentences with *wh*-object extraction. This may be explained by the fact that in Hungarian the subject and the object have the same status. Due to the rich inflectional system, not only the subject but the object may be dropped under appropriate discourse conditions. In Hungarian *wh*-questions, the subject as well as the object can be extracted across an overt complementizer. As shown by the results, accuracy in the grammaticality judgments was higher in the group of higher proficiency.

Table 2: Translation task (Accuracy scores in %)

Hungarian sentence types	Intermediate group	Advanced group
Null subjects (UnG)	95.24	97.35
a, referential	85.71	100
b, expletive	91.67	94.7
S-V inversion	91.67	96.5
Wh-subject extraction across *that*	31.43	60

As indicated by the results, the majority of Hungarian sentences with null subjects and S-V inversion were translated into L2 English appropriately. In other words, the majority of the students were able to produce the grammatically correct equivalent for these sentences. As for the translation of Hungarian sentences with *that*-trace violations, learners proved to rely on their L1 grammar heavily, as most of them applied 'mirror translation' for multiple embeddings. However, the advanced group significantly outperformed the intermediate group. To sum up, the results do not clearly justify that the properties subsumed under the null subject parameter are acquired as a cluster. They may rather be acquired individually or in a successive fashion.

3.2 Results by tasks

The results by tasks show whether there is a relationship between the learners' proficiency levels and their performance on the various tasks. The means represent overall accuracy percentage scores across all properties. They were averaged to obtain group results which are presented in Table 3.

Table 3: Results by tasks (in %)

Groups	Task 1 OPT	Task 2 GJT	Task 3 Translation
Intermediate	67.46%	59.33%	76.62%
Advanced	79.84%	68.11%	88.5%

As can be seen in Table 3 performances are higher in the more proficient group, however, participants' performances vary from task to task. While translation proved to be quite easy for the students, the GJT appeared to be much more difficult. This finding may be explained by the fact that students are usually more familiar with translation tasks, as they are often required to do them in language classes. Furthermore, the GJT, which measured learners' linguistic competence in an indirect way, required the use of more complex metalinguistic skills, which may have been less developed than L2 learners' productive skills, which were used in the translation task. To sum up, the participants did not perform consistently across task types.

Discussion

Clustering of properties

In this paper, it was proposed that Hungarian L2 learners may have problems when they acquire English, since there is a parametric difference between their L1 and L2. Participants were tested on three syntactic properties on two experimental tasks: a GJT, and a translation task. A general finding was that the three syntactic properties associated with the null subject parameter do not form a parametric cluster, but they may be acquired individually, in a particular order. In other words, these properties are acquired sequentially, rather than simultaneously in the process of L2 learning. While subjects were able to acquire overt subjects and the proper SVO word order in English, they had problems with *wh*-subject extraction, which proved to be the most complex and difficult structure to produce for both the intermediate and advanced groups. This feature seemed to be resistant to resetting, as even advanced learners performed on it poorly. These findings coincide with the claims of Towell and Hawkins (1994), referred to in the literature review.

Parameter resetting

The issue of parameter resetting is closely related to that of clustering. As has been stated, the syntactic properties under investigation did not seem to be acquired simultaneously as a cluster. The results rather indicate progressive appearance of the properties in learners' developing interlanguage grammars. It does not necessarily mean that the participants in the experiment were not able to reset the null subject parameter. As Ayoun (2000, p. 79) points out, parameter resetting may be evidenced by a partial clustering of properties as a result of progressive manifestation of parameter-setting properties. Moreover, she adds that partial clustering can be explained by two factors: the complexity and systematic application of the abstract syntactic mechanism involved for each parameter and the learners' level of proficiency.

On the whole, the results provide no clear evidence of parameter resetting. If parameter resetting is understood as the simultaneous acquisition of all the syntactic properties associated with a particular setting of a parameter, then parameter resetting has not taken place. Considering individual properties, it was shown that participants performed well on two of the properties, whereas they performed below chance level on *wh*-subject extraction (*that*-trace effect), which was the most complex feature and therefore it appeared to be resistant to resetting. It would be worth examining how near-natives would perform on this property.

Transfer effects

As illustrated by the two experimental tasks, there are certain syntactic properties that can be acquired quite early by Hungarian learners of English, such as obligatory subjects and impossibility of S-V inversion. Accordingly, in the translation task subjects showed a native-like performance in these areas. As opposed to this, the acquisition of the constraint on *wh*-subject extraction posed a serious problem for L2 learners.

The results of the translation task suggest that Hungarian students carried the L1 setting of the null subject parameter to the L2, particularly at a lower level of proficiency. Besides transfer errors, there were some developmental errors as well. In order to show areas of special difficulty for Hungarian learners of English, some examples of students' answers will be given in italics:

Typical examples:

Sentence 9: Hamarosan nagy vihar lesz.
Expected sentence: There's going to be a huge storm.

It will be a huge storm.
It's going to be a big storm.

As shown above, some of the learners seemed to be uncertain as to the use of expletives. They intuitively knew that they were required to use an overt pronouns but they confused the two types of expletives, *it* and *there*. However, the structure they produced is grammatical, but it has a different meaning in Hungarian.

Sentence 21: Úgy tűnik, hogy van néhány probléma.

Expected sentence: There seem to be some problems. OR It seems that there are some problems.

It seems to be he has some problems.
It seems that we have some problems.
It seems like, we got some problems.

As illustrated by the answers given to sentence 21, the use of expletive with a raising-predicate is quite difficult for these Hungarian learners of English. It is interesting that they tended to avoid the use of *there*, instead they used some overt referential subject pronouns. However, this may not be a transfer error, because no overt pronoun was used in the Hungarian sentence.

Sentence 18: Amint kiléptünk a házból, elkezdett esni az eső.
Expected sentence: As soon as we had left the house, it started to rain.

…., the rain started.
…, the rain started to fell.
…, began to rain

The translated versions of sentence 18 clearly indicate a transfer effect. Some students tried to translate the Hungarian sentence word-for word into English. Since in the test sentence the overt subject *rain* was used, students automatically carried it over into English. This type of error frequently is made by Hungarian learners of English at lower proficiency levels. As the third version shows, one of the students used a null subject pronoun in English.

In what follows, the kinds of errors L2 learners made when translating Hungarian *wh*-questions with subject extraction across *that* will be shown:

Sentence 20: Melyik kulcsról gondolod, hogy nyitja az ajtót? / Mit gondolsz, melyik kulcs nyitja az ajtót?

Expected sentence: Which key do you think opens the door?

Which key do you think that opens the door?
What do you think which key opens the door?

Sentence 6: Kit mondtál, hogy szereti Marit? / Mit mondtál, hogy ki szereti Marit?

Expected sentence: Who did you say loves Mary?
Who did you say, who loves Mary?
Who have you said that love Mary?
Who did you say to love Mary?
Whom did you say loves Mary?
What did you say, who loves Mary?
Who said who loves Mary?

As the translations show, students tended to apply 'mirror translation', producing structures that reflected the structural properties of Hungarian, which implies that interlanguage grammars are still constrained by UG.

In the translation task, the most difficult test sentence was #22, which contained multiple embeddings. Only one of the advanced students was able to translate it appropriately.

Sentence 22: Kiről hallottad, hogy János mondta, hogy betörte az ablakot? / Kit hallottál, hogy János mondta, hogy betörte az ablakot?

Expected sentence: Who did you hear John say broke the window?

Who did you hear John said that broke the window?
About whom John said broken the window?
Who did you hear that John said to break the window?
Who did you hear about that John said broke the window?
Who did you heard from that John had a said that he had broken the window?
Of whom did you hear, that John said that he broke the window?
Who did you hear that John said who broke the window?
Who did you hear from that John said that he broke the window?
Whom did you hear what John said that he had broken the window?

Relationship between learners' proficiency and their performance

A gradual improvement was found in participants' performance in relation to their proficiency. The higher their level of proficiency was, the better they performed on the experimental tasks. It can be assumed that the more proficient L2 learners are in the target language, the better they will use certain structural properties in the L2. As indicated by the results, there was a variability in participants' grammaticality judgments on, and production of the syntactic properties subsumed under the null subject parameter, at the two different stages of L2 development. Moreover, the difference between the performance of the intermediate and that of the advanced group may imply that their interlanguage grammars change over time.

In general, it can be stated that Hungarian L2 learners start with the L1 setting of the parameter, but as their proficiency level in the L2 increases, they may gradually readjust the value of the parameter more appropriate to the value of the L2. Data provided evidence that there were some restructuring attempts in the participants' interlanguage grammars with respect to two syntactic properties. However, the third property, that is, *wh*-extraction of subjects across *that* seemed to be resistant to resetting, even in the case of advanced learners of English. Considering this aspect, both groups of the participants produced L2 structures that were constrained by their L1 parameter setting. To sum up, in terms of some grammatical properties the participants showed native-like performance, while in other areas they had particular problems as also indicated by the significant transfer effects.

Conclusion

This paper investigated the effect of the different parameter settings on the interlanguage development of Hungarian learners of English. In other words, the study was conducted to provide insights into the nature of Hungarian L2 learners' interlanguage grammars in the context of parameter resetting. The subjects were tested on the syntactic properties subsumed under the null subject parameter. It was found that the given properties were not acquired together, but sequentially. While the participants had well-formed grammatical representations of two of the properties, they had difficulties performing the constraint on *wh*-subject extraction, which seemed to be resistant to resetting. It was also shown that Hungarian learners transferred the L1-settings of the parameter into their interlanguage grammars, particularly at lower levels of proficiency. Nevertheless, the research results indicated that Hungarian learners of English showed improvement with increasing levels of proficiency, and their interlanguage grammars conformed to principles and parameters of Universal Grammar, however, the data provided no clear evidence of parameter resetting. To obtain more conclusive and convincing results, further research is necessary.

References

Al-Kasey, T. & Pérez-Leroux, A. T. (1998). Second language acquisition of Spanish null subjects. In: S. Flynn, G. Martohardjono & W. O'Neill (Eds.), *The generative study of language acquisition* (pp. 161-185). Mahwah, New Jersey: Lawrence *Erlbaum Associates.*

Allan, D. (1992). Oxford placement test 2. New edition. Marking kit with user's guide and diagnostic key. Oxford: Oxford University Press.

Ayoun, D. (2000). Web-based elicitation tasks in SLA research. In *Language Learning and Technology, 3* (2), 77-98.

Brown, G., Malmkjaer, K., & Williams, J. (Eds.). (1996). *Performance and competence in second language acquisition.* Cambridge: Cambridge University Press.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht: Foris.

Chomsky, N. (1986). *Barriers.* Mass.: The MIT Press.

Cook, V. (Ed.) (1986). *Experimental approaches to second language acquisition.* Oxford: Pergamon.

Cook, V. (1996). Competence and multi-competence. In G. Brown, K- Malmkjaer & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 54-69)*.* Cambridge: Cambridge University Press.

Cook, V., & Newson, M. (1996). *Chomsky's universal grammar: An introduction.* Oxford: Blackwell Publishers.

Eubank, L. (Ed.) (1991). *Point counterpoint: Universal grammar in the second language.* Amsterdam: John Benjamins Publishing Company.

Flynn, S., Martohardjono, G., & O'Neill, W. (Eds.). (1998). *The generative study of language acquisition.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Gass, S. & Schachter, J. (Eds.). (1989). *Linguistic perspectives on second language acquisition.* Cambridge: Cambridge University Press.

Haegeman, L., & Guéron, J. (1999). *English grammar: A generative perspective.* Oxford: Blackwell Publishers.

Hilles, S. (1986). Interlanguage and the pro-drop parameter. *Second Language Research, 2,* 33-52.

Hilles, S. (1991). Access to universal grammar in second language acquisition. In L. Eubank (Ed.), *Point counterpoint: Universal Grammar in the second language* (pp. 305-338). Amsterdam: John Benjamins Publishing Company.

Hyams, N. (1986). *Language acquisition and the theory of parameters.* Dordrecht: Foris.

Kiss-Gulyás, J. (2004). *The acquisition of English restrictive relative clauses by Hungarian learners of English.* Budapest: Akadémiai Kiadó.

Lakshmanan, U. (1991). Morphological uniformity and null studies in child second language acquisition. In L. Eubank (Ed.), *Point counterpoint: Universal grammar in the second language* (pp. 389-410). Amsterdam: John Benjamins Publishing Company.

Liceras, J. M. (1989). On some properties of the "pro-drop" parameter: Looking for missing subjects in non-native Spanish. In S. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 109-133). Cambridge: Cambridge University Press.

Phinney, M. (1987). The pro-drop parameter in second language acquisition. In T. Roeper & E. Williams (Eds.), *Parameter setting* (pp. 221-238). Dordrecht: D. Reidel Publishing Company.

Roeper, T., & Williams, E. (Eds.). (1987). *Parameter setting.* Dordrecht: D. Reidel Publishing Company.

Towell, R., & Hawkins, R. (1994). *Approaches to second language acquisition.* Clevedon: Multilingual Matters Ltd.

White, L. (1985). The pro-drop parameter in adult second language acquisition. *Language Learning, 35,* 47-62.

White, L. (1986). Implications of parametric variation for adult second language acquisition: an investigation of the 'pro-drop parameter'. In V. Cook (Ed.), *Experimental approaches to second language acquisition* (pp. 55-72). Oxford: Pergamon.

White, L. (2003). *Second language acquisition and Universal Grammar.* Cambridge: Cambridge University Press.

Relationships Between Language Achievements in English and German and Classroom-related Variables

Marianne Nikolov

University of Pécs, Pécs, Hungary
nikolov@nostromo.pte.hu

Krisztián Józsa

University of Szeged, Szeged, Hungary
jozsa@edpsy.u-szeged.hu

Introduction

This paper analyzes a large database collected in a project enquiring into Hungarian learners' performances in English and German as a foreign language in state education. A nationally representative sample of year 6 and year 10 learners (n=20,804) was involved in a survey implemented by OKEV in May 2003 (Nikolov & Józsa, 2003). First, we give a short background to the project; then, we discuss the relationships between learners' gender, school type, number of weekly classes, years of language study, students' socioeconomic status, and achievements on proficiency tests in English and German. We explore how extra curricular language study impacts on results. Third, we enquire into how task familiarity and frequency of task type use relate to performances on tests and participants' self-assessment. Finally, we attempt to model how background variables contribute to outcomes.

Background to study

Since the early 1990s students, their parents and teachers, researchers and language policy makers as well as other stakeholders have had great expectations towards language pedagogy. It was assumed that the major change in 1989 allowing students to choose a foreign language they wish to study would boost levels of achievement within a reasonable period of time. A few years later, although there are major favourable developments, the dynamic process is still to be seen. Since then, the number of Hungarians claiming to be able to use a foreign language has tripled: in the most recent Eurobarometer survey 29 percent of Hungarians claim to be able to converse in a foreign language (*Europeans and languages*, 2005). However, this also means that seven out of ten Hungarians would not be able to do so. As the goal for European language education is to allow citizens to

become plurilingual individuals speaking two modern foreign languages besides their mother tongue (*Common European framework of reference*, 2001), there is room for development. In all European countries the majority of people claim they develop language proficiency in state schools (*Europeans and languages*, 2005). One of the more general aims of our study is to explore how realistic this expectation may be towards Hungarian schools.

In recent years several small- and large-scale studies have enquired into issues related to language pedagogy in Hungarian state education. Some of them explored students' language learning orientations and motivation (e. g., Dörnyei, Csizér & Németh, 2006; Nikolov, 2003), aptitude (Ottó, 2003; Ottó & Nikolov, 2003) and some studies triangulated data on cognitive, affective variables and performances on language proficiency tests (Nikolov & Ottó, 2006). In two large related projects in 2000 and 2002 Hungarian students' foreign language skills were assessed in English and German to monitor the levels and efficiency of foreign language education (Csapó & Nikolov, 2001, 2002; Nikolov & Csapó, 2002) in line with other major Hungarian educational research projects monitoring other school subjects (Csapó, 1998; 2002). A third, similarly large-scale project was implemented in 2003 (Nikolov & Józsa, 2003). In the present paper the datasets from this latter assessment project are analyzed.

The role of classroom variables and students' socioeconomic status in performances on English and German proficiency tests

The present paper analyzes a large database collected in a nation-wide project enquiring into Hungarian learners' performances in English and German as a foreign language. A nationally representative sample of year 6 and 10 students (n=20,804) was involved in a survey implemented by OKÉV in May 2003 (Nikolov & Józsa, 2003). The aim of the survey was to gain insights into students' developmental levels in English and German as foreign languages at the ages of 12 (year 6) and 16 (year 10).

Research questions

The present study aims to answer questions related to the relationships among classroom variables and students' achievements in English and German on a representative sample of Hungarian learners. As students participating in the study learnt English and German in years 6 and 10, we intend to explore similarities and differences between the two target languages and the two years. More specifically, the following research questions are formulated:

- What is the relationship between types of schools and students' developmental level in English and German? How do students' achievements on proficiency tests compare to one another according to the type of school they attend?

- How do boys' and girls' performances compare to one another across years and languages?
- What is the relationship between students' achievements and the frequency they claimed they practiced tasks similar to the ones in the test booklets?
- What is the relationship between performances on tasks and their perceived level of difficulty?
- How does the length of language study in years influence outcomes?
- How does the intensity of language study (frequency of weekly language classes) impact on achievements?
- What is the relationship between private tuition and performances on tests?
- What are participants' long term goals for language study?
- What is the relationship between students' socioeconomic status (SES) and other variables?
- How can these relationships be modelled?

Participants

Representative samples of year 6 and 10 students participated in the study. A total of 20,804 learners of English and German in 1,405 groups from 236 primary and 86 secondary schools filled in the data collection instruments. Besides these students an additional 162,814 learners of the two languages took the tests and their performances were evaluated locally with the help of centrally provided keys. However, in the present study results of the representative samples are used for analyses.

Data collection instruments

Three language skills (listening comprehension, reading comprehension, and writing) were assessed by paper and pencil proficiency tests in learners' regular classroom settings at the end of the school year. The English and German tests were based on the prescribed achievement targets of the national core curriculum. They were identical in their construct, structure, type of texts and tasks and length (number of items) for the two target languages; the texts were longer and more complex in year 10 than in year 6 (all test booklets are available in Nikolov & Józsa, 2003). The construct, the task and text types were similar to those used in 2000 and 2002 (Csapó, 2001; Nikolov, 2003), but the actual texts and tasks were different. The task types were expected to be familiar to the vast majority of the participants, as they were similar to published course materials used in schools.

All tasks focused on meaning (and not form), the texts were authentic, except for the listening tasks, where scripted materials were used, as listening comprehension was expected to be the least developed skill according to classroom previous observation studies and questionnaire data (Nikolov, 1999, 2003). In grade 6

writing tasks integrated reading with writing, thus, tasks required learners to choose words based on their meanings and copy them into a gapped invitation to a party and to fill in a form based on a short text with personal data. In year10 students wrote a guided composition on a dream holiday along six content points. Booklets were produced in two versions: the sequence of the tasks was different, but the actual tasks were identical. The reliability (Cronbach alpha) of these tests varied between 0.90 and 0.93 for both English and German reading tasks and writing tasks in year 6, whereas between 0.72 and 0.85 for listening comprehension. Inter-rater reliabilities for the assessment of writing tasks in year 10 were 0.91 for English and 0.95 for German.

Besides the language tests, a questionnaire was administered to all students on their family background, end of term school mark in English or German, number of weekly classes and years of study, language learning plans, and extracurricular activities. Three items related to the tasks in the booklets: students were to indicate on a 1-5 scale how easy or difficult they thought the tasks were and how often they did tasks similar to the ones in the booklets in their English and German classes.

Procedures

The paper and pencil tests were administered locally with the help of external assessors in students' own classrooms in May 2003. All students were given a code to ensure that their personality rights were respected. The reading booklets including five reading tasks were filled in a 45-minute class session, and the listening and writing tasks were accomplished in another session of the same length on the same day. A third session was devoted to filling in the questionnaires. Writing tasks were evaluated centrally by two teams of trained assessors after standard setting sessions; data were entered into files during the summer. Schools got their coded feedback on their own students' performances and the national standards during the fall and a full report was published on the internet and in print (Nikolov & Józsa, 2003).

Results and discussion

The relationship between types of schools students attend and their developmental level in English and German

Students in year 6 attend two types of institutions: the majority study in their 6th year at 8-year primary schools, whereas some of their peers attend 8-year grammar schools where they are in their 2nd year. Achievements on the tests are significantly different in the groups attending these two types of institutions. On English tests the mean score in primary schools is 56.3 percent, whereas learners in secondary schools achieved a mean score of 75.7 percent. On German tests the achievements are 44.0 and 60.0 percent respectively.

As these results indicate, learners of the two target languages performed on significantly different levels, and learners of English outperformed their peers in both groups. Figures 1 and 2, where each dot represents a group of participants (the unit of language learning), show the distributions of year 6 students and the lines indicate mean scores for the two cohorts.
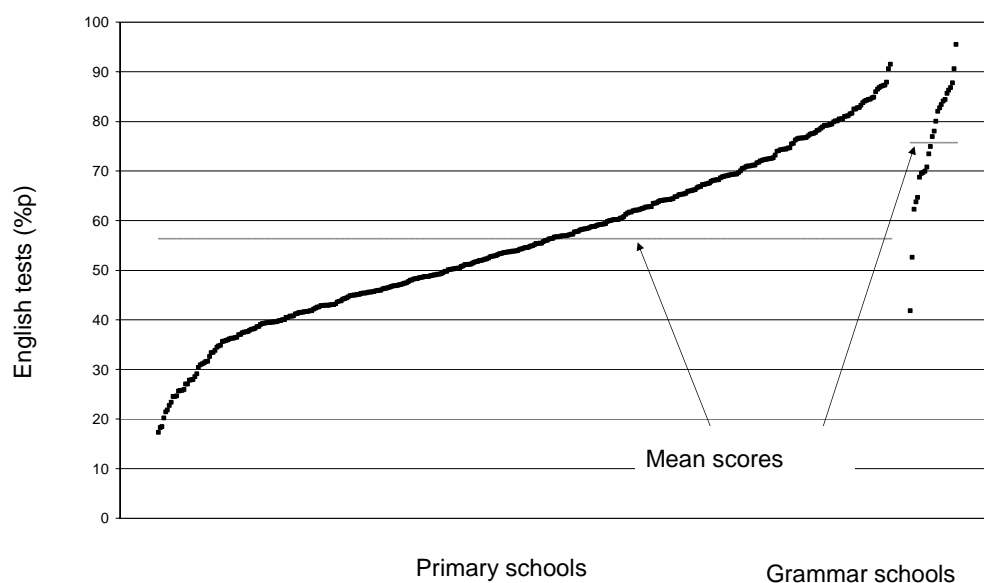

Figure 1: Distribution of year 6 English learners' groups in primary schools and grammar schools
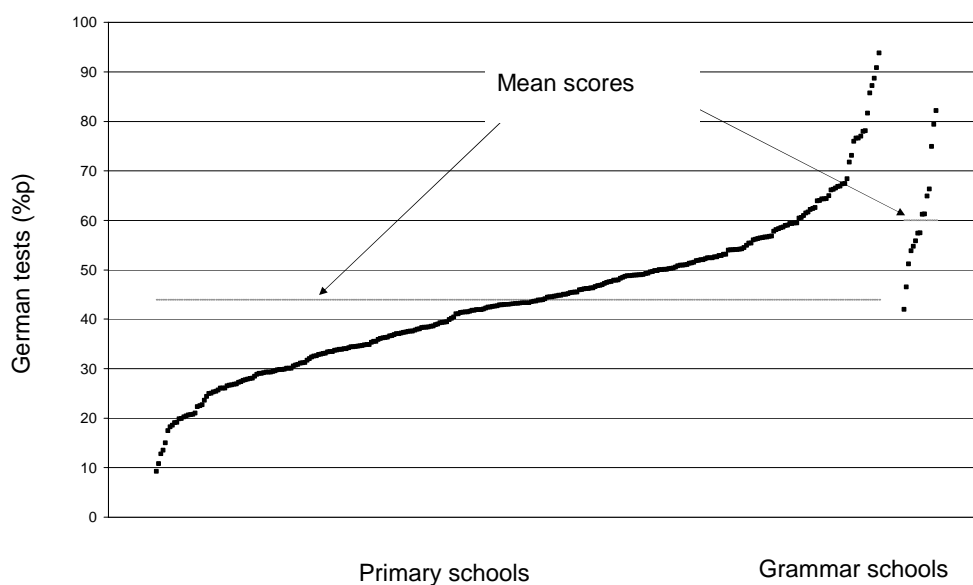

Figure 2: Distribution of year 6 German learners' groups in primary schools and grammar schools

The picture for year 10 students is more complex, as they attend three types of schools. The most prestigious and most traditional grammar schools tend to attract the more ambitious and able students intending to study in tertiary education. Comprehensive vocational schools provide a wider range of education, as they offer vocational education besides traditional secondary education. Vocational training schools cater for students with the lowest level of achievements in primary school and do not offer avenues for further education. Figure 3 and figure 4 show the distributions of all participating groups in year 10. As can be seen, students' groups are distributed along three lines with some overlaps. Significant differences characterize mean scores for the three cohorts in both languages. Learners of English in grammar schools achieved a mean score of 63.5 percent, in vocational technical (trade) schools 44.0 percent, and in comprehensive vocational schools 25.4 percent. Similarly to the younger learners, achievements are significantly lower in German than in English: the mean scores for the three types of school are 54.2, 39.7 and 23.1 percent respectively.
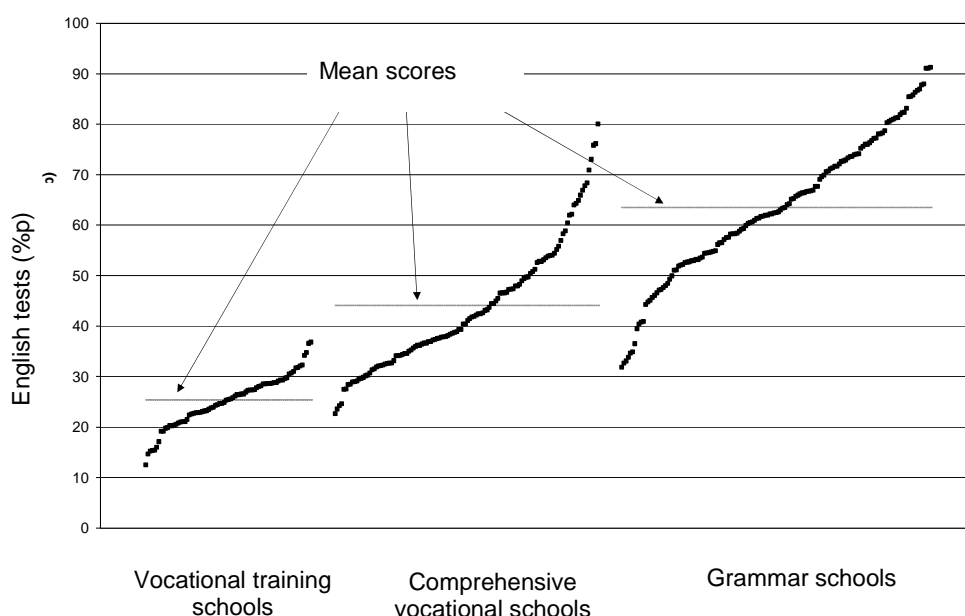

Figure 3: Distribution of year 10 students' groups studying English in three types of secondary schools
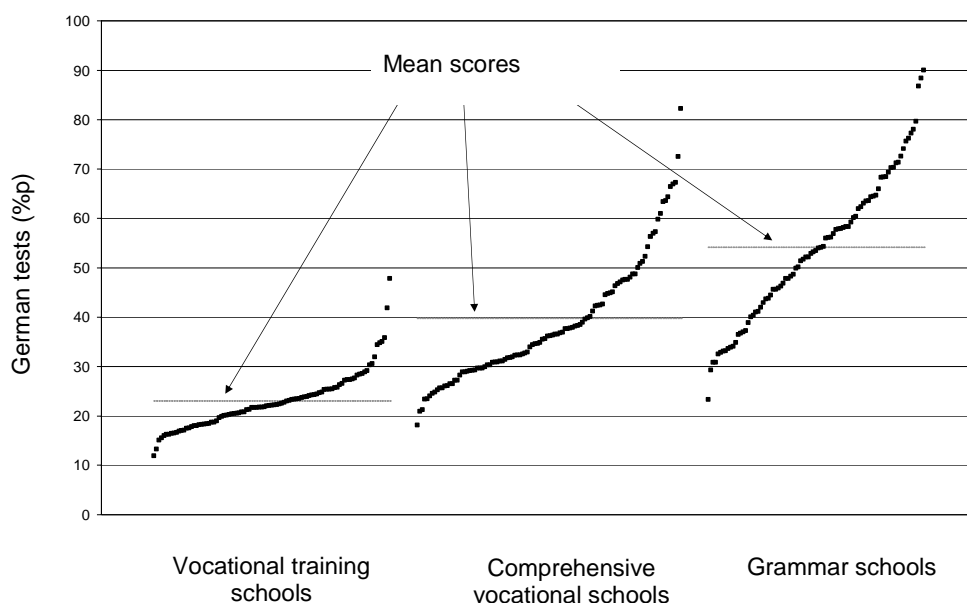
Figure 4: Distribution of year 10 students' groups studying German in three types of secondary schools

These findings are in line with previous research on school achievements in general (Csapó, 1998, 2002), and performances on language proficiency tests in English and German (Csapó, 2001; Csapó & Nikolov, 2001, 2002). Significant differences characterize learners in different types of schools; students in grammar schools tend to outperform learners in other types of institutions. However, the significant differences between achievements in the two modern foreign languages in both years indicate something different, as they are both foreign languages, none of them is related to the Hungarian language, therefore no linguistic argument supports this finding.

A comparison of boys' and girls' performances in years 6 and 10 in English and German

The relationship between learners' gender and their achievements is often examined in large-scale studies. Girls tend to achieve higher scores in all school subjects (e.g., Csapó, 1998, 2002) and in previous studies on foreign languages similar results were found (e. g., Bors, Lugossy, Nikolov, 2001; Csapó, 2001). Therefore, we expected girls to outperform boys in this study in both years and both languages.

As the data in Table 1 show, girls' achievements are significantly (p<.001) higher than those of boys for both English and German and in year 6 as well as in year 10. Similarly to the overall results, scores students achieved in English are significantly higher than in German in both years (Nikolov & Józsa, 2003). As for the three skills, girls' largest advantages are found in the writing skills in both years, whereas the differences between boys and girls are smaller in reading comprehension and listening comprehension. Boys' achievements in writing, the only

productive skill tested in the project, in German are extremely low (13%) and although girls outperformed them, their performances are also poor (21%). Standard deviations are large, indicating a wide range of performances for both genders. The largest differences are found in the writing skills in German, where standard deviation for girls is higher than for boys. In all other cases higher standard deviation characterizes boys' performances.

Table 1: The relationship between gender and performances on English and German tests in % (SD)

Skill	Gender	Year 6		Year 10	
		English	German	English	German
Reading	Girl	53 (22)	43 (19)	37 (24)	34 (22)
	Boy	51 (23)	38 (19)	36 (24)	32 (22)
Writing	Girl	66 (23)	54 (24)	36 (31)	21 (29)
	Boy	59 (26)	44 (26)	29 (31)	13 (23)
Listening	Girl	60 (20)	46 (21)	78 (17)	69 (22)
	Boy	59 (22)	43 (21)	79 (16)	67 (22)
3 skills	Girl	60 (19)	48 (18)	50 (22)	41 (21)
	Boy	56 (21)	42 (19)	47 (21)	37 (20)

These findings are similar to results of previous studies (Bors, et al., 2001, 2003; Csapó, 2001, 2002): girls tend to perform better than boys. Comparing our findings with the results of the large-scale study on English and German in years 6, 8 and 10 in the year of 2000 (Csapó, 2001) the trends are consistent: girls outperformed boys on all skills and the largest differences were found in the writing skills.

The relationship between students' achievements and the frequency of tasks in classrooms

It is important in every testing project to assure that students are familiar with the tasks and the novelty element is not a major factor influencing outcomes. As has been said, task types in the booklets were based on typical tasks used in widely spread published teaching materials and they included for example, matching tasks, gap filling, multiple choice items, and writing a short text; all were based on meaning and context. In order to examine how task familiarity and frequency of tasks in English and German classes influenced learners' performances, students were asked to indicate on a 1-5 scale how often they had opportunities to do tasks similar to the ones in the test booklets in the three skills.

Table 2 shows the distribution of respondents in percentages according to how often they claimed they practiced task types used in the proficiency tests in listening, reading, and writing. The data show clear trends: more students claim that such tasks are often or very often used in English than in German classes; and vice versa, more respondents chose never, rarely or sometimes for German than for English. As for the three skills, in harmony with classroom observations (Nikolov, 1999) and questionnaire data on frequencies of classroom activities (Nikolov, 2003), averages are consistently higher for reading and writing than for listening

comprehension tasks. Averages for listening range between 2.63 and 3.02 (Table 2). Also, all means of frequency data are higher for English than for German, therefore it is reasonable to argue that most probably higher scores in English are related to higher frequency of similar task types meaning more practice. In other words, learners of English may benefit more from classroom practice, as their experiences are more relevant to the construct of the proficiency tests. Standard deviations range between 1.06 and 1.21 indicating quite a lot of variation.

Table 2: How often students claim they practice tests similar to ones in booklets

| Year | Language | Skill | Students' ratio (%) | | | | | Mean score | SD |
			Never	Rarely	Sometimes	Often	Very often		
6	English	Listening	9	23	32	28	8	3.02	1.09
		Reading		14	24	38	19	3.54	1.08
		Writing	4	10	18	36	32	3.81	1.10
	German	Listening	12	31	31	21	5	2.75	1.07
		Reading	5	15	27	36	17	3.45	1.10
		Writing	5	11	21	35	28	3.71	1.13
10	English	Listening	11	26	32	26	5	2.88	1.08
		Reading	8	19	28	35	10	3.21	1.10
		Writing	10	23	28	29	10	3.07	1.15
	German	Listening	14	34	30	18	4	2.63	1.06
		Reading	11	23	28	31	8	3.03	1.13
		Writing	14	28	24	25	10	2.90	1.21

A comparison of data in years 6 and 10 shows that tasks in all three skills are claimed to be somewhat more frequently practiced in year 6 in both English and German than in year 10. The most often applied tasks are integrated reading and writing tasks where students are to fill in words based on their meaning: 68 percent of English learners and 63 percent of German learners claim that such tasks are often or very often used in their classes (averages are 3.81 and 3.71 respectively, see Table 2).

As a next step we examine the relationship between how often participants claim they practice tasks for each skill in the classroom and their achievements on the booklets tapping into three skills. Table 3 shows the distribution of year 6 and year 10 students' mean scores according to their claims on how often they practice similar tasks with standard deviations in brackets. The tendency is clear and consistent: more practice makes students smarter in all three skills and in both languages; this is good news for students, teachers, and testers. However, standard deviations are huge within groups of frequencies indicating that there is more to language development than simply more frequent practice, as tasks need to be tuned to learners' level to be beneficial.

Extremely large F values (ranging from 13 to 182) also indicate that huge differences exist between groups reflecting strong streaming traditions in Hungarian schools. As the numbers in Table 3 show, the most striking difference and largest F value is found in year 10 for writing. Most probably teachers assume that there is a threshold in writing: students should achieve a certain proficiency level to be able

to produce texts. The ones who never ever attempt to write short guided compositions do not stand a chance in a test.

Table 3: Developmental level of language skills according to frequency of tasks in classrooms

| Year | Language | Skill | Mean score (SD) | | | | | F* |
			Never	Rarely	Some-times	Often	Very often	
6	English	Listening	55 (21)	57 (21)	61 (21)	62 (20)	64 (21)	26
		Reading	46 (23)	47 (23)	48 (23)	54 (21)	58 (21)	55
		Writing	52 (26)	55 (25)	59 (25)	65 (23)	67 (23)	53
	German	Listening	42 (21)	43 (20)	45 (21)	49 (22)	49 (25)	13
		Reading	35 (20)	36 (19)	39 (19)	43 (19)	47 (20)	42
		Writing	39 (26)	40 (25)	46 (25)	53 (24)	56 (25)	50
10	English	Listening	68 (20)	76 (18)	80 (15)	84 (14)	84 (14)	100
		Reading	23 (18)	30 (21)	36 (23)	41 (24)	46 (26)	74
		Writing	10 (20)	20 (26)	34 (30)	44 (31)	48 (33)	182
	German	Listening	59 (22)	65 (21)	71 (21)	76 (21)	82 (18)	76
		Reading	24 (15)	29 (19)	33 (22)	37 (22)	43 (27)	51
		Writing	5 (14)	10 (20)	16 (26)	28 (31)	30 (32)	113

*F scores of variance analysis significant in all cases at $p < .001$

The relationship between students' performances on tasks and their perceived level of difficulty

Besides indicating how frequently they practiced test tasks in the three skills, participants were also asked to rate the booklets according to how easy and difficult they thought they were on a 1 to 5 scale. The results are shown in Table 4. Data indicate some striking differences: learners of German found listening tasks a lot more difficult both in year 6 and 10, although the distribution of students is more balanced in year 10. Otherwise, data are in line with trends in other areas: more students of German rated all three skills difficult or very difficult than learners of English, and fewer found them easy or very easy. Mean scores range between 2.33 and 3.85, and they are consistently higher for English than for German. Although the highest numbers tend to fall into the categories of just right and easy for English, reading in year 10 is an exception: 41 percent found the reading comprehension tasks difficult to cope with. On the other hand, in year 10 the listening tasks were found easy or very easy by 67 percent. As for German, the booklets are somewhat less favourably matched with learners' levels: they tend to be placed along the difficult to just right side of the scale, especially in year 10 reading and writing. Standard deviations are comparatively high as they range between 0.98 and 1.26 indicating quite some differences. These data suggest that students placed tasks along various levels of the scale and they differed in their assessments of difficulty. As learners performed on a lower level in both years in German than English, students' views on task difficulty seem to match this overall outcome: learners of German tended to find tasks more difficult than their peers learning English. And they were indeed more difficult for them.

Table 4: How difficult students rated test tasks in booklets

Year	Language	Skill	Students' ratio (%)					Mean score	SD
			Very difficult	Difficult	Just right	Easy	Very easy		
6	English	Listening	8	23	32	26	11	3.10	1.11
		Reading	3	16	35	32	14	3.37	1.02
		Writing	2	8	25	35	29	3.81	1.01
	German	Listening	18	33	29	14	6	2.56	1.10
		Reading	7	21	38	25	8	3.07	1.04
		Writing	5	12	29	33	21	3.54	1.10
10	English	Listening	5	8	20	32	35	3.85	1.13
		Reading	16	41	29	10	3	2.42	0.98
		Writing	23	23	27	18	9	2.69	1.26
	German	Listening	12	18	28	24	18	3.18	1.25
		Reading	20	43	26	8	3	2.32	0.99
		Writing	31	28	23	13	5	2.33	1.19

A comparison of data concerning students' perceived task difficulty and their achievements on the tasks in three skills shows consistent relationships. The distributions of students' performances according to their perception of the tasks on the very difficult – very easy scale are shown in Table 5. In both languages and in both years for all three skills the easier students perceived the tasks the better they performed on them, and vice versa. This tendency is shown by the increasing mean scores in all rows of Table 5.

Table 5: Developmental level of three language skills according to perceived task difficulty

Year	Language	Skill	Mean score (SD)					F*
			Very difficult	Difficult	Just right	Easy	Very easy	
6	English	Listening	46 (19)	50 (19)	59 (19)	67 (19)	75 (19)	295
		Reading	37 (19)	42 (20)	47 (21)	58 (21)	68 (21)	263
		Writing	33 (20)	43 (23)	54 (22)	65 (21)	77 (20)	399
	German	Listening	37 (18)	41 (19)	46 (20)	56 (21)	62 (26)	119
		Reading	31 (17)	35 (17)	40 (18)	47 (20)	53 (20)	106
		Writing	29 (21)	36 (24)	44 (23)	55 (24)	64 (23)	180
10	English	Listening	55 (21)	60 (20)	73 (16)	81 (12)	87 (11)	477
		Reading	24 (15)	31 (19)	41 (24)	57 (28)	59 (32)	234
		Writing	7 (16)	18 (23)	41 (27)	58 (26)	61 (29)	761
	German	Listening	50 (20)	54 (19)	66 (19)	77 (16)	88 (14)	468
		Reading	23 (13)	29 (17)	38 (22)	55 (30)	55 (35)	200
		Writing	3 (11)	9 (17)	27 (29)	43 (31)	43 (33)	400

*F scores of variance analysis significant in all cases at p< .001

The most striking extremes are found for writing in year 10 for both English and German: the mean scores for learners of English and German finding the writing task the most challenging are 7 and 3 percent respectively, whereas their peers scoring the same tasks as very easy achieved mean scores of 61 and 43 respectively. The standard deviations are high in every category, but especially for students of English and German in year 10 rating reading and writing tasks just right

to very easy. F values, indicating large differences between groups, are higher for English in both years, and extremely high for writing skills in English in year 10. The results show that students tend to have a clear and realistic picture of what levels of challenges tasks represent for them and their perceptions are in line with how they perform on them.

To take the analyses one step further, Table 6 shows correlations between students' performances on the tests, their ratings of task difficulty, and frequencies in classroom practice. As for the relationships between achievements and perceived difficulty, correlations range between .31 and .64 indicating weak to strong relationships between these two variables. These relationships are somewhat stronger for English than for German in both years (with the exception of writing in year 10 where they are the same and listening in year 10, in case of which the correlation for German is higher), stronger in year 10 than in year 6. From among the three skills, relationships are consistently weaker in writing than for the other skills, and strongest in reading comprehension.

Table 6: Relationships between performances on tests, perceived difficulty and frequency of tasks in classes

	Year 6		Year 10	
	English	German	English	German
Correlations between achievements and perceived difficulty				
Listening	.42	.32	.55	.59
Writing	.39	.31	.42	.42
Reading	.47	.39	.64	.54
Correlations between achievements and classroom frequencies				
Listening	.13	.11	.28	.28
Writing	.18	.20	.26	.23
Reading	.19	.22	.38	.33
Correlations between perceived difficulty and classroom frequencies				
Listening	.26	.32	.36	.42
Writing	.31	.35	.45	.48
Reading	.30	.33	.27	.27

Correlations between students' performances and frequencies reflect very weak to modest relationships (.11 to .38); they are lower in year 6 than in year 10, and similar in the two languages. As to correlations between perceived task difficulty and claimed frequencies in classrooms, relationships range from weak to modest (.26 to .48), and tend to be stronger for German than English.

Relationships between time variables and outcomes

So far we have explored variables related to skills, tasks and students' perceptions of frequencies and difficulty, and students' performances. In this section our concern is the relationship between variables related to time: how the length of language study in years and the frequency of weekly language classes influence outcomes. In the questionnaire students were asked to indicate how many years they

had learnt English or German and in how many classes a week. Correlations in Table 7 show relationships between the length of language study in years and students performances on the tests in three skills in years 6 and 10. The figures suggest similarly weak relationships in both years, both languages and across the three skills. One would have expected stronger relationships, as there are certain patterns characterizing learners in this respect. Although compulsory foreign language study starts in year 4, over half of the population starts learning a foreign language in lower primary years (Vágó, 2005). This means that some learners in year 6 have learnt English or German for 6 years, while others for 5, 4 or 3 years.

The situation in year 10 is even more complex, as the longest period of language study is 10 years, whereas the shortest may be 2 years only. In the latter case students started the target language in year 9, most probably after having learnt another language in primary school. A word of caution is in place here. Although the *National Core Curriculum* (2003) prescribes a foreign language for every learner from grade 4 (age 10), in fact 10 percent of the population does not start any foreign language, and even in year 9, 6 percent learn no foreign language, as they are exempted (Vágó, 2005). The low correlations in Table 7 mean that the longer one studies English or German, the better they perform on the booklets is a weak rule. These low correlations mean serious criticism towards the efficiency of programmes these students attend. Also, one would have expected a stronger relationship between years of study and scores on tests in listening comprehension, as its development is the primary goal of early years. This does not seem to be the case. As a next step, let us have a closer look at data according to the number of years students have been exposed to English and German and their performances on the tests.

Table 7: Correlations between performances and years of foreign language study

Skills	Year 6		Year 10	
	English	German	English	German
Reading	.24	.29	.23	.23
Writing	.21	.19	.22	.19
Listening	.23	.22	.25	.25
3 skills	.25	.27	.26	.25

Table 8 shows the distribution of year 6 students in percentages according to the length of their language study in years. As can be seen, the distributions are similar for the two languages: 45 percent of learners studying English and German learnt the target language for 3 years, the compulsory length of time. Eight percent of the learners had fewer years of exposure than mandatory, whereas 47 percent of the 12-year-olds learnt English or German for more years than expected, with little variation in the groups studying for 4, or more years. It is interesting to note that 4 percent of the students learning English and 2 percent of German learners started their language experience in kindergarten years. These figures show that schools launch foreign language programmes 1 to 3 years before the compulsory start (which is year 4). The reason is simple: all schools are sponsored per capita, therefore the more students (parents) they manage to attract the more financial support

they get. As early foreign language programmes are highly appreciated by parents when they decide where their child will attend school, they offer modern foreign languages early. In addition to this general tendency, German is supported with extra funds as an ethnic minority language; therefore, more extra money is allocated to the teaching of German (Imre, 1999).

Table 8: Year 6 students' distribution according to number of years studying English or German (%)

Language	≤2	3	4	5	6	7≤
English	8	45	19	6	18	4
German	8	45	17	6	22	2

Table 9 contains data on participants' achievements in percentages (with standard deviations) in year 6 according to the number of years they have studied English and German. The numbers increase in each row between 3 and 6 years. Learners who started learning English or German in year 4 tend to achieve somewhat lower scores than their peers with 4, then 5, and 6 years of language study. This latter group contains students who started in grade 1 at age 6. Interestingly, learners who started in kindergarten (more than 7 years of study) do not show any advantage (except for reading in German), meaning that young learners do not seem to benefit from exposure to English or German before primary school. On the contrary: learners in the group with 6 years of study achieved somewhat better scores. The column with 2 years or fewer is thought provoking, as the scores are higher than for learners in the 3-year column. Most probably these students study English or German as a second foreign language. It is known from statistics, that although only one foreign language is mandatory in primary school, some students study two. These students have been found to outperform their peers, as they transfer study skills successfully and they tend to have better cognitive skills (grade point averages) (Bors, Nikolov, Pércsich & Szabó, 1999).

Table 9: Year 6 students' performances on language tests in relation to years of language study (%)

Language	Skill	Years of language study					
		≤2	3	4	5	6	7≤
English	Reading	49 (25)	45 (20)	56 (22)	58 (22)	60 (22)	58 (23)
	Writing	58 (28)	57 (23)	68 (23)	70 (23)	70 (23)	68 (25)
	Listening	56 (24)	55 (19)	64 (20)	63 (22)	68 (20)	66 (21)
	3 skills	54 (24)	52 (18)	62 (19)	64 (20)	66 (19)	64 (21)
German	Reading	36 (19)	36 (18)	42 (19)	44 (21)	50 (19)	52 (22)
	Writing	46 (28)	45 (25)	51 (26)	52 (26)	58 (23)	53 (26)
	Listening	45 (25)	41 (19)	44 (19)	49 (23)	53 (21)	53 (22)
	3 skills	42 (21)	41(17)	45 (18)	48 (20)	54 (18)	52 (20)

Overall, these results are consistent with the low correlations in Table 7. Performances do vary systematically with the year of study, but the relationships are modest. This is not surprising, as it is well known from research that young

learners develop very slowly, therefore it would be unrealistic to expect year 6 students after 6 years of study to achieve the double of what their peers achieve after 3 years.

A similar analysis is necessary for students in year 10 to examine the relationship between the length of study and students' achievements in English and German. Table 10 shows the distribution of participants according to the number of years they have learnt the target languages. The majority of students have studied English and German for over 6 or 7 (32 and 35 %), or more than eight years (35 and 43%, respectively). This means that the length of study in years is significantly higher for German: 82 percent have spent six or more years studying German, whereas the ratio for English is 67 percent. One would expect learners of German to be better than learners of English. As has been shown, this is not the case: achievements in English are consistently higher. Therefore, some other factors must be responsible for the differences.

Table 10: Year 10 students' distribution according to number of years studying English or German (%)

Language	≤2	3 or 4 or 5	6 or 7	8≤
English	23	10	32	35
German	10	7	39	43

To better understand the data Table 11 shows year 10 students' achievements in percentages for both English and German. Performances do increase in each row, although they do vary in the three skills. Standard deviations are high (the highest in the case of the writing skill) showing huge differences in performances. F values also show extremes in variance between groups.

Table 11: Year 10 students' performances on language tests in relation to years of language study (%)

Language	Skill	Years of language study				F*
		≤2	3-5	6-7	8≤	
English	Reading	27 (18)	34 (23)	37 (23)	43 (26)	94
	Writing	21 (25)	29 (30)	33 (31)	40 (33)	81
	Listening	71 (20)	78 (17)	79 (16)	83 (13)	110
	3 skills	40 (18)	47 (20)	49 (21)	55 (22)	117
German	Reading	23 (15)	27 (20)	31 (20)	38 (24)	64
	Writing	6 (16)	11 (23)	16 (26)	21 (29)	43
	Listening	56 (22)	63 (24)	67 (21)	73 (21)	72
	3 skills	28 (15)	34 (19)	38 (19)	44 (21)	76

*F scores of variance analysis significant in all cases at $p < .001$

As for the data on the frequency of weekly classes, they are visualized in figures 5-8. Each dot in the figures represents a group of learners participating in the project. The lines show the mean scores for learners according to the number of weekly hours in which they study English or German.
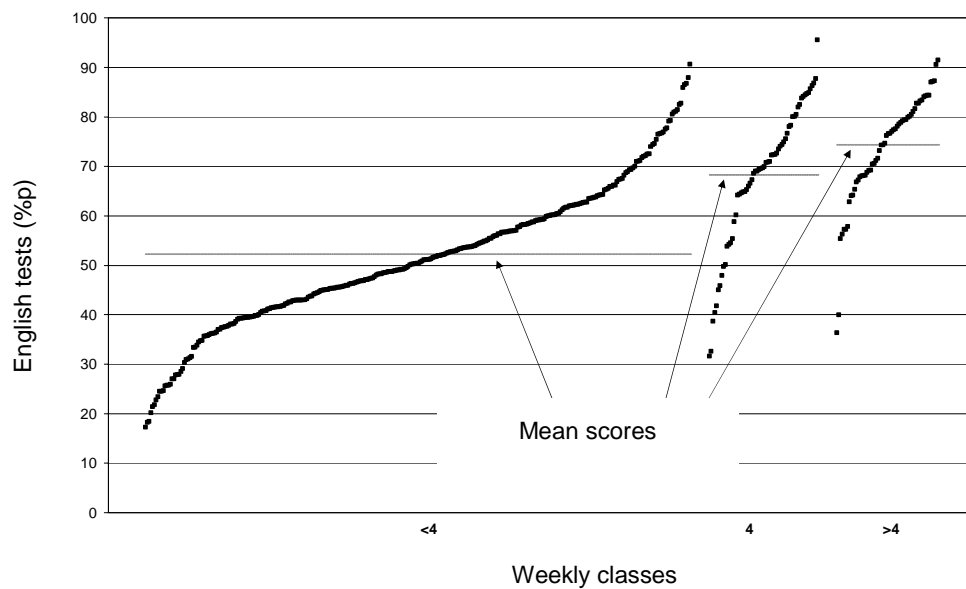
Figure 5: Year 6 students' groups learning English in fewer than four, four, and more than four classes a week
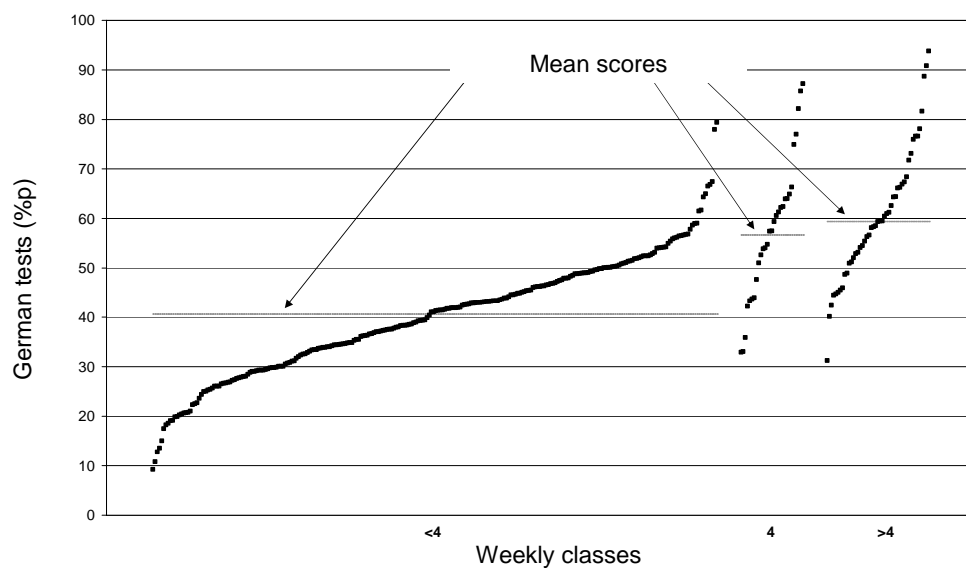

Figure 6: Year 6 students' groups learning German in fewer than four, four, and more than four classes a week
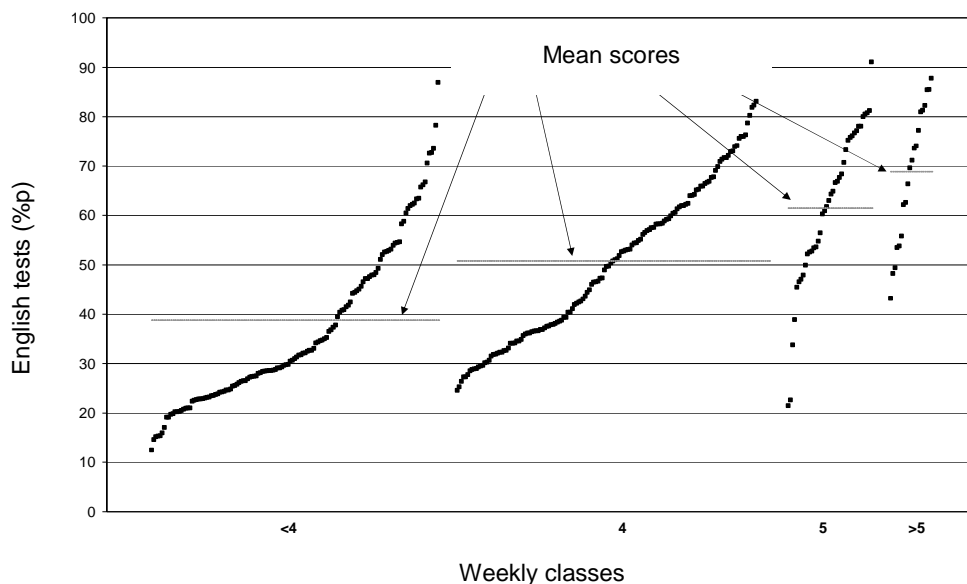
Figure 7: Year 10 students' groups learning English in fewer than four, four, six, and more than six classes a week
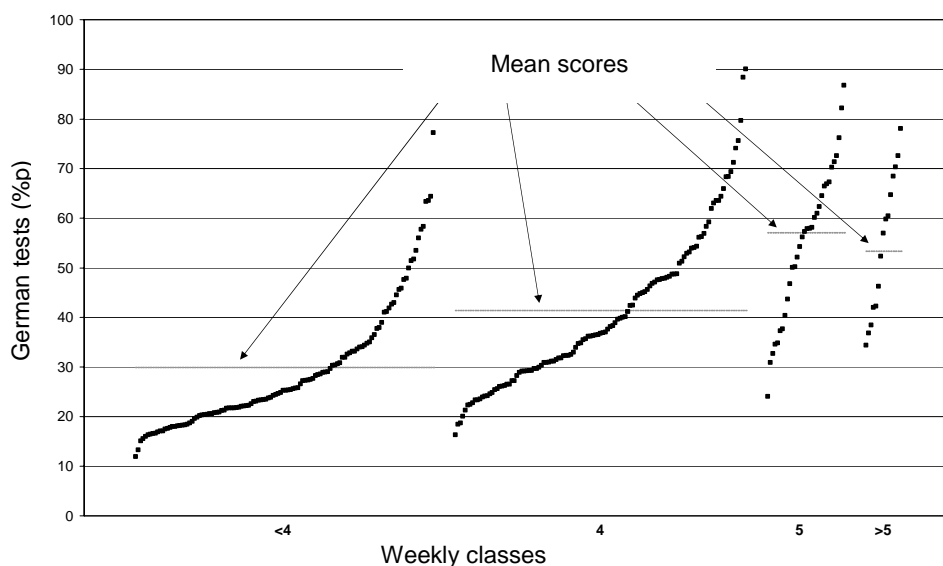

Figure 8: Year 10 students' groups learning German in fewer than four, four, six, and more than six classes a week

Students in year 6 are categorized in three groups, whereas learners in year 10 are put in four different categories according to the weekly hours. The trend is obvious in all the four figures: the more frequent the weekly exposure the higher the mean scores in both years and both languages, although overlaps across

groups are large. There is however a remarkable difference in figure 8: the mean score of learners of German in more than 6 hours a week is on a lower level than that of their peers getting instruction in 6 classes. This is a remarkable and puzzling finding, but it is impossible to find out why this is the case. Most probably students' lower level of motivation and aptitude might be responsible for the phenomenon and the quality of teaching might also play a role.

To sum up the relationships between time in years, weekly classes, and students' performances, a systematic but weak relationship characterizes them in both languages and in both years. There is one exception to the rule: in year 10 students studying German in the most intensive courses performed on a lower level than their peers learning German in fewer classes.

What is the relationship between private tuition and performances on tests?

Extracurricular activities related to language study are assumed to contribute to students' achievements in schools, as this is the reason why parents pay for private tuition. In order to explore the correspondence between data on such support and achievements, students were asked to indicate if they attended private classes. In year 6, 30 percent of students learning English, and 23 percent of learners of German claimed to attend private language classes. The ratios for year 10 students are 27 and 24 percent, respectively. These percentages are similar for the two target languages and the two years.

In Table 12 participants' performances in percentages are categorized according to their answers to the question whether they got private tuition. In year 6 the contribution of private language classes is significant in English in all three skills, although the differences are small. However, this is not the case with learners of German: students' performances are very similar in the two categories. Students in year 10 are indeed on a higher level in case they attend private language classes and this is true for both languages. The differences are large and significant between the two groups. Standard deviations are bigger in the case of the older age groups and the most obvious contribution to performances is found in learners' writing skills – the only productive skill assessed in the project. The mean score for learners of English in year 10 with extra classes is 43 percent, whereas their peers with no extracurricular tuition averaged 28 percent. The relationship between the achievements of the two groups of German learners is similar: students with special support achieved a mean score of 27 percent, without private tuition 14 percent. Most probably the value of private classes lies in tuning scaffolding to learners' individual needs, while in schools teachers focus on groups' needs rather than the individual. Also, there must be a lot of individual variation in what students meant by private tuition. Some may have attended private classes regularly over years, whereas others may apply for help to get support with problematic homework or before high-stakes tests.

Table 12: Relationships between performances on tests and private tuition

Skills	Private tuition	Year 6		Year 10	
		English	German	English	German
Reading	Yes	55 (22)	42 (20)	44 (24)	41 (23)
	No	50 (22)	41 (19)	34 (23)	31 (21)
Writing	Yes	65 (24)	49 (26)	43 (30)	27 (31)
	No	61 (24)	49 (26)	28 (30)	14 (24)
Listening	Yes	62 (21)	44 (21)	83 (14)	75 (20)
	No	59 (21)	45 (21)	77 (17)	66 (22)
3 skills	Yes	61 (20)	45 (19)	56 (21)	48 (22)
	No	57 (20)	45 (19)	46 (21)	37 (19)

Mean scores in % (SD).

All in all, this significant contribution of one to one teaching pitched at learners' levels may show that teachers in schools find it difficult to cater for individual differences and many parents realize that more appropriately tuned support is necessary for their children to succeed. As to what happens in classrooms, research has shown that on the one hand, pair and group work tasks are infrequent in Hungarian schools and teachers apply frontal classroom work in a lockstep fashion most frequently with individual work (Nikolov, 2003; Nikolov & Ottó, 2006). Individual work, however, does not mean tasks tailored to learners' individual needs, but everyone does the same task and is expected to achieve the same results to succeed. On the other hand, despite the fact that students are streamed into ability groups for modern language classes from the earliest possible age, teachers expect them to be on the same level and face difficulties when managing learners' individual differences. Teachers assume that groups should be homogeneous so that they can develop them at the same pace (Nikolov, 2002).

Participants' long term goals for language study

Goal setting is a decisive factor in foreign language study and it may determine how students persist in the effort to become more proficient (Dörnyei, 2001). Participants were asked to indicate in the questionnaire what their long-term goals were with the language they studied. They were offered six options reflecting typical instrumental motives Hungarian students tend to identify with: five options indicate positive documented achievements along a prestige scale, whereas one is negatively worded. In the Hungarian context the school-leaving examination is available to everyone but is not seen as a major indicator of language proficiency, though a recent school-leaving examination reform is supposed to change the situation (Fekete, Major & Nikolov, 1999). Intermediate and advanced level external exams, however, are highly appreciated in society and schools, students, parents, and the job market value them. Degrees in the target language reflect not only a pragmatic value (e. g., student may find better job opportunities) but also intrinsic motivation (i.e. they enjoy the actual study of the language) on part of the student: a 4- or 5-year university degree in English and German is considered as a

top achievement in language study, whereas a college degree is somewhat less prestigious.

The figures in Table 13 show the distributions of students in percentages according to what plans they identified with. The data reflect different patterns for the two target languages: learners of English tend to strive higher, as the ratio of students aiming for external proficiency exams is larger. In year 6, 62 percent of English learners and 51 percent of German learners set such an exam (intermediate or advanced) as a goal for themselves, whereas the ratios for year 10 students are 66 and 52 percent respectively. Younger learners tend to be more optimistic, as more of them identified with the advanced degree than the older learners. In other words, in year 10 students seem to see options more realistically. As for the ratio of students aiming for a school-leaving exam, more learners of German chose this option and these ratios are very similar in the two age groups. The highest level of devotion to language study is represented by students aiming to get a college or university degree in English and German. The percentages indicate younger learners' higher level of optimism, and learners of English seem to be slightly more ambitious. An equal level of interest is reflected in year 10 for both languages.

Table 13: Students' distributions according to their plans to continue language study (%)

Plans with language study	Year 6		Year 10	
	English	German	English	German
Give it up as soon as possible	10	16	10	18
Take school-leaving examination	18	25	19	26
Take proficiency exam at intermediate level	34	34	47	42
Take proficiency exam at advanced level	28	17	19	10
Get a college degree	4	4	2	2
Get a university degree	6	4	2	2

Looking at the negative end of the scale of goals, the ratio of students wishing to give up language study as soon as possible is worryingly high in both years and a significantly less favourable picture emerges for German than for English. In years 6 and 10, one out of ten learners of English set no meaningful goals for themselves, whereas 16 percent of year 6 and 18 percent of year 10 students want to quit German programmes as soon as possible. This is a very high ratio indicating low level of self-esteem in foreign language study and a threat to the aim of learning foreign languages set for state education in national and European language policy documents. To become a plurilingual individual is definitely an unrealistic goal for these students.

The higher ratio of German learners in this negative category is controversial: most probably many of these students want to give up German not only because they do not experience success, but since they would have liked to study English, therefore, they lack motivation. This seems to be a recent but increasing problem in Hungarian schools: as language teachers are tenured in their posts and many

language teachers were retrained to teach English and German in the 1990s, schools are cornered. The overwhelming enthusiasm towards English (Dörnyei, Csizér & Németh, 2006) forces them to introduce criteria along which they select students they place in English groups, whereas the others are "placed" in German groups. The criteria they apply include students' grade point average and/or test results on cognitive skills, reflecting their learning abilities and motivation. This vicious circle may be responsible for lower levels of achievement and higher ratios of frustrated learners of German. Learners of German are less able, less motivated, and their socioeconomic status is less favourable than those of learners of English (Csapó & Nikolov, 2002).

These findings on learners' language learning goals are in line with results of a previous large-scale study. In the year of 2000, very similar patterns were identified for representative samples of year 6, 8, and 10 learners of English and German with the help of the same data collection instrument (Nikolov, 2003, p. 65-67). However, there are some differences. In the present study the ratio of students wanting to give up language learning is higher in both languages (in the dataset in 2000 the row contained 12, 14, 6, and 12 %), whereas the ratio of students aiming for a higher education degree is about half of ratios found in 2000. Students have become more optimistic over three years in their goals of proficiency exams in year 6, but aspirations remained the same in year 10. In other words, a slight shift can be identified in the extreme categories: more students want to give up language study and fewer want to major in languages in tertiary education.

This negative shift is extremely important in the case of unsuccessful students, and classroom research would be necessary to explore the trend in depth. However, the lower interest in language degrees is understandable. On the one hand, speaking a foreign language has recently become a tool and a must in studying other content areas in tertiary education and for finding good job opportunities. The most proficient users of foreign languages take up other challenges (e. g., business, law, computer, medical studies) and this is a favourable trend. On the other hand, a degree in English or German is less often seen as a prestigious one, as there are hardly any teaching posts and they do not open up valuable avenues for other jobs.

Table 14: Students' performances according to their plans to study English or German (%)

Plans with language study	Year 6		Year 10	
	English	German	English	German
Give it up as soon as possible	37 (16)	31 (14)	26 (11)	24 (11)
Take school-leaving examination	46 (17)	38 (16)	34 (12)	30 (12)
Take proficiency exam at intermediate level	58 (18)	48 (16)	52 (18)	45 (19)
Take proficiency exam at advanced level	71 (17)	59 (17)	67 (19)	64 (21)
Get a college degree	64 (18)	51 (19)	58 (19)	52 (21)
Get a university degree	65 (19)	56 (19)	66 (23)	61 (26)
F*	446	256	530	398

To compare how students' language learning goals and their performances relate to one another let us take a closer look at data in Table 14. The figures show the distributions of students' achievements in percentages (with standard deviations in brackets) along their plans for their future language study. Systematic patterns can be seen: the more successful the learners, the higher their goals are set in both languages. Within the larger category of tertiary degrees, a college degree is envisaged by lower achievers, and a university degree is aimed for by higher achievers. The highest performances fall into the advanced proficiency exam category in both languages and both years. The lowest achievers wish they could give up language learning as soon as possible, and obviously their reason must be related to their low levels of achievement.

These data support the claim that Gardner's (1985; Gardner & MacIntyre, 1993) model of language learning motivation misses an important link: the interaction between language learning success and motivated behaviour. The achievements of learners in the most pessimistic category are the lowest and standard deviations are also the smallest in these groups. F values are huge for both languages in both years, and higher for English than for German.

The role of students' socioeconomic status (SES) as reflected by their parents' level of education

So far we have examined how variables related to the learning of foreign languages at schools contribute to students' achievements. As is known from large-scale studies (Andor, 2000; Csapó, 1998, 2002; Nikolov & Józsa, 2003), students' progress at school is related not only to what they are exposed to in their classes, but also to their socioeconomic status. In studies enquiring into the relationships between Hungarian students' school achievements and their socioeconomic status the most widely used indicator is parental education. This is what we explore in this part of the paper.

Table 15 shows the distribution of students according to their mothers' and fathers' level of education, as well as their mean scores and standard deviations. Parents' educational levels range from 8-year primary schools, to vocational trade schools, secondary schools, college and university degrees. The figures show that the ratio of learners with more educated parents is higher for English than for German. To put it differently: the higher the socioeconomic status of the learners the more probable it is that they study English, and vice versa: more of the less educated parents' children study German than English both in year 6 and 10. The ratio of children learning English whose parents graduated from university is the double of those learning German. The ratio of English learners with parents in the three more educated groups is higher in both years, whereas the ratio of German learners with parents in the least educated groups is higher. Therefore, more learners of English come from more favourable family backgrounds, and more learners of German may get less support from home.

Table 15: Distribution of participants according to their parents' level of education (SES) with their mean scores (%) and standard deviation

| Level of education | Mother | | | | Father | | | |
| | Year 6 | | Year 10 | | Year 6 | | Year 10 | |
	E	G	E	G	E	G	E	G
8-year primary	15.3	24.8	11.8	16.6	10.3	16.4	8.0	9.3
Skilled worker	24.4	28.9	22.8	30.9	40.2	48.1	38.6	49.3
School-leaving exam	33.9	30.0	36.8	34.8	26.9	21.5	29.1	26.3
College	17.5	11.9	19.8	13.6	11.2	7.6	12.5	8.4
University	8.8	4.4	8.7	4.1	11.4	6.4	11.7	6.6
Mean score	2.80	2.42	2.91	2.58	2.73	2.39	2.81	2.54
SD	1.16	1.12	1.14	1.05	1.15	1.05	1.13	1.00

E: English; G: German

To allow us to examine the relationships between students' achievements on the proficiency tests and their school achievements indicated by their end of term grades on a 1 to 5 scale, Tables 16 and 17 show data according to parents' educational categories for years 6 and 10.

Table 16: Students' mean scores on proficiency tests (%) and their school grades (on 1 to 5 scale) according to their mothers' level of education (SES) in year 6

Mean scores		8-year primary	Skilled worker	School-leaving exam	College	University
English	Proficiency tests	40.6	50.5	61.4	68.7	75.7
	School grade	3.0	3.5	3.9	4.2	4.4
German	Proficiency tests	33.8	43.1	49.6	56.3	63.2
	School grade	2.8	3.3	3.8	4.2	4.2

Table 17: Students' mean scores on proficiency tests (%) and their school grades (on 1 to 5 scale) according to their mothers' level of education (SES) in year 10

Mean scores		8-year primary	Skilled worker	School-leaving exam	College	University
English	Proficiency tests	31.3	38.7	51.3	59.7	66.8
	School grade	3.0	3.1	3.5	3.7	4.0
German	Proficiency tests	28.1	33.2	43.5	51.4	60.4
	School grade	2.8	2.9	3.2	3.5	3.8

The relationships are systematic: the higher the educational level of the parent, the higher the figures are for both languages, for achievements on the proficiency tests and school grades in both years. There are huge differences between mean scores achieved by the two extreme groups: children of parents with university degrees tended to achieve twice the mean of learners of parents with 8-year primary school education: for example, in year 10 learners of English whose parents graduated from university achieved a mean score of 66.8, whereas the offspring of the least educated group scored 31.3 percent. The same tables also show that school grades in German tend to be lower than in English even within the same educational categories. These trends must mean something beyond the scope of

our study, as they must be related to attitudinal, motivational, and methodology issues.

Finally, correlations between students' socioeconomic status and some other factors are worth exploring. Table 18 shows correlations between students' parental educational levels and their achievements on proficiency tests by skills, their school grades reflecting school achievements, the number of weekly classes and the length of language study in years. On the one hand, all correlations are higher for English than for German in both years. On the other, all relationships are significant: strong relationships characterize learners' socioeconomic status and their performances on the tests, correlations are somewhat lower for school grades, and finally, weak but significant relationships characterize both the number of weekly classes and the length of language study in both years and for both English and German. Interestingly, the correlations for German are somewhat higher in the two latter time-related categories in year 6, and the correlation is very low for how many years students learnt German in year 10. These relationships mean that the more educated the parents the more probably their children study English or German in more intensive courses for a longer time.

Table 18: Correlations between parents' level of education (SES) and other variables

Variables	Year 6		Year 10	
	English	German	English	German
Reading	.53	.46	.50	.43
Writing	.45	.32	.47	.40
Listening	.49	.40	.48	.36
3 skills	.55	.47	.54	.45
School grade	.41	.41	.29	.24
Weekly classes	.22	.24	.21	.18
Years of language study	.16	.21	.17	.06

All correlations are significant p<0.01

Piecing the puzzle together: the relationships between various classroom variables and outcomes

So far we have looked into relationships among different variables. In order to see the larger picture we need to put these relationships into a model. Table 19 shows correlations between variables and students' performances on the test booklets. In the background questionnaire students were asked to provide data on their school grades in the target language, as well as their parents' highest level of education. This latter variable has been found as a powerful predictor of students' success in education in general, and in language learning in particular in Hungary (Csapó, 1998, 2001, 2002; Csapó & Nikolov, 2003; Nikolov & Józsa, 2003). As the figures in Table 19 illustrate, the strongest relationships are found between school grades and achievements. This is good news, as it means that the tests tapped into similar skills students develop in classes and teachers' assessments are in line with achievements on proficiency tests. However, these correlations are not very high and they are lower for German than for English. As has been shown in connection

with task familiarity and frequencies of classroom tasks, in the long run it would be necessary to ensure that classroom practice and assessment are more similar to external assessment.

The relationship between students' socio-economic status represented by their parents' level of education is quite remarkable. The correlations are the same in both years but significantly higher for English. The figures suggest a stronger relationship in the present study than in the previous large-scale project implemented in 2000 (Csapó, 2001, p. 33). The correlations between achievements and the number of weekly classes are modest (.39 - .43), and the ones between years of language study even lower (.25 - .29). Data on private tuition indicate weak relationships in year 10, and hardly any in year 6.

Table 19: Correlations between performances and other variables

Variables	Year 6		Year 10	
	English	German	English	German
School grade in language	.61	.58	.56	.49
Weekly classes	.42	.39	.43	.42
Years of language study	.29	.28	.26	.25
Private tuition	.09	.00ns	.22	.22
Parents' education (SES)	.54	.47	.54	.45

Regression analyses with the same database result in interesting findings. As the figures in Table 20 show, four variables in the model explain between 31 and 44 percent of variance. Two of the variables are responsible for significant differences in the ratio of explained variance between the two languages: students' SES and the number of weekly classes. Parental education by itself explains the largest ratio of variance in the four models, and it is clearly more powerful in the case of learners of English.

Table 20: Regression analyses with developmental level of language skills as dependent variable (rβ%)

Independent variables	Year 6		Year 10	
	English	German	English	German
Parents' education (SES)	25	18	24	17
Weekly classes	13	10	14	13
Years of language study	3	3	4	4
Private tuition	ns	ns	2	2
Variance explained (%)	41	31	44	36

It is remarkable that the two time-related variables offered by schools (years of study and frequency of weekly classes) together explain less of the variance than SES by itself. The data on the length of language study is particularly disappointing from the perspective of early start programmes, as they do not seem to contribute to success in the long run in a very powerful way. Private language classes play a role in year 10 only, whereas in year 6 their contribution to the explained variance is not significant.

Finally, the variance explained leaves us with a large portion of variance unexplained. Most probably the quality of language teaching and affective and cognitive factors are responsible for the rest of variance and this is what further research needs to examine in the future. Another, equally important avenue for explorations is to find out how language teaching in individual groups counterbalances for what seems to be decisive — learners' SES — over years.

Conclusions

Learners of English outperformed their peers learning German in both years on all three measures; girls performed better than boys; and huge differences have been found between groups and types of schools. As has been shown, students' performances on English and German proficiency tests are formed by a number of variables schools can and cannot influence. The two most widely accepted quantifiable variables are the number of years students study a foreign language and the intensity of the language course, i.e. how many weekly classes are devoted to language study. In our datasets weak relationships characterize these variables with outcomes, and extracurricular tuition does not seem to play an important role in how students perform on proficiency tests. As for classroom practice, students learning English seem to get more appropriate instruction. Students' perceptions of task frequencies and difficulty are in line with their achievements in both languages and years. End of term school grades in German are typically lower in both years. Apart from these variables, participants' long-term language learning goals and their socioeconomic status have also been examined. Learners of English set themselves higher goals, their SES has been found significantly more favourable, and stronger relationships have been identified between their SES and other variables than for learners of German.

This quantitative study allows us to point towards areas where more research is necessary. Most probably, the teaching processes and outcomes are entangled with learners' attitudes, motivation, anxiety, and most importantly, the quality of learning opportunities. It is clear that students are streamed into not only school types, but also into ability groups. Within this trend more of the more able, more motivated tend to study English in more intensive and more appropriately tailored courses – no wonder they are more successful.

The limitations of the study are manifold and these problems indicate the way forward: neither teachers' not students' views have been examined on these issues; no classroom observation has been applied; and the study is cross-sectional. There is definitely room for further research.

Acknowledgements

References

Andor, M. (2000). A nyelvtudás szociális háttere [Social background to language knowledge]. *Educatio*, *9* (4), 717-728.

Bors, L., Lugossy, R., & Nikolov, M. (2001). Az angol nyelv oktatásának átfogó értékelése pécsi általános iskolákban [An overall evaluation of the teaching of English in Pécs schools]. *Iskolakultúra, 11* (4), 73-88.

Bors L., Nikolov M., Pércsich R., & Szabó G. (1999). A pécsi nyolcadik osztályosok idegen nyelvi tudásának értékelése [Assessment of year 8 students' proficiency in English in Pécs schools]. *Magyar Pedagógia, 99* (3), 289-306.

Common European framework of reference for languages: Learning, teaching, assessment. (2001). Cambridge University Press, Cambridge.

Csapó, B. (Ed.) (1998). *Az iskolai tudás* [School knowledge]. Budapest: Osiris.

Csapó, B. (Ed.) (2002). *Az iskolai műveltség* [School literacy]. Budapest: Osiris.

Csapó, B. (2001). A nyelvtanulást és a nyelvtudást befolyásoló tényezők [Variables influencing proficency and langauge learning]. *Iskolakultúra, 11* (8), 25-35.

Csapó, B., & Nikolov, M. (2002, April). The relationship between students' foreign language achievement and general thinking skills. Paper presented at the American Educational Research Association Annual Meeting, New Orleans, USA.

Dörnyei, Z. (2001). *Teaching and researching motivation.* London: Longman.

Dörnyei, Z, Czizér, K., & Németh, N. (2006). *Motivation, language attitudes and globalisation: A Hungarian perspective*. Clevedon, Avon: Multilingual Matters.

Europeans and languages (2005). http://europa.eu.int/comm/public_opinion/ archives/ebs/ebs_237.en.pdf

Fekete, H., Major, É., & Nikolov, M. (Eds.). (1999). *English language education in Hungary: A baseline study*. Budapest: British Council.

Gardner, R. C. (1985). Social psychology and second language learning: The role of attitude and motivation. London: Edward Arnold.

Gardner, R. C., & MacIntyre, P. D. (1993). A student's contributions to second-language learning. Part 2: Affective variables. *Language Teaching, 26,* 1-11.

Imre, A. (1999). Az idegennyelv-oktatás kiterjedésének hatása a nemzetiségi-nyelv-oktatásra [The impact of the spread of foreign language teaching on the teaching of ethnic languages]. In I. Vágó (Ed.), *Tartalmi változások a közoktatásban a 90-es években* [Changes in the content of state education in the 1990s] (pp. 175-202). Budapest: OKIKK Okker Kiadó.

National Core Curriculum [Nemzeti alaptanterv]. (2003). Budapest: Oktatási Minisztérium.

Nikolov, M. (1999). Classroom observation project. In H. Fekete, É. Major, & M. Nikolov (Eds.), *English language education in Hungary: A baseline study* (pp. 221-246). Budapest: British Council.

Nikolov, M. (2002, November). What do teachers of young learners claim and do? An empirical study of their claims and practices. Paper presented at the Dijete i Jezik Danas/Child and Language Today. 2nd International Conference, Osijek, Croatia.

Nikolov, M. (2003). Angolul és németül tanuló diákok nyelvtanulási attitűdje és motivációja. [Attitudes and motivation of learners of English and German]. *Iskolakultúra, 13* (8), 61-73.

Nikolov, M., & Csapó, B. (2002, April). Twelve-year-olds' attitudes towards classroom activities and their performances on tests of English and German as a foreign language. Paper presented at the American Association of Applied Linguistics Annual Conference, Salt Lake City, USA.

Nikolov, M., & Józsa, K. (2003). *Idegen nyelvi készségek fejlettsége angol és német nyelvből a 6. és 10. évfolyamon a 2002/2003-as tanévben* [Developmental levels of year 6 and 10 learners' in English and German in the 2002/2003 academic year] On-line document retrieved on 7 July, 2006. http://www.om.hu/letolt/okev/doc/orszmer2003/idegen_nyelv_beliv.pdf

Nikolov, M., & Ottó, I. (2006). A nyelvi előkészítő évfolyam [The year of intensive language learning]. *Iskolakultúra, 16* (5), 49-67.

Ottó, I. (2003). A nyelvérzék és mérése [Aptitude and its measurement]. *Alkalmazott Pszichológia, 5* (2), 57-64.

Vágó, I. (2005). Kilencedik évfolyamos diákok idegennyelv-tanulási útjai [Language learning paths of year 9 students]. Unpublished manuscript. Budapest: Országos Közoktatási Intézet Kutatási Központ.

To Write or Not To Write in the First Grade – That Is the Question

Gloria Vickov

Faculty of Philosophy, University of Split, Croatia

glorija.vickov@ffst.hr

Introduction

According to the literature related to foreign language teaching methodology, there should be no doubts concerning the benefits of early foreign language learning, particularly in the area of pronunciation. One of the issues insufficiently covered for a long time, and in practice left more or less to teachers' intuition and arbitrary evaluation, is the early introduction of writing in a foreign language, in our case, English. Whether early writing in a foreign language has a negative influence upon writing in the mother tongue, what age is considered to be optimal for acquiring initial literacy in a foreign language, whether first graders are capable of writing in a foreign language.– these are the questions which seem to be of extremely high importance for foreign language teachers in the Republic of Croatia, because three years ago, they became active participants in the process of early foreign language teaching when a foreign language, English in particular, became a compulsory school subject from the first grade of elementary school.

Answers to the stated questions, as explained by Vickov (2005), can be viewed from the perspective of two opposite schools of thought. While the first school of thought recommends postponing the introduction of writing skills to later stages of learning, primarily due to a possible interference from the mother tongue, the other advocates introducing it during the early years (at the beginning of learning a foreign language). However, both approaches seem to avoid specifying the age limit as the beginning of introducing writing, which is particularly dubious in the approaches which advocate early writing. As a matter of fact, while Croatian authors speak in most cases of early school age (Kruhan, 1999; Mađarić, 1996; Vilke, 1991), and the foreign ones use pretty much the same terms – "young school children, young learners" (Brewster, 1992; Dunn, 1984; Scott & Ytreberg, 1990; Slattery & Willis, 2001), it is not infrequent that, in their explanations, they vaguely refer to the age between seven and eleven, without stating precisely the optimal age.

As a starting point of our study we have taken the latter approach, i.e. the theories which advocate early writing in a foreign language. To be more specific, our study is based on suggestions (Dunn, 1993; Perren, 1972) supporting the age of six and seven as a possible start for learning writing in a foreign language. The basic arguments supporting early writing in a foreign language are the following: First, the nature of writing

itself, as stated by Vygotsky (1977), does not necessarily need to go through the developmental stages of the written language in the mother tongue or the developmental stages of the oral language in the foreign language. A child is ready to write as soon as his ability to memorise has reached a level at which the child can memorize the names of the letters of the alphabet, when the attention span has developed to the point that the child can concentrate for a while on something that does not interest him when cognition has matured enough to comprehend the relationship between written symbols and sounds which these stand for (Vygotsky, 1977, p. 229).

Similarly, Kuvač (2006) explains that once the child reaches the final level of phonological awareness, he is able to start learning to write. The final level of phonological awareness is characterized by the ability to present each phoneme by a corresponding grapheme, which normally happens between the age of 4 and 7. As experienced by Croatian first-grade teachers, many six- and seven-year-olds seem to have mastered the basics of reading and writing skills in the mother tongue by the time they enter the first-grade. Generally speaking, foreign language learning, as further stated by Vygotsky, (1977, p. 276), enables a child to understand the mother tongue as a special case of the language system, therefore, it enables the child to generalize the features of the mother tongue. In terms of Croatian as the mother tongue and English as the foreign language, we can conclude that this enables the child to become aware of the phonological nature of the Croatian orthography and to recognize it as being significantly different from the non-phonetic spelling in English.

Second, it is fairly possible that a child can learn two different writing codes simultaneously; this can be even less confusing than sequential learning. This has proven to be possible in Eastern Africa where six- and seven-year-old children learn to read and write in English and Devanagari or Urdu at the same time (Perren, 1972). The same has been experienced in the THRASS program (Teaching Handwriting, Reading and Spelling Skills), a new phonics approach for teaching English in Africa (Meade & Griffiths, 2006).

Third, it seems that secret languages and codes fascinate young children and they are therefore very interested in writing in the English language, which they take as another code (Dunn, 1993). Writing in English ensures written "proof" of development both for the child and his parents and may create motivating satisfaction.

Finally, experience has shown (Kruhan, 1999; Mađarić, 1996) that if one does not start working on getting acquainted with the non-phonetic spelling at the very beginning of learning English as a foreign language, later it appears to be rather difficult to relate such a non-phonetic graphic representation to the phonological picture of the Croatian orthography according to which pupils inevitably tend to create perceptions of written words in English. This very often results in forming misperceptions within the domain of phoneme-grapheme associations. According to the above stated authors, once such orthographic misperceptions have been made, it seems to be very hard, if not even impossible, to correct them at a later stage.

An empirical study on first-graders' development of writing skills

As has been pointed out at the beginning of this paper, these approaches to the initial EFL writing are the basis of our work, in which a study on a sample of first-graders has been carried out. The first-graders belonged to the generations of pupils who preceded the introduction of English as an obligatory school subject into the first grade of elementary school. The purpose of our study was, on the basis of a systematic research, to gain an insight into introducing writing with the first-graders at the beginning of their English language learning, as well as to answer the following research question: How does writing in English as a foreign language in the first grade of elementary school influence the development of writing skills in Croatian as the mother tongue?

Taking into account that from the early formation of letters to crafting an essay, writing involves perhaps more subskills than any other academic task (Levine, 1992), we would like to emphasize that the term *writing skills,* as used in this paper, refers to the initial stages of writing, characteristic of the first grade of elementary school. However, when considering Croatian as the first-graders´ mother tongue and English as their foreign language, one should bear in mind that initial writing in the two languages does not cover completely the same subskills that first-graders are supposed to acquire at the beginning of their formal schooling. Within Croatian as an obligatory school subject in the first grade of elementary school, first-graders develop not only handwriting and spelling skills, but also begin writing in cursive and gradually incorporate standards of capitalization, punctuation and orthography. On the other hand, introducing writing in English as a foreign language in the first grade primarily incorporates developing spelling skills and raising an awareness of non-phonetic writing in English.

In order to achieve the above stated goal, our study started from the hypothesis that introducing and developing writing skills with first-graders at the very beginning of learning English does not negatively influence their development of the same skills in the Croatian language.

Participants

Our study involved sixty first-graders in two groups: an experimental and a control group. The experimental group included thirty first-graders who were learning English in a private foreign language school in which they were developing all four language skills from the beginning of the school year with a syllabus of two lessons per week. Taking into consideration their acquiring initial literacy in Croatian within elementary school lessons based on the Croatian national curriculum, we can state that the first-graders from the experimental group were developing initial writing skills both in Croatian and in English at the same time. Thirty first-graders from in the control group were not learning any foreign language. Both groups were homogeneous at the onset of the study not only according to the results of the test measuring their readiness for school, but also according to the educational qualification of their parents.

Data collection

As to the school readiness test, it was delivered as a semi-structured interview (conducted by school counsellors) consisting of five subtests. Each subtest focused on particular aspects of the children's cognitive abilities. The first subtest concentrated on their grapho-motor abilities. The second subtest measured their understanding of quantities. The third one focused on reasoning on the basis of similarities and differences. The purpose of the fourth subtest was to check whether the children could distinguish colours and parts of the body. A test of lateralization was also conducted within the fourth subtest (children were supposed to show their left arm, right foot, to put their hands together and the like). The final subtest focused on their ability to analyze and synthesize sounds, which is considered to be crucial for initial reading and writing. The statistical analysis of the data concerning the school readiness test results and the parental educational background for both groups is presented in histograms (Figures 1 to 4).

In order to gain an insight into the first-graders´ initial writing skills, we used four dictations in Croatian and four dictations in English as a means of measurement. The research was conducted during the summer semester of the school year 1999/2000 (dictations being written at approximately one-month intervals). The first dictation, both Croatian and English, was written in March, 2000 and the fourth, the last one, at the end of the first grade in June, 2000. The pupils from the control group wrote four dictations in Croatian, and the experimental group wrote the same four dictations in Croatian and four dictations in English (for tasks see Appendixes). In relation to the topics and contents, the Croatian dictations were adjusted to the interests of the pupils, as well as to the Croatian curriculum for the first grade of elementary school. Therefore, each of the four Croatian dictations included language items which were supposed to be acquired by first-graders and which had been practiced as separate teaching units in both groups before the implementation of dictations (capital initial letters, punctuation marks, distinguishing between ″č″ and ″ć″ sounds and correct handwriting). The first Croatian dictation consisted of 33 words (11 monosyllables functioning as conjunctions, prepositions and auxiliaries) forming 6 short sentences and were concerned with first-graders´ initial impressions of school and letter learning. The second dictation consisted of 54 words (20 monosyllables) forming 7 sentences and reported on two boys who went on a trip by school bus. The third dictation comprised 44 words (16 monosyllables) forming 12 sentences and was about boys playing football. The fourth dictation encompassed 57 words (20 monosyllables) integrated into 9 sentences dealing with a school trip undertaken by a first-grade class and their teacher. (For all dictations see Appendix A).
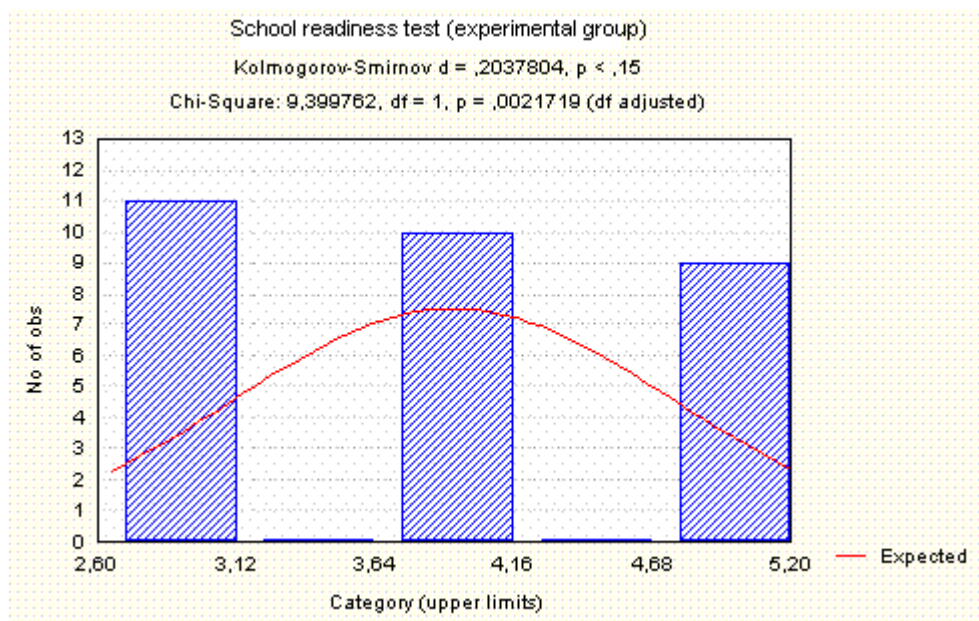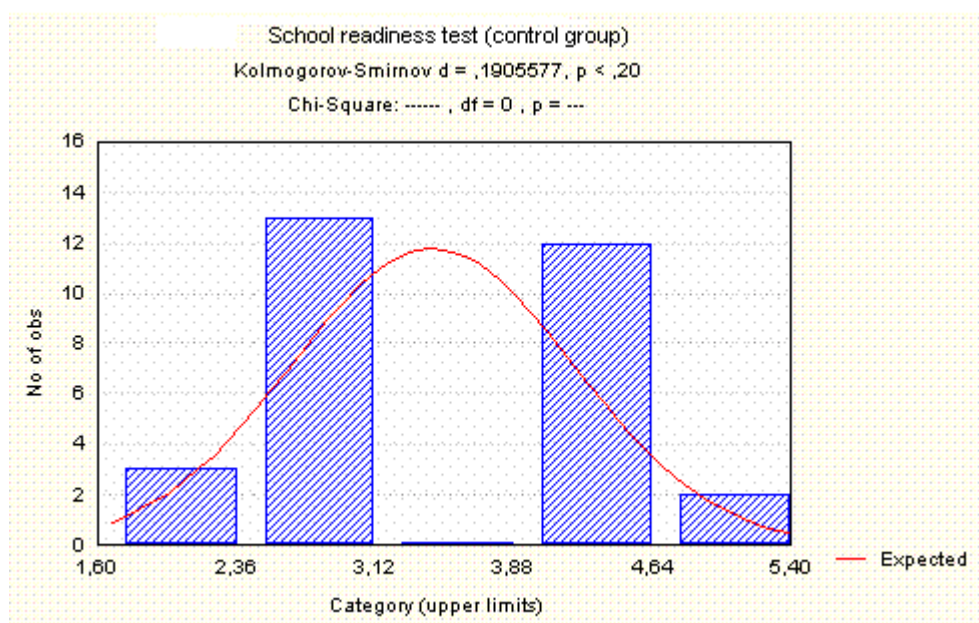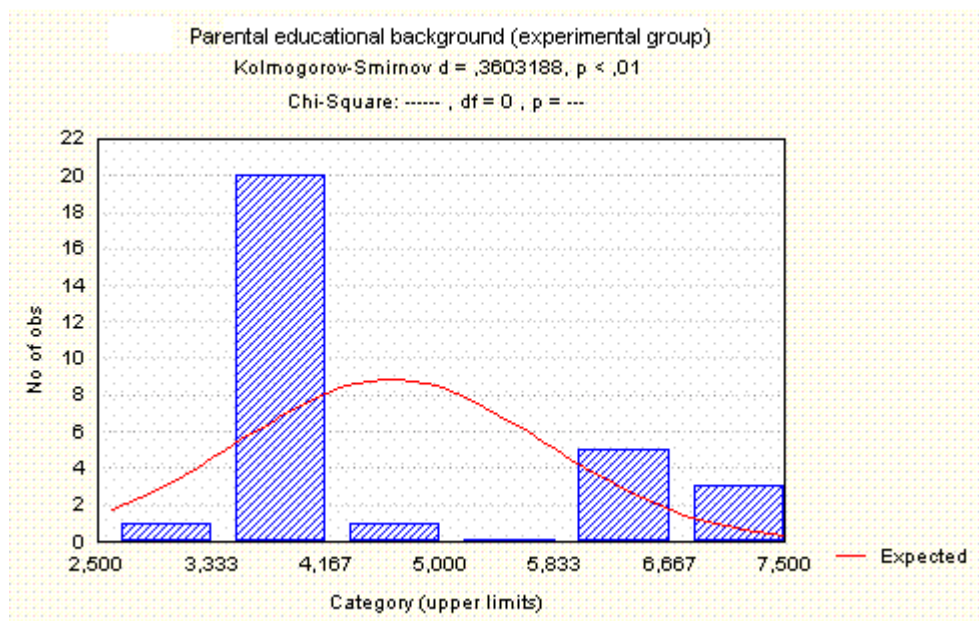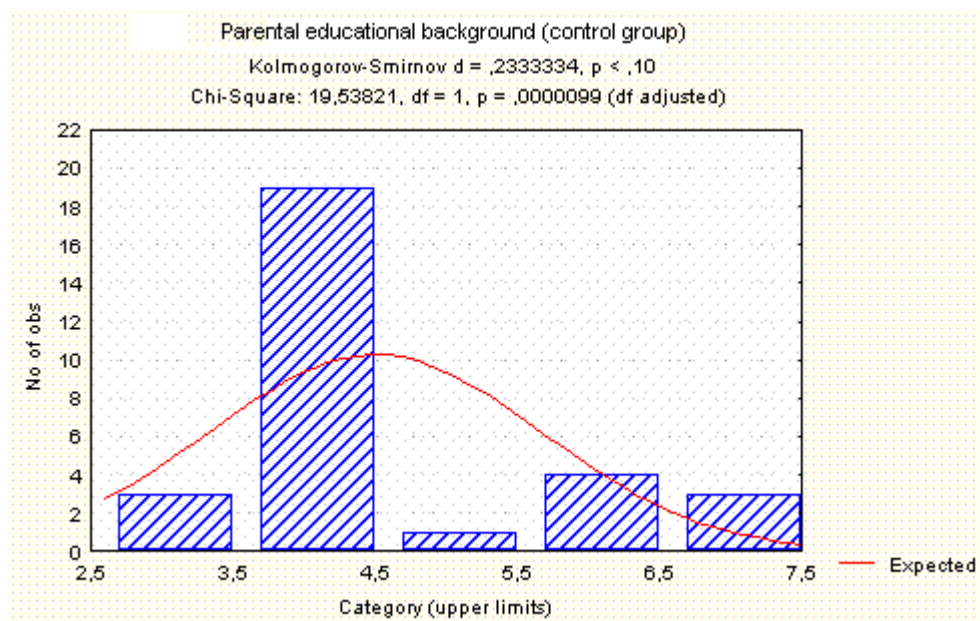
Figure 1

Figure 2

Figure 3

Figure 4

English dictations could not thematically follow the Croatian dictations, simply because the first-graders, involved in our research, were exposed to a limited amount of English vocabulary characteristic of the beginner level (for example, numbers to ten, names of colours, family members, greetings, introducing oneself). Therefore, the four English dictations consisted of isolated words and short sentences using taught English vocabulary items. The teaching material in the experimental group pupils was *I-Spy 1* (Ashworth & Clark, 1996). The first English dictation consisted of 22 words presented as 10 isolated words and 4 short sentences thematically related to pets. The second English dictation included 34 words of which 9 were presented as single words; the rest formed 7 sentences related to introducing oneself, asking for the name and age, identifying favourite colours and likes/dislikes in relation to food. The third English dictation consisted of 31 words of which 22 single words (colours, numbers and animals) and 2 sentences related to the identification of pupils' favourite number and colour. The fourth English dictation included 34 words incorporated into 9 sentences regarding food and animals (see Appendix).

Before starting to write the Croatian dictations, three teachers, who conducted the dictations, would always draw the pupils´ attention to specific orthographic features, i.e. rules which had previously been covered in detail. The dictations would then be scored according to those features. The four Croatian dictations checked to what degree the pupils had acquired a particular language feature: in the first dictation, it was correct handwriting, in the second, the accurate use of capital initial letters, in the third dictation, emphasis was on punctuation marks and the final dictation distinguished between the sounds "č" and "ć". Therefore, in each dictation teachers marked only the errors which they were focusing on. For the purpose of this study in the four Croatian dictations in each group we analysed the total amount of errors (not just those already mentioned), added them and, following the criterion "one word – one error", the ratio was calculated in relation to the maximum number of possible errors. This scoring procedure was chosen in order to compare as easily and accurately as possible the writing of Croatian dictations to the writing of English dictations (in the experimental group) and to draw as precisely as possible parallels between the experimental and control group in the writing of the Croatian dictations. In English dictations the errors were analysed only on the basis of number, with no regard to whether they were orthographic or grammatical errors. The percentage of errors in relation to the maximum number of errors possible per dictation was specified as well, following the criterion "one word – one error", like in the Croatian dictations.

Before the teachers started dictating a text, they read it out the whole text. Then, they dictated one sentence after another (or part of a sentence at a time). While dictating, the teachers endeavoured to dictate the text as clearly as possible and were careful for the pace of delivery to be neither too fast nor too slow. They also paid attention not to repeat particular words or sentences too often while the pupils were writing, otherwise the pupils´ concentration could have been disrupted. It was for this same reason that, while dictating, the teacher would stand in front of the pupils (avoiding walking around the classroom). On completing the dictation the teacher would read it all again, which provided the pupils with the opportunity to go through the entire text one more time, in order to check their writing and correct any mistakes. Due to the crucial importance of

the dictation results for our research, particular care was taken to prevent any incidence of copying. We would like to point out that due to technical limitations it was not possible for all the dictations to be conducted by one teacher. 30 pupils from the control group were taken from two different first-grade classes of an elementary school in the town of Split and were chosen according to the homogeneity with the experimental group pupils in relation to the school readiness test and parental educational background. None of the pupils in the control group learnt a foreign language. The experimental group examinees were first-graders attending elementary schools in Split as well, but they also attended a beginners´ EFL course at a private foreign language school. These pupils wrote both Croatian and English dictations under the guidance of their English teacher (who was a Croatian native speaker) in groups of ten (private foreign language school lessons are held in small-size classes). It must be emphasized that during the whole process of the investigation there was close collaboration between the three teachers, especially regarding the way in which the texts were read (for example, deliberately emphasizing the specific position of the lips when dictating words with the "č" and "ć" sounds) as well as the pace and rhythm of delivery. They also agreed on which particular parts of the dictations to repeat.

Results

Upon having assessed all the dictations, the results were subjected to the statistical analysis which included: basic statistical parameters of all the variables (arithmetic mean, minimal and maximal number of errors, standard deviation), and a presentation of error distribution in the form of histograms. Then, a Mann – Whitney test was applied for the purpose of identifying statistically significant differences among variables between the experimental and the control group as well as a tabular presentation of correlations between writing Croatian and English dictations and motivation for writing in English (Spearman Rank Order Correlations).

The variables that are included in the basic statistical analysis and the Mann-Whitney test in the Croatian dictations (in both groups) are:

1 handwriting
2 capital letters
3 punctuation
4 "č" and "ć" sounds
5 dictation 1 – C
6 dictation 2 – C
7 dictation 3 – C
8 dictation 4 – C
9 dictation C.

The first four variables denote the total amount of the errors which, as previously explained, refer to the particular language structures each single dictation was focused on. The first variable represents the total amount of errors made by the pupils in handwriting, the second variable shows the total amount of errors in the correct use of capital letters task, etc. Variables 5, 6, 7 and 8 relate to the total amount of errors in each single Croatian dictation (in chronological order, from the first to the fourth dictation). The ninth variable denotes the total amount of all the errors in all the four Croatian dictations.

The variables in the English dictations in the experimental group are as follows:

1 dictation 1 – E
2 dictation 2 – E
3 dictation 3 – E
4 dictation 4 – E
5 dictation E.

Following the pattern of the variables in the Croatian dictations, the first four variables refer to the total amount of errors in each single English dictation while the fifth variable includes the total amount of errors in all the English dictations.

The following tables (Tables 1 and 2) present the basic statistical parameters of the variables of the Croatian dictations for both experimental and control group.

Table 1: Basic statistical parameters of all the variables related to the Croatian dictations in the experimental group

Variable	Number of examinees	Mean	Minimum	Maximum	Std. Dev.
HANDWRITING	30	1.50	0.00	6.00	1.74
CAPITAL LETTERS	30	1.37	0.00	6.00	1.75
PUNCTUATION	30	0.77	0.00	3.00	0.90
"Č" & "Ć"	30	1.00	0.00	8.00	1.68
DICTATION 1 - C	30	2.10	0.00	8.00	2.41
DICTATION 2 - C	30	2.87	0.00	12.00	3.03
DICTATION 3 - C	30	2.33	0.00	11.00	2.64
DICTATION 4 - C	30	3.47	0.00	14.00	3.35
DICTATION C	30	10.76	1.00	41.00	10.01

According to the data regarding the first four variables in Table 1, the experimental group achieved very good results in acquiring the basics of Croatian orthography. The mean score which indicates the average amount of errors per dictation is 1.5 in the first variable (errors in handwriting), and in the second variable (capital letters) it amounts to 1.36. In the third variable (the total amount of errors with regard to the correct use of punctuation marks) the mean score indicates that pupils in the experimental group made only 0.7 errors on average, and in using "č" and "ć" sounds only one error. This

means that in relation to the errors regarding the language items dictations were focused on, the experimental group pupils made errors between zero and 1.5.

Table 2: Basic statistical parameters of all the variables related to the Croatian dictations in the control group

Variable	Number of examinees	Mean	Minimum	Maximum	Std. Dev.
HANDWRITING	30	3.93	0.00	12.00	3.45
CAPITAL LETTERS	30	2.97	0.00	13.00	4.01
PUNCTUATION	30	3.93	0.00	10.00	3.76
"Č" & "Ć"	30	3.27	0.00	8.00	2.43
DICTATION 1 – C	30	4.70	0.00	14.00	4.09
DICTATION 2 – C	30	7.10	0.00	27.00	7.28
DICTATION 3 – C	30	10.43	0.00	33.00	8.26
DICTATION 4 – C	30	8.53	1.00	24.00	5.73
DICTATION C	30	30.76	3.00	89.00	23.10

Similar results are found in the following four variables (dictation 1-C, dictation 2-C, dictation 3-C and dictation 4-C). The mean score per dictation ranges from 2.1 to 3.4, which implies that the pupils made approximately 2-3 errors per dictation. Taking into account the minimal amount of errors in the first eight variables, it is obvious that the experimental group had the least possible amount of errors (0) in each variable.

The maximum amount of errors is 6 both in the first (handwriting) and the second variable (capital letters), only 3 in the third variable (punctuation) and 8 in the fourth one ("č" and "ć" sounds). In the fifth, sixth, seventh and eighth variable (the dictations in chronological order with the total amount of errors, with no regard to their nature), the maximum amount of errors is, as expected, somewhat larger: in the first dictation (the fifth variable) it amounts to 8, in the second 12, in the third 11, and in the fourth dictation 14.

According to the basic statistical parameters of the same variables in the control group (Table 2), there are significant differences in all the parameters, except for the minimum amount of errors. For example, the average amount of the errors in the first eight variables ranges from 2.9 to 10.4, that is to say, from almost 3 to 10 errors. More precisely, the control group pupils had made 3.9 errors on average in handwriting, 2.9 errors in capital letters, 3.9 in punctuation and 3.2 errors in the diacritical sounds "č" and "ć".

As to the total amount of all the errors, the mean score is 4.7 in the first dictation (the fifth variable), 7.1 in the second dictation (the sixth variable), 10.4 in the third dictation (the seventh variable), and 8.5 in the fourth dictation (the eighth variable).

The maximum amount of errors per variable is also larger than in the experimental group. In the variable *handwriting*, the maximum amount of errors is 12 (twice that of the experimental group), in *capital letters* it amounts to 13 errors, in *punctuation* 10, and in "č" and "ć" 8 (as in the experimental group). In the next four variables (all the four

dictations with the total amount of all the errors, with no regard to their nature), the maximum number of errors ranges from 14 (in the first dictation) to 33 (in the third dictation). The pupils from the control group achieved the minimum amount of errors 0 in almost all the variables, except in the eighth variable (dictation 4) in which the minimum amount of errors is 1 (also achieved by the experimental group).

In terms of the basic statistical parameters, particularly arithmetic mean, we have noticed that in all the included variables there are significant differences between the two groups, specifically in the sense of better results achieved by the experimental group. By means of further statistical analysis we have found that those differences are statistically significant and thus suitable for establishing certain interpretations. The Mann-Whitney Rank Sum Test was run in order to determine if the error analysis revealed statistically significant differences in the results obtained by the two groups. The results of the test (the sum of ranks within each single variable, z, p and N values) are reported in Table 3. The results reveal that there are statistically significant differences in terms of error analysis results ($p < 0.05$). The differences are in favour of the experimental group.

Table 3: Mann Whitney test results for the experimental and control group

Variable	Sum of ranks Experimental group	Sum of ranks Control group	Z adjusted
HANDWRITING	689,50	1140,50	-3,39
CAPITAL LETTERS	846,00	984,00	-1,06
PUNCTUATION	714,00	1116,00	-3,09
"Č" and "Ć"	647,50	1182,50	-4,06
DICTATION 1 - C	708,00	1122,00	-3,10
DICTATION 2 - C	754,00	1076,00	-2,40
DICTATION 3 - C	618,00	1212,00	-4,43
DICTATION 4 - C	635,00	1195,00	-4,16

Variable	p-level	Value N Experimental group	Value N Control group
HANDWRITING	0,00	30	30
CAPITAL LETTERS	0,29	30	30
PUNCTUATION	0,00	30	30
"Č" and "Ć"	0,00	30	30
DICTATION 1 - C	0,00	30	30
DICTATION 2 - C	0,02	30	30
DICTATION 3 - C	0,00	30	30
DICTATION 4 - C	0,00	30	30

It is obvious that the differences in almost all the variables between the experimental and control group are statistically significant. The difference is not statistically significant only in the variable *Capital letters,* (p > 0.05) which has not surprised the teachers as it appears to be a very demanding orthographic category that needs to be practised and reinforced until the fourth grade of elementary school. Namely, in the first grade children perceive a sentence as a whole having its own beginning and ending relatively slowly. Besides that, while being exposed to writing dictations, children at this age, seem to concentrate more on exact transforming a sound into a letter rather than on observing orthographic rules. If we consider all the remaining variables, it is evident from the statistics that the experimental group achieved considerably better results than the control group.

Such an outcome directly confirms our assumption that introducing and developing spelling skills as one of the basics of writing in English (Levin, 1992) at the beginners´ level in the first grade of elementary school does not negatively influence acquiring and developing the writing skills (characteristic of first-grade pupils) in Croatian as the mother tongue. First, the pupils in the experimental group did not perform any worse in the Croatian dictations than the ones from the control group. Second, the results also suggest that there is a tendency on the part of the pupils from the experimental group to accomplish even better results in writing the Croatian dictations than do the control group pupils. This can be easily noticed by observing the statistical data in the following tables presenting percentages of the average number of errors in both groups in relation to the total number of errors possible. This statistical procedure was based on a "one word – one error" criterion (each individual word had to be spelled correctly or it was considered to be an error). Consequently, the total amount of errors possible in the first Croatian dictation was 33, and in the second one 54. In the third dictation, which was focused on the correct use of punctuation marks, the total amount of errors possible included also the twelve punctuation marks, so that the total number of errors possible in this dictation amounts to 56 (as all the other Croatian dictations involved only clauses of statement, punctuation marks were not taken into account as errors in those dictations). The total amount of errors possible in the fourth dictation was 57.

Table 4: Error percentage in Croatian dictations in the experimental group

Variable	Number of examinees	Mean scores	Minimum	Maximum	Standard deviation
DICTATION 1-C	30	6,36 %	0,00 %	24,24 %	7,31 %
DICTATION 2-C	30	5,31 %	0,00 %	22,22 %	5,60 %
DICTATION 3-C	30	4,16 %	0,00 %	19,64 %	4,72 %
DICTATION 4-C	30	6,08 %	0,00 %	24,56 %	5,88 %
DICTATION C	30	5,38 %	0,50 %	20,50 %	5,01 %

DICTATION %-C = average error percentage in all the four dictations

Table 5: Error percentage in Croatian dictations in the control group

Variable	Number of examinees	Mean scores	Minimum	Maximum	Standard deviation
DICTATION 1%-C	30	14,24 %	0,00 %	42,42 %	12,38 %
DICTATION 2%-C	30	13,15 %	0,00 %	50,00 %	13,48 %
DICTATION 3%-C	30	18,63 %	0,00 %	58,93 %	14,75 %
DICTATION 4%-C	30	14,97 %	1,75 %	42,11 %	10,06 %
DICTATION %-C	30	15,38 %	1,50 %	44,50 %	11,55 %

DICTATION %-C = average error percentage in all the four dictations

As can be seen in Tables 4 and 5, the average error number percentage in relation to the percentage of their total number possible (100%) ranges in the experimental group from 4,16% (in the third dictation) to 6,36% (in the first dictation). On the other hand, in the control group, this percentage ranges from 13,15% (in the second dictation) to 18,63% (in the third dictation). The data show that there is a tendency for the experimental group examinees to write better Croatian dictations (with considerably less errors) than the pupils in the control group, with regard to both Croatian spelling and the correct use of orthographic features specific to the developmental stages of Croatian writing skills for the first year of elementary school (capital letters, punctuation marks and the diacritic marks "č" and "ć").

Despite the methodological limitations of our research, as already mentioned previously, concerning the teacher factor and the differences in class size, judging by the results obtained, we can draw the conclusion that introducing and developing writing skills in English as a foreign language at an elementary level, i.e. from the very beginning of the first grade, has no negative effects whatsoever upon acquiring and developing writing skills in Croatian, the pupils' mother tongue, since the results of the experimental group proved to be better than those achieved by pupils in the control group. We would like to emphasize once again that both groups were taught in Croatian the same way, within the framework of Croatian as an obligatory school subject in the elementary school. The better results observed in pupils of the experimental group, i.e. those who were simultaneously developing writing skills in both languages, can bring one to a conclusion that early writing in English has, to a certain extent, a positive influence upon the development of basic writing skills in Croatian as well.

The experimental group was additionally exposed to writing four dictations in English. Upon having performed the statistical analysis, a quite considerable level of competence in spelling skills in English has been perceived. Following the pattern of the statistical analysis of the Croatian dictations, Table 6 reveals the basic statistical parameters of the variables related to the English dictations. As already explained, the variables refer to the total amount of errors in each single English dictation.

Table 6: Basic statistical parameters of the variables related to the English dictations in the experimental group

Variable	Number of examinees	Mean	Minimum	Maximum	Std. Dev.
DICTATION 1 -E	30	4.40000	0.000000	18.00000	4.20673
DICTATION 2 -E	30	8.03333	1.000000	28.00000	6.69783
DICTATION 3 - E	30	9.16667	0.000000	27.00000	7.08122
DICTATION 4 - E	30	6.40000	0.000000	23.00000	5.56838

The results presented in Table 6 show that the experimental group pupils had from 4 to 9 errors per dictation on average, which is, by all means, more than in the corresponding variables of the Croatian dictations (the fifth, the sixth, the seventh and the eighth variable). This does not, however, mean that these first graders are considerably weaker at writing English dictations. The point is that the Croatian and English dictations are two completely different forms of assessment in regard to the amount of words and sentences, the breadth of vocabulary and the content itself. From this it is evident that the success/failure of the English dictations on the part of the experimental group pupils cannot be discussed in terms of the average amount of errors compared with the average amount of mistakes in the corresponding variables of the Croatian dictations. Therefore, it was decided to calculate the percentage of the average amount of errors per variable in relation to the total number of errors possible, adjusting them according to the criterion "one word – one error". According to this, the total amount of errors possible in the first English dictation is 22, in the second one 34, in the third dictation 31, and in the fourth dictation 34. The data concerning the above stated percentages are presented in Table 7.

Table 7: Error percentage in the English dictations in the experimental group

Variable	Number of examinees	Mean	Minimum	Maximum	Std. Dev.
DICTATION 1%-E	30	20.00%	0.00%	81.82%	19.12%
DICTATION 2%-E	30	23.63%	2.94%	82.35%	19.70%
DICTATION 3%-E	30	29.57%	0.00%	87.10%	22.84%
DICTATION 4%-E	30	18.82%	0.00%	67.65%	16.38%
DICTATION %-E	30	23.14%	2.48%	79.34%	17.81%

Although the percentages shown in Table 7 are considerably larger than the ones in the variables of the Croatian dictations in the same group of examinees (see Table 4), we cannot stated that these pupils are bad at writing English dictations. If we consider the percentage results in each of the four English dictations separately, we can see that the percentage of the average amount of errors in relation to the total amount possible, ranges from 18% to 29%, with the percentage in the variable Dictation%-E (the total amount of errors in all the four English dictations) being 23.1%. We are inclined to think that the percentages stated above support the idea that the experimental group pupils

were relatively successful at writing English dictations, particularly, if we take into account the following. Firstly, the young learners wrote the dictations in a language that is almost completely new to them, both at an oral and written level. Secondly, while writing in English, the pupils could not rely on the sound image of a word which could then be simply phonetically transformed into writing, as they can almost always do when writing in Croatian. The third difficulty is regarding the four letters of the English alphabet which do not exist in the Croatian alphabet: q, w, x and y. As these letters have only been practices within the English lessons at the private foreign language school only twice a week, we may assume that they additionally burdened the pupils during the writing process. The English teacher reported that the pupils had frequent questions such as: "What does "y" look like? How should I join "x" to the other letters in the word?"

Finally, if we add to all this the fact that writing dictations requires the skills our first graders have only started to develop both in Croatian and English, it becomes clearer just how demanding and difficult it was for them to write dictations in English. This makes the results they achieved in our study all the more significant.

Attitudes and motivation for writing in English

Since the first-graders included in the experimental group were developing all four language skills from the beginning of the school year, we were interested in finding out how they were experiencing writing in English in terms of motivation. For this purpose we chose to conduct a structured interview which consisted of 6 questions (For the interview, see Appendix C). In answering the questions, the pupils had to choose between two possible answers. The first two questions focused on pupils' attitude towards writing in English as well as on finding out whether they preferred writing in English to writing in Croatian. By means of the rest of the questions we tried to gain an insight into the intensity of the motivation of the young learners for writing in English compared to that for writing in Croatian. Apart from the structured questions, the interview included two additional questions, the first of which focused on the cause of learners' motivation for writing in English. The second question was concentrated on the intensity of the stated motivation. The pupils were interviewed individually, in the classroom where they usually learn English. As we were dealing with the first-graders, we had to make sure that the interview was conducted in as normal an atmosphere as possible, and that nothing hindered the pupils´ inclination to cooperate and their truthful, spontaneous answers. Therefore, the interviewer recorded all the answers using a tape recorder, which is normally present and used in the classroom during English lessons, thus not being out of the ordinary. The recording was conducted as discretely as possible so that it did not disrupt the pupils´ attention.

The results of the statistical analysis of the interview are presented in the form of the variable frequency table (Table 8) designed for estimating the pupils' motivation. The variables refer to answers to six questions where each question allows for two answers of which one should be chosen. The answers (variables) in the frequency table assume

the values marked as 0 and 1 and indicate the presence of motivation for writing in English. The table further shows the number of respondents opting for particular answers and their percentage in relation to the total number of respondents (30).

Table 8: Variable frequencies for estimating motivation

	Variable value	Number of respondents	Percentage
Do you like writing in English?	0	3	10%
	1	27	90%
Do you prefer writing in English to writing in Croatian?	0	7	23.33%
	1	23	76.67%
Do you prefer writing to speaking in English?	0	24	80%
	1	6	20%
Do you prefer writing to singing songs in English?	0	21	70%
	1	9	30%
Which do you prefer in Croatian – writing or speaking?	0	17	56.67%
	1	13	43.33%
Would you like to write more or less in English?	0	4	13.33%
	1	26	86.67%

The first question in the interview was directly related to the pupils' attitude towards writing in English: *Do you like writing in English?* As can be concluded from surveying the tables, 3 first-graders gave a negative answer, marked as 0, while 27 answered affirmatively, designated as 1, which means that 90% of respondents from the experimental group claimed that they liked writing in English. Such a positive attitude towards writing in English as displayed by a great majority of the first grade pupils undoubtedly constitutes an important basis on which motivation for learning can be built and developed.

The second question: *Do you prefer writing in English to writing in Croatian?* indicates again a significant disproportion in choosing the answer. Only 7 respondents claimed that they preferred writing in Croatian (the answer marked as 0), while as many as 23 respondents said that they preferred writing in English (the answer marked as 1). The reasons for the answers as given by the latter group were looked into by posing an additional question: *What do you like about writing in English?* In this question no possible answers were suggested to the children: they gave their answers arbitrarily and in a descriptive way, sometimes giving more than one reason for their preference.

By summing up the answers the following has been established: the most frequent reason for preferring writing in English (11 pupils) is to be found in, as pupils themselves put it, in "strange" letters which do not exist in the Croatian language, accurately spelling w, y, x in their answers. (The fact that not a single pupil mentioned the letter q might be taken as indicative, and could be explained by a lower frequency of that letter in English words the children were being daily exposed to, as well as its rare occurrence in the basic vocabulary as presented to the children by the teaching materials within the English curriculum). In addition to this, in explaining their attitude, some pupils mentioned "strange" words (8 pupils), some of them claiming English words to be "different", "strange", "shorter", having double letters, that it was fun drawing them (possibly resulting from the fact that in the process of introducing new words the children were asked to draw the pertaining object), the fact that English words contained no "č" or similar letters also possibly featuring among the reasons for their preference. The reasons listed hereby seem to confirm the opinion expressed by Dunn (1993) claiming that many children at the early school age are very interested in writing in English or any foreign language for that matter, primarily because they see it as a sort of a "secret code". Becoming familiar with the "secret code" and being able to start using it may be one of the ways to satisfy children's curiosity which is one of the strongest motivating factors at this age. In the further explanation of their preference two pupils mentioned the non-phonetic nature of the English language, that is, the relation between reading and writing as being basically different from the one existing in Croatian, their mother tongue. Three pupils, however, seemed to be unable to explain why they preferred writing in English, while one girl said she liked it because it was "more difficult". Another girl claimed she liked it because it was graded much less frequently than was the case with writing in Croatian.

The third and the fourth questions were aimed at establishing to what extent our first-graders were affectively related to writing in English by comparing it to speaking or singing. As a matter of fact,in the contemporary FLT methodology the two activities, seem to be the most frequent forms of teaching in early foreign language learning.

Consequently, in answering the third question (*Do you prefer writing to speaking in English?)* 80% or the respondents opted for speaking.

A similar outcome was obtained in answers to the fourth question (*Do you prefer writing to singing songs in English?)* where 70% of the children said they preferred singing.

These answers should not come as a surprise since it has been empirically proved (Sironić-Bonefačić, 1999) that children at this age (six and seven-year olds) simply enjoy classes based on playing, movement, singing and dramatisation in which priority is given to the spoken utterance associated with real-life situations while stimulating the children's imagination and creativity. The teaching based on this methodological approach, in order to increase the degree of language learning motivation attempts to correspond to the basic components of children's psychology – curiosity, playfulness, imagination, the need for movement and self-affirmation.

Expressing a greater affinity to speaking and singing songs should be taken as a sign of warning. Namely, the writing skill demands, in general terms, a much greater concentration and the entire psycho-physical engagement on the part of the child. This is

why it may be viewed from the standpoint of seven-year-olds as being more difficult and more demanding. It can therefore be assumed that writing, unless planned in detail and carefully doled out, could very easily lead to saturation and exhaustion, thus becoming a demotivating factor.

The preference for speaking as expressed by the majority of pupils probably lies in the fact that this skill, unlike writing, has been for a long time completely mastered by children in their mother-tongue, thereby gaining the necessary affirmation in their environment Meanwhile, in that-same environment they are becoming increasingly aware of the importance of English as a ubiquitous language appearing, mostly in its spoken form, on TV and video, in their parents' business contacts, conversations with foreign tourists (numerous Croatian families are engaged in family tourism as one of the main touristic branches, especially in the area along the Adriatic coast) etc. all of which seems to be sending them a clear message how important proficiency in English is in everyday life.

Finally, the speaking skill in its numerous forms has always been definitely much better represented in the curriculum than the writing skill, thus becoming closer to children, more familiar and therefore also preferable.

The fifth question coincides with the third in its contents, differing from it only in as much as it relates to Croatian, the children's mother-tongue: *Which do you prefer in Croatian – writing or speaking?* A little more than half of the respondents (57 %) said they liked speaking better, while 43% preferred writing. We can see that in Croatian speaking has proved to be the most familiar skill, closest to the children's idea of learning language in general. In this case it is not surprising at all, especially in view of the fact that speaking is the very skill through which children develop and affirm themselves in every aspect of life since birth. Speaking is thus undoubtedly the language activity in which they feel most secure and by which they become part of the adult world in the fastest and easiest way.

However, it is evident that, when it comes to the mother tongue, there was no marked disproportion in answers favouring speaking over writing as was the case with English addressed in the third question. Here, as has been pointed out above, only a little less than half of the respondents favoured writing over speaking. We deem that a possible reason could lie in the fact that speaking Croatian does not present a particular challenge to first-graders since it is a skill entirely mastered a long time ago, practised almost at all times. On the other hand, writing as a new way of expression and thinking in the first grade of primary school could be seen as symbolising the beginning, not only of formal literacy, but of education, bringing into the child's world a new dimension of seriousness and importance of the written word in the adult world, arousing a desire for self-affirmation and recognition, at the same time satisfying their curiosity.

By drawing a parallel with the answers our respondents gave to the third question, where a great majority claimed that they preferred speaking to writing in English, we are inclined to think that initial writing in Croatian is basically different from the development of the same skill in a foreign language, i.e. English. Namely, as experienced by Croatian primary teachers, what with their families and what in the pre-school institutions, a large number of today's first-graders mastered the basics of the Croatian alphabet, both in initial writing and reading, prior to starting primary school. Since writing in

Croatian is no novelty, and since it probably makes them more confident, they do not tend to see writing in comparison to speaking as being considerably more demanding or difficult, which might be the reason why they seemed to favour it.

Initial writing in Croatian may be regarded as not significantly more demanding than speaking because of the phonetic principle almost consistently implemented in Croatian orthography which is a very reliable partner in introducing young learners into the world of writing. Here the first-graders can almost always rely on the familiar auditory image of words by simply transferring it into its written form which has proved particularly useful, for example, in writing dictations.

Our respondents could have been in favour of writing also due to additional confidence gained by being more frequently exposed to this skill in Croatian than in English within the regular curriculum, as well as by doing a lot of homework, not only in Croatian but a number of other school subjects. We assume that the crucial role played by writing in Croatian in mastering the majority of basic subjects and, in general, the children's survival in the literate community they are surrounded by, could be regarded as an additional reason for higher motivation in acquiring this skill in the mother tongue.

The sixth question (*Would you like to write more or less in English?*) was basically aimed at verifying the initial positive attitude taken by the majority of our respondents towards writing in English as expressed in answering the first three questions. From the above variable frequency tables for estimating attitudes and motivation it follows that 87 % of the respondents said that they would like to write more, thereby confirming the positive attitude expressed at the beginning of the interview.

At the end of the interview, in answering the additional, descriptive question (*What is your favourite activity in the English language class?*) the pupils mentioned several language activities, listing them in the order corresponding to their preferences. Exactly half out of the pupils put singing first, whereas four assigned the second place to writing. Writing as the favourite activity was recognised by seven pupils, two of them saying they preferred watching English cartoons on the video; one pupil opted for speaking; one girl claimed she liked playing games, while another girl gave the most prominent place to listening to audio cassettes. The favourite language activity, as mentioned by two pupils, consisted of learning the so-called *passwords*: isolated words, a new one each time, without the naming of which, as envisaged by the early language learning curriculum and the teaching materials entitled *I-Spy* (Ashworth & Clark, 1996) no English lesson could begin. One little girl, on the other hand, said she liked it best when their teacher read stories in English.

The answers pupils gave to the last question seem to be consistent with their answers to the previous questions in which they clearly showed preference for the language activities based on the spoken variants, e.g., singing. It is by no means accidental that such activities basically contain dynamic elements, such as movement and playing games, so necessary in creating and enhancing interest and concentration in the young learners. Namely, at this age the concentration and attention span are invariably short, not exceeding five to ten minutes on the average, after which saturation and boredom set in, preventing the child from actively participating in the current activity. It is precisely in this context that Vilke points out that "for a seven-year-old playing is genuine and serious work, therefore the teacher has to present most of the planned activities as a

sort of game in which the child will take an active part, both by mime, movement and a language expression" (Vilke, 1999, p. 25).

The results obtained by our interviews seem to be leading us towards a conclusion that the young learners comprised by the experimental group were very motivated to write in English which should, considering the complexity and high demands posed by this particular language skill, be duly recognised and carefully cultivated and stimulated in the process of foreign language learning lest it should be inferior to other language skills, primarily speaking (which, as experience has shown, is not a rarity), and for which the children displayed most interest and affinity, as has been confirmed by our interview.

Based on the results obtained by a statistical processing of Croatian and English dictations of the experimental group pupils, as well as the ones of the interview aimed at examining the young learners' attitudes and motivation for writing in English, coefficients have been calculated to show correlations between these variables with the aim of establishing their possible connection.

Table 9: Correlations between variables (Spearman Rank Order Correlations)

Variable	Dictation-C	Dictation-E	Motivation for writing in English
Dictation-C	1,00	0,64	-0,14
Dictation-E	0,64	1,00	-0,25
Attitudes and motivation for writing in English	-0,14	-0,25	1,00

Dictation-C = total number of errors in all four Croatian dictations
Dictation-E = total number of errors in all four English dictations

As the correlation coefficients obviously show, a significant connection has been observed only between the variables Dictation-C and Dictation-E, which in our research represent the degree of acquiring the basics of writing skill in Croatian and English, respectively. Although it might be disputable to draw conclusion about cause-effect relations based on the correlations, it only seems logical to assume that those young learners of the experimental group who were successful in Croatian dictations (making a small number of errors) will be similarly successful when it comes to English dictations. In other words, an increase in the number of errors in one language is accompanied by a proportional increase in the other (foreign) language, the correlation coefficient being 0.64. So, the writing skills (as assessed by dictation and ratio of errors) are interrelated in L1 and L2 at age 6.

At any rate, a positive response of the first-graders to writing in English, together with the considerable results achieved in English dictations, should represent a signal to their teachers to introduce writing at an elementary level, from the very beginning of the first grade, paying close attention to detailed planning and careful determining of the amount of time dedicated to writing activities. Namely, writing skills require a much higher degree of concentration and an overall effort on the part of the pupil, as well as the sitting position, which is not particularly popular with children of this age. It could,

therefore, from the viewpoint of seven-year-olds, be regarded as more difficult and more demanding. It may thus be assumed that developing writing skills, unless it has been planned in great detail and very carefully measured, could easily lead to tiredness, resulting perhaps in the loss of motivation as well. Potential dangers of inadequate evaluation and assessment of the pupils' competence and abilities in the area of writing, as well as excessively demanding and vaguely formulated tests measuring the acquired skills, should by all means be taken into account.

According to the experience the English teacher had with the experimental group pupils from our study, basic writing in a foreign language at this age should be based on copying isolated vocabulary items previously practiced orally, frequent homework assignments which should be stimulatingly evaluated, various language activities based upon visual referents (flashcards and appropriate wordcards). We should by all means bear in mind that at this age, the emphasis should be placed on perceiving and getting used to the non-phonetic nature of the English language.

In this respect, dictations have two functions: practice and checking to what degree particular language features have been mastered.

While being formally exposed for the first time to the complex skill of writing in English, young learners will find an additional source of confidence and support in their teacher who should be patient and understanding of their first insecure steps into the world of letters. The teacher should also be capable of arousing their curiosity, motivating their interest and sharing with them the joy of their first success, that is, their first words written in English.

References

Ashworth, J., & Clark, J. (1996). *I-Spy 1*, Oxford: Oxford University Press.

Brewster, J., Ellis, G., & Girard, D. (1992). *The primary English teacher´s guide*. London: Penguin Books.

Dunn, O. (1993). *Developing English with young learners*. London: Modern English Publications.

Kruhan, M. (1999). Vještine čitanja i pisanja u nastavi stranih jezika [Reading and writing skills in foreign language teaching]. In Y. Vrhovac (Ed.), *Strani jezik u osnovnoj školi* (pp. 139-157). Zagreb: Naprijed.

Kuvač, J. (2004). Uloga vrtića u razvoju jezičnih sposobnosti i predvještina [The role of Kindergarten in the development of language skills]. In L. Cvikić (Ed.), *Priručnik s radnim listićima za poučavanje hrvatskoga kao nematerinskoga jezika* [Manual for teaching Croatian as a foreign language]. Zagreb. Online: www.osi.hu/esp/rei/ Documents/Jezicak-Manualforbilingualteaching-Croatianlenguagteam.doc

Levine, M. (1992). *Developmental variation and learning disorders*. Cambridge: EPS.

Mađarić, R. (1996). Usvajanje vještina čitanja i pisanja u nastavi engleskog jezika u ranoj dobi [Developing reading and writing skills in EFL teaching at an early age]. *Strani jezici, 25* (3-4), 190-205.

Meade, M. & Griffiths, C. (2006). The THRASS synthetic phonics programme – A truly inspirational resource, issued by THRASS UK News Media Centre www.thrass.co.uk/nm.htm

Perren, G. (1972). New languages and younger children. *ELT Journal, 26* (3), 229-238.

Scott, W.A. & Ytreberg, L. H. (1990). *Teaching English to children*. New York: Longman.

Sironić-Bonefačić, N. (1999). Planiranje i programiranje. (Planning and programming) In Y. Vrhovac (Ed.), *Strani jezik u osnovnoj školi* (pp. 39-51). Zagreb: Naprijed.

Slattery, M. & Willis, J. (2001). *English for primary teachers*. Oxford: Oxford University Press.

Vickov, G. (2005). Postoji li negativan utjecaj ranoga uvođenja pisanja na stranome jeziku na pisanje na materinskome jeziku? [Is there a negative influence of introducing writing in a foreign language at an early age on writing in the mother tongue?] *Strani jezici, 34* (4), 285-302.

Vilke, M. (1991). *Vaše dijete i jezik* [Your child and language]. Zagreb: Školska knjiga.

Vilke, M. (1999) Djeca i učenje stranih jezika u našim školama [Children and foreign language learning in our schools]. In Y. Vrhovac (Ed.), *Strani jezik u osnovnoj školi* (pp. 17-29). Zagreb: Naprijed.

Vygotsky, L. (1977). *Mišljenje i govor* [Language and thought]. Beograd: Nolit.

Appendix A – Dictations in Croatian

Dictation 1
Jas am učenik prvog razreda. Rado idem u školu. U školi je uvijek veselo. Šalimo se ali I učimo. Naučili smo sva slova. Sad mogu baki napisati pismo I pročitati svoju omiljenu priču.

Dictation 2
Učenici Ines Marić I Vedran Pavić pošli su na izlet. Autobusom su putovali od Trogira do Omiša. Ines je pričala o gradiću Sinju I selu Muću. Vedran je pričao o svom psu Reksu I mačku Tomu. U Omišu su vidjeli Jasnu Grčić. Ona je došla iz Kaštel Sućurca. Svi će posjetiti kanjon rijeke Cetine.

Dictation 3
Dječaci igraju nogomet. Trče za loptom I viču. Dodaj! Pucaj! Što čekaš? Odjednom lopta udari u prozor. Tko je kriv? Tko je razbio staklo? Što li će reći roditelji? Kako će platiti učinjenu štetu? Hoće li biti kažnjeni? Šutjeli su i gledali se nijemo.

Dictation 4
Osvanuo je sunčan dan. Učiteljica nas je pozvala na izlet. Autobus nas je čekao ispred škole. Prešli smo preko mostića do bistrog potočića. Livada je bila prepuna raznobojnog cvijeća. S drveća se čuo cvrkut ptičica. Pod čempresom smo pojeli doručak. Dječaci su igrali igraničara, a djevojčice pravile vjenčiće od tratinčica. Kad je pala noć, krenuli smo kućama.

Appendix B – Dictations in English

Dictation 1

girl	boy
car	bed
tree	house
apple	egg
ice-cream	umbrella

Have you got a pet?
Yes, I have.
No. I haven't.
Hello!

Dictation 2:
What´s your name? My name is Marina. How old are you? I'm seven. My favourite colour is blue.
lion, horse, elephant, frog
apple, an ice-cream, pizza, chips, egg

I like bananas. I don't like pears.

Dictation 3:
What colour is it?
blue, red, green, yellow, orange, pink, purple, black
What's your favourite number?
one, two, three, four, five, six
kite, car, book, bed, dog, cat, moon, bin, sun

Dictation 4:
What do you like? I like bananas and apples. Do you like pears? No. What's your favourite food? My favourite food is pizza. Do you like animals? Yes. I like dogs, cats and birds.

Appendix C – Interview questions

1: Do you like writing in English?
2: Do you prefer writing in English to writing in Croatian?
3: Do you prefer writing to speaking in English?
4: Do you prefer writing to singing songs in English?
5: Which do you prefer in Croatian – writing or speaking?

Two additional questions:

1: What do you like about writing in English?
2: What is your favourite activity in the English language class?

Anchors Aweigh! An Analysis of the Impact of Anchor Item's Number and Difficulty Range on Item Difficulty Calibrations

Gábor Szabó

University of Pécs, Hungary

gszabo@btk.pte.hu

Introduction

Item Response Theory (IRT) is often regarded as a useful tool for achieving goals in assessment which would otherwise—on the basis of Classical Test Theory (CTT)—be nearly or even entirely impossible to reach. One such goal is the construction of item banks containing items calibrated on a common difficulty scale, making it possible to accomplish a wide variety of assessment tasks from the administering of equated test versions to computer adaptive testing (Baker, 1997, p. 50).

In the process of item bank building it seems crucial to guarantee that the new items added to the bank are linked to the already calibrated ones. This can be achieved by means of anchoring procedures, which are based on the inclusion of calibrated anchor items in the test containing the new items (Hambleton, Swaminathan & Rogers, 1991, p. 128). While the literature provides detailed descriptions of the actual procedures, very little—if any—advice is offered concerning the number of anchor items employed.

The present paper intends to investigate the number of anchor items necessary for successful anchoring on the basis of data gathered from a language test item bank building project carried out at the University of Pécs, Hungary.

First, the procedures of anchoring will be presented, then the stages and operations of the anchor study will be described. Next, the steps and results of the analysis will be discussed. Finally, conclusions will be drawn on the basis of the findings.

Anchoring procedures

An IRT-based item bank is based on the idea that all the items in the bank can be calibrated to be placed on the same difficulty continuum (Willmot & Fowles, 1974, p. 49). This can be achieved by using the same reference point for all the items in the calibration process. Obviously, this means that the numerical difficulty estimates associated with the items are not objective or independent. However, within the bank, all items are comparable to all the other items.

Practically, this means that the first items to be placed in the bank serve as the reference point for all items to be added later, but the numerical values associated with these first items are, in fact, arbitrary. From the first items on, however, all items will relate to the same reference point, making it possible to place all the items on the same scale of difficulty.

In the process of bank building one key issue is the way new items are added. As the purpose is to place all items on the same scale, it is necessary to provide a link between the new items and the ones already in the bank. This link is formed by including calibrated items from the bank (often referred to as anchor items) in the test which contains the new items (Henning, 1987, p. 131). In the course of calibration the anchor items' calibrated difficulty estimates are used as the reference point in relation to which the new items' difficulty estimates are calculated. This process is illustrated by Table 1.

As can be observed, Test 1 and Test 2 contain ten common items. These are the anchor items, which are supposed to provide the link between old items (in Test 1) and new items (in Test 2). It can also be detected that the calibrated difficulties for the anchor items are different in Test 1 and Test 2. This is the result of calibration without reference to earlier item difficulty estimations. If, however, the anchor items' original difficulty estimates are used in the calibration process (Test 2 [Adj.]), linking is accomplished, and the new items' (items 11-15 in Test 2) difficulty estimates become directly comparable to those of the old items not included in Test 2 (items 1-5 in Test 1).

Using this procedure it is possible to increase the number of items in the bank, and the set of anchor items may be different in each new test, as long as they are all selected from among the calibrated items in the bank.

Table 1 raises two important questions, however. First, when examining the number of items, one can observe that in this example ten out of fifteen items are anchor items. This would suggest that the number of new items that can be added to the bank with each test administration may be seriously limited by the fact that a fairly large number of anchor items may be needed for successful linking. While Table 1 is admittedly a mere illustration, the question could be raised: how many anchor items are necessary for reliable linking between old and new items.

Second, the difficulty estimates of the anchor items in Table 1 represent a fairly wide range of item difficulty (from –3450 to 2.00). While, again, Table 1 is clearly just an example of such a procedure, it seems legitimate to wonder whether such a wide range is necessary, or, indeed, whether the range of difficulty estimates for anchor items matters at all in the calibration process.

While these questions seem obvious, the literature appears not to answer either of them. Despite the fact that item bank building, anchoring and item calibration are frequented topics in IRT-related sources (see e.g., Baker, 1997; Brown, 1997; Choppin, 1981; Hambleton et al., 1991; Hemker, Sijtsma, Molenaar & Junker, 1996; Henning, 1986; de Jong, 1986; Pollit, 1999; van der Linden & Hambleton, 1996; Wainer & Mislevy, 1990) the number and characteristics of anchor items are not discussed in detail, except for suggestions concerning the statistical parameters (such as discrimination or model fit) determining the suitability of items to serve as anchors.

In the following, then, in the light of the analysis of data gathered in an item bank building project, I will attempt to examine how the questions raised can be answered, and whether the answers offered can be utilized in the practice of anchoring.

Table 1: Sample linking procedures (adapted from Henning, 1987, p. 132)

	Test 1		Test 2		Test 2 (Adj.)	
	Item No.	Difficulty	Item No.	Difficulty	Item No.	Adj. Difficulty
	1	4.50				
	2	4.00				
	3	3.50				
	4	3.00				
	5	2.50				
C						
O	6	2.00	1	4.50	1	2.00
M	7	1.50	2	4.00	2	1.50
M	8	0.00	3	3.50	3	0.00
O	9	-1.50	4	3.00	4	-1.50
N	10	-2.00	5	2.50	5	-2.00
	11	-2.50	6	2.00	6	-2.50
I	12	-3.00	7	1.50	7	-3.00
T	13	-3.50	8	0.00	8	-3.50
E	14	-4.00	9	-1.50	9	-4.00
M	15	-4.50	10	-2.00	10	-4.50
S						
			11	-2.50	11	-5.00
			12	-3.00	12	-5.50
			13	-3.50	13	-6.00
			14	-4.00	14	-6.50
			15	-4.50	15	-7.00

Mean link Test 1 = -17.50/10 = -1.75
Mean link Test 2 = 17.50/10 = 1.75
Translation Constant = Test 1 mean link minus Test 2 mean link = -1.75 - (1.75) = -3.50
Adjusted Difficulty = Test 2 Calibrated Difficulty plus Translation Constant

Research design

The data to be analyzed were gathered in the course of the item bank building project at the University of Pécs. In this project a language test item bank is being built, which is to include multiple choice grammar and usage items for the purpose of filtering first-year English majors through proficiency testing. Each year new items are added to the bank after having been used in that year's version of the filter test. For analysis purposes the BIGSTEPS computer program was used (Wright & Linacre, 1992).

As the bank is based on one-parameter IRT difficulty estimates, the process of linking each new version to the bank through the use of anchor items has been applied. Each new version includes anchor items that have been calibrated in earlier versions, and item difficulty estimates are always based on the earlier difficulty estimates of anchor items.

The study was based on the assumption that to check whether or to what extent the number and difficulty range of anchor items influences calibration it seems best to lower the number of anchor items dramatically. Theoretically, the lowest number of anchor items would have been one. However, as the study was also to examine how the anchor items' difficulty range affects calibrations, it was decided that two anchor items would be used, as this is the lowest possible number of items producing a range.

The actual study consisted of re-calibrating item difficulty estimates for each version of the test between 2000 and 2005 in two stages. In stage one, two anchor items were selected whose difficulty estimates showed a minimal difference in the group of anchor items used originally in that particular version. Hence, besides lowering the number of anchor items to the minimum, the range of anchor items' difficulty was also reduced to the lowest possible for that year.

In stage two, again, two anchor items were selected, but this time it was the two items showing the biggest difference in terms of difficulty. This way the biggest possible range was provided. At the same time, it should be noted, this also meant the preservation of the original range of item difficulty, as any range is based on two items assuming extreme values on a scale regardless of the number of items in between. A summary of Stage 1 and Stage 2 anchor item selections is presented in Table 2.

As can be observed, there appears to be a major difference in terms of range between Stage 1 and Stage 2 data. Even the smallest difference (in 2004) was a hefty 2.66 logits. By comparison, the largest difference (in 2002) was 3.66 logits. The number of anchor items in different versions was stable except in 2001, when nine items were used, instead of ten anchor items in all the other test versions.

Table 2: Anchor item difficulty ranges (in logits)

	2000	2001	2002	2003	2004	2005
original number of anchor items	10	9	10	10	10	10
original range of anchor item difficulty	3.23	3.09	3.69	3.59	2.70	3.30
Stage 1 range of anchor item difficulty	0.13	0.07	0.03	0.03	0.04	0.01
Stage 2 range of anchor item difficulty	3.23	3.09	3.69	3.59	2.70	3.30

After the selection of different pairs of anchor items for Stage 1 and Stage 2, the actual analyses took place. Both in Stage 1 and Stage 2 item difficulty estimates were calculated for all the items in the test. Next, Stage 1 and Stage 2 item difficulty estimates were compared to the original difficulty measures calculated on the basis of all the original anchor items' calibrated difficulty estimates. This took place in two steps. First, average item difficulty estimates were computed for all three sets of data—i.e. for the original estimates, as well as for Stage 1 and Stage 2 difficulty figures—and then the differences between the mean averages of original difficulty estimates and those of Stage 1/Stage 2 estimates were calculated.

The second step was the checking of whether the differences detected this way were statistically significant. For this purpose paired t-tests were conducted, using SPSS. After completing these steps it became possible to evaluate to what extent the assumptions concerning the impact of reduced anchor item number and range held true.

Results

In this section the results of the analyses will be dealt with in the form of examining the findings related to each test version separately. Test versions will be discussed in a chronological order.

The 2000 version in its original form contained ten anchor items. With this number being reduced to two, it was interesting to see how item difficulty figures changed in general. This is depicted in Figure 1.
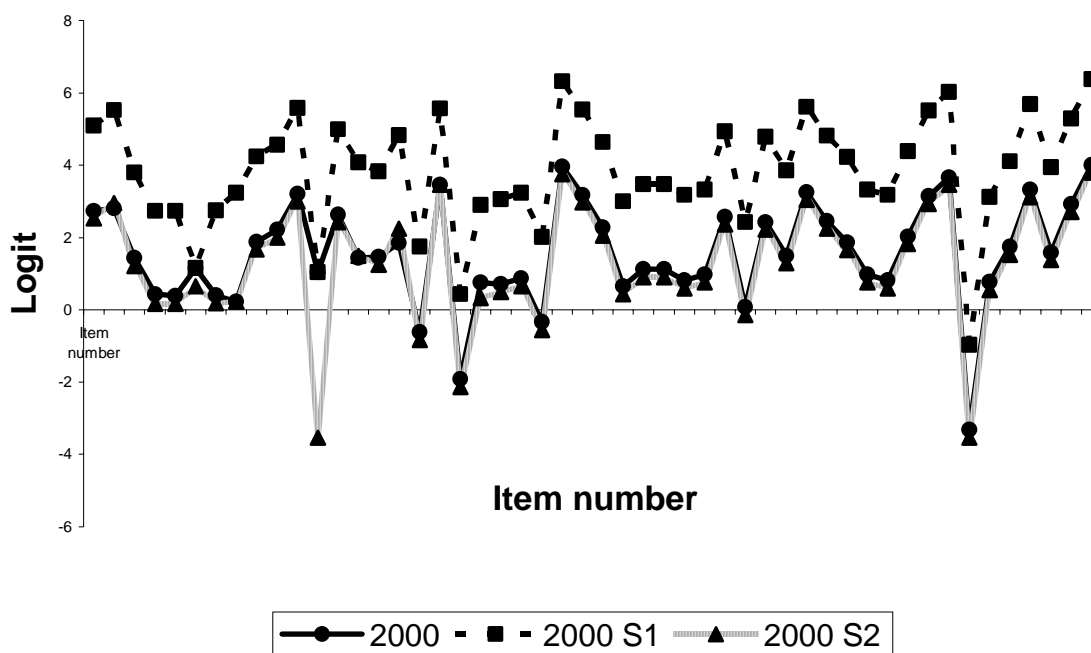
Figure 1: Item difficulty estimates in the 2000 version

As can be observed, the original and the Stage 2 results do not appear to be very different, while Stage 1 results tend to be higher, though still following the original item difficulty pattern as defined by the original data. Indeed, it was found that mean item difficulty figures showed a small difference between the original and Stage 2 results (0.278 logits) and a larger difference between the original and Stage 1 results (2.290 logits). Interestingly, however, both figures were identified as significantly different from the original mean item difficulty estimate at p<0.001 in the case of Stage 1 and at p<0.01 in the case of Stage 2. At the same time, both Stage 1 and Stage 2 results showed an extremely high, significant (at p<0.001) correlation with the original results (0.948 and 0.921, respectively). It would seem then that the reduction of the number of anchor items resulted in a significant change, but its magnitude depended on the range of anchor items chosen. Moreover, the re-estimated difficulty figures showed an extremely strong relationship with the original estimations. This, however, would be a mistake to generalize on the basis of only one test version. To attempt to answer the research questions the other results should be examined as well.

The 2001 version shows a very similar picture. While the differences in mean anchor item difficulty were not particularly high (Stage 1: 0.049, Stage 2: 0.307), both figures were found significantly different, Stage 1 at p<0.01, while Stage 2 at p<.001. Just like in the case of the 2000 results, Stage 1 results were somewhat higher, while Stage 2 results lower than the original estimates. The difference, however, was quite small in both cases, as can be observed in Figure 2.
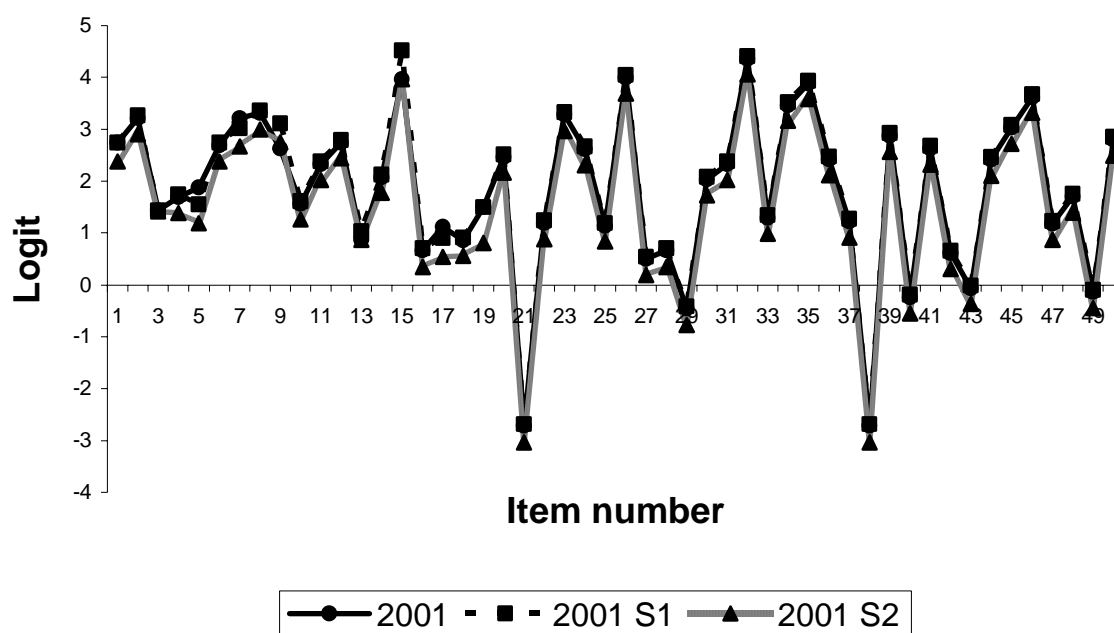
Figure 2: Item difficulty estimates in the 2001 version

Once again, both re-calibrated difficulty sets showed a very high (Stage 1: 0.997; Stage 2: 0.996), significant (p<0.001) correlation with the original estimates.

The analysis of the 2002 data reveals an even more interesting picture. As can be observed in Figure 3, the results of all three calibrations appear to be virtually the same.

Indeed, the difference in mean difficulty compared to the original difficulty estimates was minute in both Stage 1 and Stage 2 (0.219 and 0.046, respectively). Yet, both of these figures were found to be significantly different from the original mean of estimates at p<0.001. In 2002 it was the Stage 2 results that were somewhat higher than the original figures, unlike in the previous two years. Yet again, both recalibrated sets of estimates showed an extremely high, significant correlation with the original estimates (0.999 at p<0.001 in both cases).

Apparently, then, a pattern seems to be emerging. While recalibrations do change the numerical values of item difficulty estimates in a way that is clearly not attributable to chance, in all but one case (Stage 1 2000) the actual mean item difficulty estimates change very slightly, and in all cases the recalibrated estimates show a spectacular correlation with the original estimates. To see whether this pattern holds true for the other three versions to be analyzed, let us take a closer look at those results as well.

The year 2003 results appeared to be very similar to the year 2002 ones. As is presented in Figure 4, the two recalibrated sets of estimates appear once again nearly to coincide with the original difficulty figures.
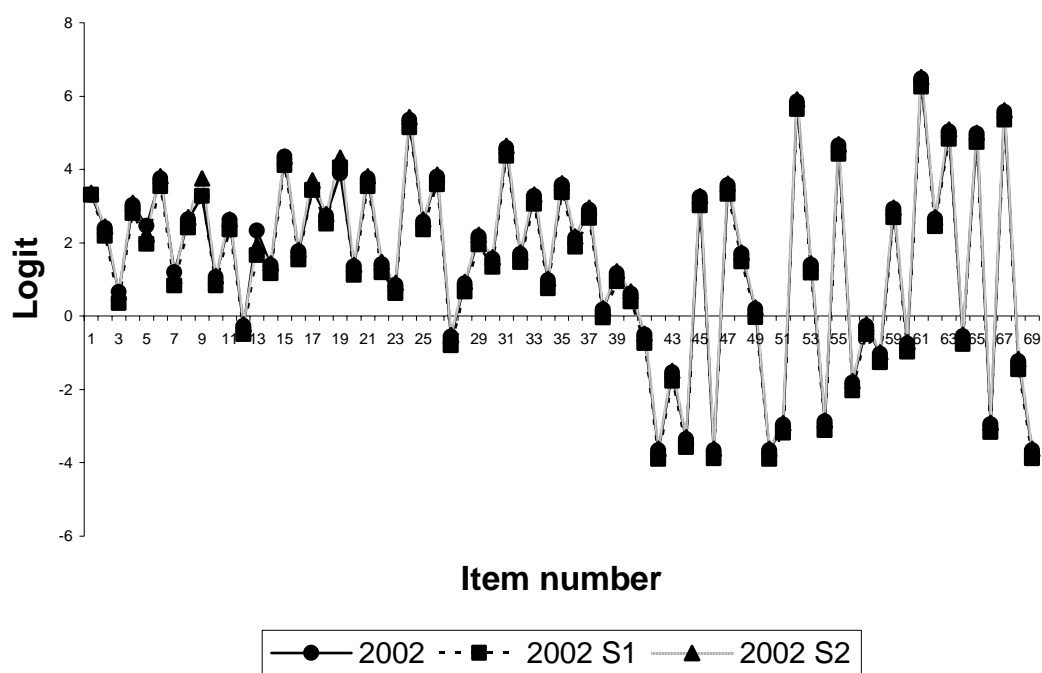
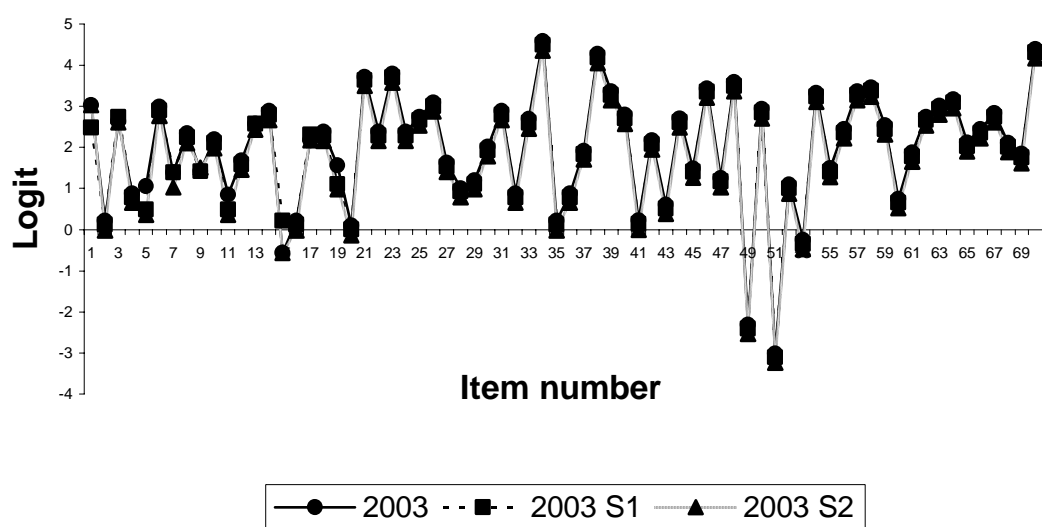Figure 3: Item difficulty estimates in the 2002 version

Figure 4: Item difficulty estimates in the 2003 version

The results of the t-tests were also similar to earlier years'. The difference in mean item difficulty was small for both Stage 1 and Stage 2 (0.088 and 0.222, respectively), and both sets of estimates were found to be significantly different from the original at p<0.001. Similarly, correlation figures were significant (at p<0.001) and extremely high again (Stage 1: 0.994; Stage 2: 0.997).

The year 2004 results, however, reveal a slightly different pattern. As can be observed in Figure 5, Stage 1 results this time were, again, nearly identical to the original estimates. Stage 2 results, on the other hand, appear to be markedly different.
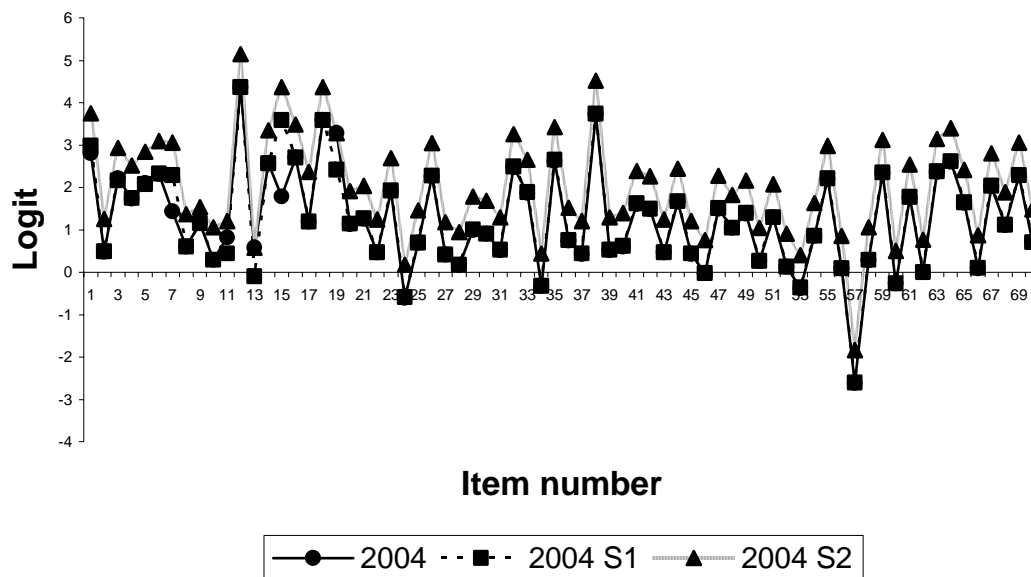
Figure 5: Item difficulty estimates in the 2004 version

Upon close observation it was revealed that this time Stage 1 results were even more similar to the original estimates (mean difference: 0.018) than in previous years, and this difference was not found significant. Stage 2 results, however, showed a greater mean difference (0.788) than in previous years, which was significant at p<0.001. Correlation figures for both Stage 1 and Stage 2 were, yet again, high (0.972 and 0.971, respectively) and significant at p<0.001.

Why Stage 1 results did not significantly differ from the original figures in this year's data seems difficult to tell. The procedures, the items and the candidates did not appear to be different from earlier years; hence, there seems to be no easy and obvious explanation. Instead of attempting to explain this unique similarity—which, if we only consider these results, may end up being little more than speculation—let us examine the last observed year's results and attempt an explanation on the basis of all data. This year brought about yet another pattern of results, as depicted in Figure 6.
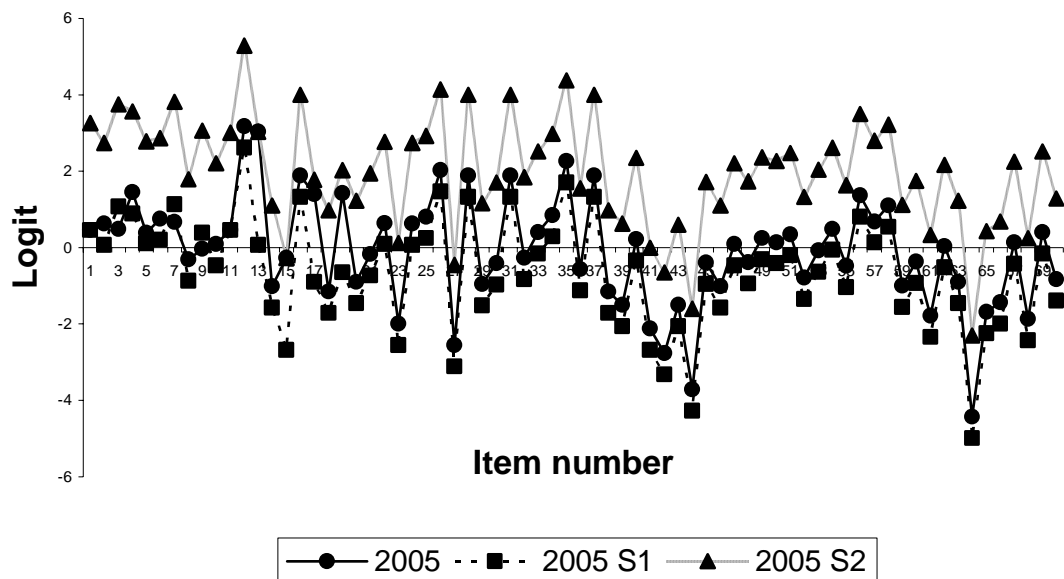
Figure 6: Item difficulty estimates in the 2005 version

Interestingly, this time Stage 1 results were somewhat more different than in previous years (mean difference: 0.601), and Stage 2 results showed an extremely large mean difference (2.077) comparable only to the Stage 1 mean difference in 2000. Yet, both Stage 1 and Stage 2 results still show a strong (0.934 and 0.935, respectively) and significant ($p<0.001$) correlation with the original difficulty estimates.

Considering all these results one can observe immediately that there seems to be quite a bit of variation in terms of how different Stage 1 and Stage 2 results were compared to the original estimates, and no clear pattern seems to be emerging concerning the relationship between Stage 1 and Stage 2 results either. In the following, an attempt will be made to examine what may lie behind these results.

Implications

In the previous section a detailed account was given of how the Stage 1 and Stage 2 anchoring procedures changed the calibrations of item difficulty in different versions of the test. Before drawing the general conclusions it seems appropriate to summarize the implications the various anchoring designs carried, which can best be achieved with the help of an overview of the findings. Accordingly, a summary of the results of the analyses is offered in Table 3.

Table 3: Mean item difficulties and differences

	2000	2001	2002	2003	2004	2005
original mean item difficulty	1,562	1,835	1,508	1,976	1,244	-,088
Stage 1 mean item difficulty	3,852	1,885	1,289	1,887	1,263	-,690
Stage 2 mean item difficulty	1,284	1,528	1,555	1,754	2,033	1,988
Stage 1 difference in mean item difficulty	2,290*	0.049*	0,219*	0,088*	0,018	0,601*
Stage 2 difference in mean item difficulty	0.278*	0.307*	0,046*	0,222*	0,788*	2,077*

* significant at p<0.001

Based on the data, one may quickly jump to the conclusion that, since in the over-whelming majority of the cases the re-calibration of item difficulty indices produced significantly different mean item difficulty figures (regardless of the range of the newly selected anchor items), such a drastic lowering of the number of anchor items is not feasible. Yet, upon observing the data more closely, one may find reasons to be cautious.

First, in all but two cases (Stage 1, 2000 and Stage 2, 2005) the mean difference figures are below 1 logit, indicating that the actual differences are relatively small. This is all the more confirmed by the fact that in two thirds of the cases (Stage 2, 2000; Stage 1 and 2, 2001-2003; Stage 1, 2004) the difference is smaller than 0.5 logits, and in one third of all cases (Stage 1, 2001; Stage 2, 2002; Stage 1, 2003-2004) this figure is below 0.1 logits. What this seems to imply is that while reducing the number of anchor items does result in a change in mean difficulty estimates that cannot merely be attributed to random variables, the actual distortion in the results is quite small.

Second, not only are the changes very small, but highly systematic as well. As shown in Table 4, all Stage 1 and Stage 2 results demonstrate extremely high, significant correlations with the original estimates.

Table 4: Stage 1 and 2 correlation figures with original estimates

	2000	2001	2002	2003	2004	2005
Stage 1 results' correlation with original	0.948*	0.997*	0.999*	0.994*	0.972*	0.934*
Stage 2 results' correlation with original	0.921*	0.996*	0.999*	0.997*	0.971*	0.935*

* significant at p<0.001

Apparently, then, while the reduction of the number of anchor items may have resulted in small but significant changes in item difficulty estimates, these changes seemed to have virtually no impact on the relative positioning of items on the difficulty continuum.

What follows from here is an important conclusion. It seems like applying few anchor items results in some inaccuracy, but virtually no inconsistency. In other words, when the number of anchor items is reduced dramatically, to the smallest number still producing a range, the resulting difficulty estimates merely appear to be shifted on the difficulty continuum, but they preserve their relative position to one another.

A set of questions remain to be answered, however. First, it is unclear from the present study what exactly caused the fluctuations in terms of how much the Stage1 and 2 results differed from the original estimates. While in many cases the differences were small, and in one case so minute as being insignificant, in some other cases this was not so. Based on the data there seems to be no factor that could easily be identified as being responsible for this pattern of results.

Second, it is also unknown how results would have changed with re-calibration if more than two anchor items had been selected. Three, four or even five items would still have meant a major decrease in the number of anchor items. Of course, if this had been done, selection criteria would have been needed to decide which anchor items to drop as anchors.

Third, it is also unclear how results would have been different if the original number of anchor items had been higher. A set of, say, twenty items may have provided a stronger link between already banked and new items, but it is also possible that such an increase would have been unnecessary.

Yet another consideration is how results may have changed if the selection of Stage 1 and 2 anchor items had been based, instead of difficulty and range, on some other item characteristic, for instance on degree of fit to the model.

While these questions are impossible to answer on the basis of the present study, they could well form the basis for further research.

Conclusion

In this paper, in the light of empirical data, I examined the process of linking new items to already calibrated ones by means of anchor items in order to clarify issues related to the number and characteristics of these anchor items. Having examined the implications of the data, it seems appropriate now to attempt to answer the questions raised earlier in this paper.

The first question was related to the number of anchor items necessary for reliable linking. In the light of the current anchor study it seems that no conclusive answer can be provided. What seems clear is that generally two items are not enough to guarantee that difficulty calibrations will be the same as in the case of using ten anchor items. What this means, however, is that in some cases even two items can do the trick, while in some other cases the calibrations do change, but only to a very limited extent, albeit in a statistically significant way. Clearly, the

key question is what causes these differences, and how the changes can be controlled. Also, as indicated in the previous section, many more questions can be asked regarding the number of anchor items to use even if the number of such items is quite low. Further research may clarify these points.

The second question concerning the possible impact of the range of anchor items seems to have generated a negative answer. Based on the findings of the present study, there seems to be no reason to assume that the range of items has any perceivable impact on difficulty calibrations. This conclusion, of course, is based on the fact that the results of the study indicated no clear pattern of range impact. In certain cases a wide range produced more similar calibrations, while in some other cases the opposite was true. It is important to note, however, that the issue of number and range of anchor items may, in fact, be related. If two items cannot be used for anchoring effectively, range-related assumptions cannot be examined in this design, either. In a different design, with more items, on the other hand, range may be found significant. Also, once some explanation is provided to the fluctuation described above, and a method is proposed for anchoring with two items, the issue of range may well be revisited, too. Once again, further research will need to be conducted to shed light on these issues.

Obviously, the conclusions delineated above are to be handled with caution for several reasons. First, the study described is based on the performance of a relatively small population. Second, the data were gathered from a language test item bank building project focusing on the construction of a grammar and usage item bank made up exclusively of multiple choice items. These circumstances generate a somewhat more limited variety of items, which may have had an impact on the process of anchoring as well: the results may have been influenced by a somewhat unnatural homogeneity of items as compared to even other fields of language assessment.

This may be all the more true of item clusters banked and re-used together. In listening or reading tasks, for instance, individual items, in theory, cannot be separated from the others. Hence, the reduction of anchor items is only possible in terms of testlets or whole tasks.

Whatever the answers will be to these questions, however, the issue of the number and difficulty range of anchor items is clearly a significant one. From a theoretical perspective the idea of two items defining a scale of reference is perfectly acceptable. In a two dimensional system once a distance has been defined by two points, this distance can be used as a reference to measure any other distance as well. As the difficulty continuum can essentially be regarded as a two dimensional system, two-item anchoring sounds like an acceptable hypothetical frame of reference. While this study's results do not support the acceptability of such a frame of reference, more empirical research needs to be carried out in order to test the hypothesis.

As future studies hold countless important answers to all these questions, it seems it is time now to embark on a journey of research to measure distances unknown and to explore uncharted continua. Anchors aweigh!

References

Baker, R. (1997). Classical test theory and Item Response Theory in test analysis. Special Report No 2: *Language Testing Update*.

Brown, J. D. 1997. Computers in language testing: Present research and some future directions. *Language Learning and Technology*, *1* (1), 44-59.

Choppin, B. H. (1981). Educational measurement and the item bank model. In C. Lacey & D. Lawton (Eds.) *Issues in evaluation and accountability* (pp. 87-113). London: Methuen.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park: SAGE Publications.

Hemker, B. T., Sijtsma, K., Molenaar, I. W., & Junker B. W. (1996). Polytomous IRT models and monotone likelihood ratio of the total score. *Psychometrika*, *61* (4), 679-693.

Henning, G. (1986). Item banking via DBASE II: the UCLA ESL proficiency examination experience. In C. W. Stansfield (Ed.), *Technology and language testing* (pp. 69-77). Washington, D.C.: TESOL.

Henning, G. (1987). *A guide to language testing*. Boston, Mass.: Heinle & Heinle.

de Jong, J. H. A. L. (1986). Item selection from pretests in mixed ability groups. In C. W. Stansfield (Ed.), *Technology and language testing* (pp. 91-108). Washington, D.C.: TESOL.

Pollit, A. (1999). Rasch measurement in latent trait models. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education* (Vol. 7). London: Kluwer Academic Publishers.

van der Linden, W. J., & Hambleton, R. K. (1996). *Handbook of modern Item Response Theory*. New York: Springer-Verlag.

Wainer, H., & Mislevy, R. J. (1990). Item response theory, item calibration and proficiency estimation. In H. Wainer, N. J. Dorans, R. Flaughter, B. F. Green, R. J. Mislevy, L. Steinberg & D. Thissen (Eds.), *Computerized adaptive testing: A primer* (pp. 61-101). Hillsdale, NJ: Lawrence Erlbaum.

Willmott, A. S. & Fowles, D. E. (1974). *The objective interpretation of test performance*. Slough: National Foundation for Educational Research.

Wright, B. D. & Linacre, J. M. (1992). *BIGSTEPS Rasch model computer program. Version 2.3*. Chicago: MESA Press.

The Role of Cooperation in Pre-Service Education and In-Service Development in TEFL

Stefka Barócsi

Kispesti Deák Ferenc Secondary School, Budapest, Hungary

flower74@freemail.hu

Introduction

This paper seeks to discuss a pilot study on the role of cooperation in pre-service education and in-service development. The context refers to teacher preparation leading to team teaching as a special form of teaching in the course of teaching practice, and ultimately cooperation in the teaching profession.

The emphasis throughout this discussion is on two key notions: collaboration and cooperation. These terms are defined in the *Oxford Advanced Learner's Encyclopedic Dictionary* as follows: cooperation refers to "working together for a common purpose" (Hornby, 1994, p. 200); whereas collaboration is related to "working together with somebody to create or produce something". (Hornby, 1994, p. 174). Clearly, there is a close resemblance between the two terms. The first definition does not contain the element of product, while the second definition does not mention the element common purpose. Both definitions, however, imply the missing elements and it may be said that in reality, the three factors of mutual work, personality and common purpose are always interdependent.

Furthermore, it is necessary to clarify what is meant by the two terms in educational research. Collaboration is generally regarded as consistently working together and sharing responsibility in order to accomplish a task (Knezevic & Scholl, 1996; Medgyes & Malderez, 1996). The notion of cooperation is assumed to mean individuals working together to achieve a task or particular goal for which responsibility is not necessarily shared (Medgyes & Malderez, 1996; Nunan, 1992; Slavin et al., 1985). If we consider this plausible distinction and follow it, it seems that cooperation refers to a more general and abstract level; whereas collaboration involves more day-to-day work in which tighter bonds are possible. In fact, when used in the context of collaborative learning and teaching, the distinction between the two terms becomes difficult to sustain. It should be noted here that the use of the two terms is not consistent in the literature (Gwyn-Paquette & Tochon, 2002; Medgyes & Malderez, 1996; Medgyes & Nyilasi, 1997; Nunan, 1992; Wilhelm, 1997). Therefore, it is not only necessary to compare and contrast, but also to integrate the two terms.

Before launching into the review of the research, it is useful to briefly consider what is meant by cooperation and collaboration in this paper. Following the above considerations, the present study uses the two key notions interchangeably. They refer to a process of working together and sharing responsibility in order to try to achieve a mutual goal. This broader perspective on the main concepts of cooperation and collaboration is meant to provide a useful tool to facilitate the understanding of the multi-faced issue of the process of one person working with another; an issue which deserves a more profound investigation.

Rationale for the study

The spark that lit the fire for this research is my involvement in teacher training at the Centre for English Teacher Training (CETT) at Eötvös Loránd University in Budapest. Up to the present, I have worked with fifty student teachers in the course of thirteen years. My work as a teacher trainer is highly relevant to me and has led me into new areas of knowledge and skills. The teaching practice is a not an uncommon phenomenon in teacher education, however, it can be looked at as the management of a complex scheme and appropriately addressed are its aims, structure, content and roles of participants involved. Somewhat simpler is to see the participants as links in a chain. This chain integrates the claims of theory and practice, society and the individual. The entire period of teaching practice is really too great a manifestation of hard work, effort, interest, enthusiasm and emotions.

My experience as a teacher trainer over the years has proved that although the concept of teaching practice is simple, in reality, the interaction between participants creates a complex situation. Thus emerges the need for fruitful cooperation in order to sustain the system of teaching practice. My interest in this area resulted in research on cooperation in teacher education (Barócsi, 1998). The conclusion was that cooperation in the teaching practice is a valuable experience. What the study touched upon was cooperation in the teaching profession. The involvement in the research process and the insights into my daily practice as a teacher trainer have made me particularly interested in the notion of cooperation. With regard to the teacher trainer who is the researcher in this particular situation, the knowledge and understanding of the process of cooperation of pre-service and in-service teachers is crucial. This is an issue, which needs thorough examination. The current pilot study attempts to serve this purpose. It involves my previous work on the validation of the long qualitative interview schedule (Barócsi, 2005). The findings of these studies had implications for the further exploration of the topic leading to the main research project on the area of my interest. Ultimately, the research is regarded as a call for cooperation in teacher preparation and the teaching profession.

Research topic and research questions

My experience as a school-based teacher trainer has shaped the one key idea to hold onto, which is that of cooperation in the course of the teaching practice, where the fundamental goal is student-teachers' professional formation. The concern in my role has always been to foster cooperation within the framework of teaching practice. In line with my work, I felt it was important to identify what makes cooperation in teacher training successful. Furthermore, I have always been interested in finding out if a special model of educating foreign language teachers with a teaching practice longer than usual in the Hungarian teacher training institutions can have an impact on collaboration in the teaching profession. In this respect, the research questions are formulated as follows:

- What factors interact in cooperation in pre-service teacher preparation?
- What factors interact in cooperation in the subsequent in-service teacher development?
- What is the influence of pair or team teaching in the training of teachers on cooperation in their careers?

In order to answer these questions I investigated the field of cooperation. The aim was to explore cooperation in pre-service teacher education and how it could affect relationships in a career. The purpose of referring to the Centre for English Teacher Training in Budapest was two fold: to look at the pair or team teaching side of the special mode of teacher training and to investigate how trainees develop and manage their professional lives after graduation. As the focus was on cooperation, it was relevant to see if they applied cooperative teaching techniques or had opportunities to cooperate with their colleagues. My intention was to go deeper into the issue of collaboration and explore it both theoretically and practically. The ultimate aim was to determine the values inherent in collaboration.

Previous studies on cooperation in pre-service TEFL

An examination of the research literature on the topic of cooperation is essential in order to understand the important issues. The literature I explored covered mainly three areas: cooperation in the learning process, cooperation in teacher education and cooperation in the teaching process.

Cooperation in the learning process

Cooperative learning is widely supported at the present time (Goddard, Hoy & Woolfolk, 2004; Gwyn-Paquette & Tochon, 2002; Liang, 2004). The most direct evidence for the value of cooperating to learn has been provided by Slavin (1985); who interprets the effects of working cooperatively on student achievement. The data from his research indicate that working together toward a common goal has

positive effects on group relations. The common goal to achieve success contributes to the process of learning because students tend to help each other (Gwyn-Paquette & Tochon, 2002; Slavin et al., 1985). The importance of collective beliefs in addition to group goal attainment has been perceived as essential for collective efficacy (Goddard, Hoy & Woolfolk Hoy, 2004). Drawing on investigations of the learning process, Nunan (1992) points out that collaboration in group learning encourages the learners to learn better. This is to suggest that collaborative learning provides opportunities for raising students' awareness. Awareness of oneself as a learner and awareness of the learning process are obviously essential if any progress is to be made and the increase of responsibility is a purpose (Goddard et al, 2004; Nunan, 1992). Within the positive view of group work (Liang, 2004) students' perceptions of cooperative learning are given serious attention. Edge (1992a), for example, reminds us that learners learn in different ways and suggests that the articulation of thoughts and feelings must be taken into consideration. In addition, exchanging ideas helps people to communicate with other people. As the cooperative approach provides more opportunities for interaction, it appears crucial because it promotes learning (Edge, 1992a, 1992b; Gwyn-Paquette & Tochon, 2002). It is even more so in the process of teaching. This important question forms the basis of the next sections.

Cooperation in teacher education

First and foremost it is essential to clarify what is needed for successful teacher education (Gwyn-Paquette & Tochon, 2002). As a matter of fact, there appears to be a general recognition (Freeman, 1989; Kennedy, 1993; Swan, 1993, Wallace, 1991) of the basic dimensions, namely skills, knowledge, awareness, practice, and understanding. However, there is much more to development, as much more is needed in order to gain success in language teacher education. Kagan (1992) describes professional growth of novice teachers as changes over time in behaviour, knowledge, images, beliefs or perceptions. In this respect, as Kagan (1992) recognises, development involves a multidimensional change. In the process of learning to teach, on the one hand, trainees are to achieve practical knowledge and form concepts, on the other hand they have to develop interpersonal skills and face affective issues (Kerry & Mayes, 1995). The recognition (Kagan, 1992; Kerry & Mayes, 1995) that the teacher learning process has a multidimensional aspect explains why it is a long process.

   Schemes of pre-service teacher education are always responses to the specific need to produce teachers. Team teaching exists as a new and more beneficial approach to gaining the main ingredients of professional training as a teacher. Team teaching is regarded (Bodóczky & Malderez; 1994; Medgyes & Malderez, 1996; Medgyes & Nyilasi, 1997) as a useful form of education. It is a prolonged activity and its effectiveness results from the fact that the participants' language proficiency, language awareness and pedagogical skills are enhanced in the process of interaction and negotiation. Combined experiences encourage reflection. Working together in the preparation of classes and in sharing of views after the lessons de-

velops decision-making about teaching, raises awareness and reinforces the process of reflection (Gower & Walters, 1983). The underlying assumption is that partner support during the teaching experience relieves anxieties, fosters the feeling of security as well as the learning process (Bodóczky & Malderez, 1994; Medgyes & Malderez, 1996). The question of how future teachers develop their skills, knowledge, confidence and capacity for reflection remains relevant. The emerging key issue is that collaborative work between trainees has a major effect on student teachers' professional development (Knezevic & Scholl, 1996). On the whole, the implications (Claxton, 1989; Nunan, 1992) are that cooperation exists in teacher education where student teachers work together. Additionally, trainees are provided with support throughout their practice by their trainer. The primary task is to ensure the successful operation of the teaching practice. On the whole, systematic discourse and on-going opportunities lead to important mentor-trainee cooperation. Finally, it is worthwhile making an attempt to look at what goes beyond the scope of teacher preparation.

Cooperation in the teaching process

We move to a consideration of the role of cooperation in the teaching profession. At present, it is mostly cooperation in learning and cooperation in teacher education that is in the focus of research. The wealth of literature on the field of education offers inadequate information addressing this topic. The notion of collaboration is deeply rooted in the combat against isolation (Brandes & Ginnis, 1992; Claxton, 1989; Slavin et al., 1985; Underwood, 1987). One of the central issues is that there is a need to alter this present day situation (Claxton, 1989; Nunan, 1992). A demand (Nunan, 1992) is created for promoting the idea of cooperation rather than competition in language education where teachers, learners, and curriculum specialists can collaborate for various reasons and purposes. As cooperation possesses invaluable advantages, an appeal is made to set up a sort of cooperation among teachers of English as a foreign language (Underwood, 1987). Medgyes (1994a; 1994b), for example, offers thoughtful insights into collaborative approaches to teaching, the main philosophy of which is based on cooperation between teachers. There is an additional and very important reason: collaboration helps people learn, therefore, it has become a requisite for professional growth among beginning teachers (Kagan, 1992). Cooperation promotes professional development in the sense that teachers can learn from each other by means of professional interactions. This underlies the need for collaboration in the realm of the teaching profession.

Research design

Research method

In order to collect data for the research on the role of cooperation in pre-service education and in-service development, a qualitative approach was adopted. The circumstances justified the choice of interviews rather than questionnaires. The choice of method used to collect and consequently analyse data was determined by first, the low number of respondents; second, an interest in the participants' detailed experiences and personal impressions about team work; third, the main aim to achieve in-depth investigation. This qualitative research included elements of ethnographic study with regard to the teacher trainer who was the researcher in this situation. The research design for the pilot study can be seen in Appendix A.

Participants

The core of data was gained from one pre-service and one in-service teacher. The participants were randomly chosen to take part in the study. Respondent 1 was a student teacher at CETT at Eötvös Loránd University in Budapest. As a double-major student this participant followed a new four-year program of study. Respondent 2 was a previous student teacher who graduated from CETT, attending a three-year program with one major. It must be noted here that, regarding curriculum issues, i.e. the kind and number of classes students received, the two programs were practically the same. The main difference lay with the time of the teaching practice component within the programme. It used to be in the third year and lasted for one whole year. As for the students following the new program, the teaching practice took one semester and was in the fourth year. For both programs the course of study at CETT is related to innovation that involves team teaching as a special form of teaching in an extended teaching practice-a period leading to personal and professional growth.

For the pilot study two persons were interviewed. Both participants were female. Respondent 1 was at the beginning of the teaching practice and was going to work with a partner to teach a class for one term. Respondent 2 had a full-time job, working as a teacher of English at a secondary grammar school. This participant had already worked with the researcher during her teaching practice. The investigator in this particular situation relied on the successful relationship established during that period. As far as team teaching was concerned, Respondent 2 had been involved in real-life collaborative teaching during one year long teaching practice, whereas Respondent 1 had never been involved in real-life collaborative teaching before. The several cases of team teaching in the form of micro-teaching for Respondent 1 were not to be completely ignored. However, they were not regarded as significant previous experience in team teaching. Obviously, from the point of view of experience in collaborative teaching and work experience, the participants did not present a homogeneous group. Vital for the success of the interviews were the participants' tolerance, willingness to cooperate, empathy and

openness. The respondents greatly elaborated on the questions and proved a source of invaluable ideas.

Evidently, the focus of my research was on the teachers, who were related to the special form of collaborative teaching. The students were regarded as additional participants. The pilot study did not have data on their opinions. The students' perspective had been somewhat neglected and not included in the research design. The students, however, appeared indirectly mentioned in the participants' reflections. There was more than a suggestion from the respondents' interviews that students' perceptions needed consideration. The pilot study revealed the assumption that including students' perceptions was important. This indicated the necessity for another aspect of the process when looking at possible directions for further research.

Description of data collection

Although the pilot study dealt with a limited number of participants, an attempt was made to carry out a research which was value bound (Johnson & Onwueg-buzie, 2004). In fact, the foundation from which the research method started to develop was Mc Cracken's, (1988) four-step model for carrying out a long qualitative interview. Mc Cracken's model outlines four priorities: first, the study of literature; second, the cultural context and the object of the study; third, the construction of the questions, creating the prompts and the interview procedure; fourth, the analysis of qualitative data. One of the central parts of the study is the review of the literature on the main topics concerned. Examination of the research literature on the topics was seen essential in order to understand the important issues. This stage was justified by the complexity and broadness of the area of research. The project involved prolonged engagement, persistent observation, triangulation and thick description of the collected data (Lazaraton, 2003).

In order to answer the research questions, two audio-recorded interviews were conducted. Two sets of questions were designed: thirty-nine questions were suitable for the beginning of the teaching practice (Appendix B) and forty questions were used for the teaching profession (Appendix C). In terms of content and format the two sets of questions were similar, but not identical. The primary condition for the inclusion of the same questions in both interviews was an intention to cover particular areas of interest. For example, it seemed plausible to ask the participants to express their opinion about important issues such as cooperative learning, education at CETT, pair or team teaching and cooperation during the teaching practice (Appendix B and Appendix C). Important considerations influenced the differences. The selection of appropriate questions depended on the participants' experience. Admittedly, trainees bring their own experiences to the learning to teach process. They also possess some knowledge about teaching in general and team teaching in particular. However, some time is needed to collect ideas in order to be able to express opinion. This consideration determined the nature of the first set of questions (Appendix B). As for the people working in the teaching profession, the knowledge about relationships becomes a real experience.

It is, therefore, crucial to trigger discussions about work experience and coopera-
tion in the teaching profession (Appendix C).

Furthermore, an attempt was made to ensure reliability, validity and credibil-
ity of the research, therefore, the qualitative interview schedule was improved in
the process of validation (Barócsi, 2005). The methods of validation were based on
several recent publications (Alderson & Banerjee 1996; Block, 1998; Brown, 2001;
De Capua & Wintergerst., 2005; Kormos, Hegybíró-Kontra, & Csölle, 2002; Petric
& Czárl, 2003). In order to support the research findings, particularly data trian-
gulation was used (Mc Groarty & Zhu, 1997). With the intention to produce a
picture as complete as possible, information from the transcripts of the interviews
was triangulated with other sources of data. The interview transcripts were trian-
gulated with one student teacher's diary and observational field notes taken dur-
ing classroom observation. Due to the limitations of space, the transcripts of the
interviews, the student teacher's diary and the observational field notes were not
included in this paper; however they may be available through personal contact
with the author. The researcher's field notes contributed to the better under-
standing of the processes in the classroom, whereas the entries from the student
teacher's diary were included for the purpose of receiving more reflections to gain
as much knowledge as possible about the particular areas. Apart from these
sources of collecting data, triangulation within the main research will also include
data collection from tape-recorded planning sessions during the period of teach-
ing practice as well conversations with the previous students' colleagues.

Procedures

Data were collected through two audio-recorded semi-structured interviews
(September 16 and September 20, 2005). After finding out about the participants'
availability, the time was fixed. I met the informants personally at a mutually con-
venient time. The interviews took place in a classroom at Deák Ferenc Secondary
Grammar School. The same procedure was followed in the two interviews. Strict
confidentiality and anonymity were guaranteed before and throughout the proce-
dure. Before the interviews, there was a friendly conversation for the purpose of
establishing relaxed atmosphere and better rapport with the respondents. After
the warm up, I explained the purpose of the interview and introduced the proce-
dure. The respondents were offered to read the questions before the actual inter-
view, however, only Respondent 1 took the opportunity and had a quick glance at
the interview schedule. The respondents could listen to the interview questions
and not see them. There were no time limits and the respondents could talk as
long as they wanted to, thus the respondents were given the fullest opportunity to
speak. When the interviews were completed, the respondents were asked to give
feedback about the interview questions. As this was not tape recorded, I took
notes about the respondents' observations. After the respondents left, I wrote
down my impressions. With regard to the interviews, in order to get meaningful
results, they were recorded and transcribed. The duration of the interviews was
twenty and seventy minutes respectively. The language of the interviews was

English as this was the language used for discussions throughout the teachers' practices.

Data analysis

An inductive approach to data analysis was adopted. Hypotheses were not generated prior to data analysis. The data were analysed using the constant comparative method described by Maykut and Morehouse (1994). This method involves many stages. First, it was necessary to prepare the data for analysis. The raw data were transferred into readable form, namely the audio-taped interviews were transcribed and the diary was typed. The researcher's classroom observation field notes were not typed since they were suitable for interpreting the information and many of the comments were redundant. This stage was followed by coding and organizing the data. The type and the source of the data were included and chunks or units of meaning were identified. The more work was done the more efficient coding became. Furthermore, some physical arrangements were necessary to help the process of identifying the emerging themes. This stage of discovery also meant that the concepts were arranged.

Again, the constant comparing method was implemented with inductive category coding, refinement of categories, relationships and patterns across categories, integration of data, and finally writing up the research. As the interview sets of questions referred to two different contexts, the analysis of the data proved very complex and time-consuming but this could not be avoided because of the different settings the respondents came from. The analysis of the researcher's classroom observation field notes and the student teacher's diary proceeded in the same way as for the previous data. The reason for following the same approach was the nature of the data. The different sources required much more time to handle.

Findings

As mentioned above, the questions as a whole were designed to cover particular areas of interest. Thus, analysis of the data revealed a number of major topics. Apart from confirming the information presented in the review of the literature, the pilot study came up with several findings.

Cooperative learning

This section gives a summary of the main points of the results from the interviews. First and foremost, the respondents were asked to reflect on cooperative learning related to their own school experience as learners in general and learners of languages. In recent years, there has been much discussion about how crucial the role of cooperative learning is in the teaching/learning process. It was tempting to ask

about cooperative techniques in order to attribute the findings to the concept in teacher training and the teaching profession. The interviews underlined the point that cooperative tasks are less popular in secondary schools than at the CETT, where plenty of them are used. However, there is a need to distinguish between learning a second language and learning other subjects, where such techniques are applied less. The language environment can be an important area of difference between other foreign languages and English, as the latter is taught in a more communicative way. Although cooperation might evoke uncertainty at an early stage or even mean bad experience for some people, eventually it seems to take positive attitude as it creates better opportunities to plan, collect ideas or solve problems. With respect to sharing responsibilities, cooperative learning promotes confidence and security. Cooperation is seen useful for successful task completion. The responses to the questions showed that the process of cooperation in learning is perceived as dependent on the people involved and their opinions.

Cooperation in teacher education

The findings that follow draw primarily on the interviews but are supported by the researcher's classroom observation field notes and the student teacher's diary. The participants in the pilot study were asked to define team teaching. At CETT, another term is widely used, pair teaching, originating in the fact that two student teachers work together. Both respondents named the form of teaching as pair teaching instead of team teaching, though both terms were known. On the whole, the respondents defined team teaching as a mode of teaching involving not only one, but two, three or more teachers, who in order to teach a group of learners, work together before, during, and after classes. Obviously, giving a definition is not an easy task to perform. It is even more challenging to define a complex issue in an interview. On the one hand, the purpose may not be immediately evident; on the other hand, an interview demands spontaneous responses that may be deprived of thoughtfully laid out details. That was the reason why an additional question was included. To ask about the areas that team teaching covers was done with the purpose of clarifying the definition of the concept. The findings hold significant applications that apart from the actual teaching process, team teaching involves collaboration in observation and evaluation.

To seek for a definition of team teaching was closely related to my purpose to look at the teacher trainer's role in the process. The findings suggest that the teacher trainer is perceived as effectively performing an important role. This may reflect the fairly widespread view that the trainer has a key role in the professional development of student teachers. In addition, the participants' reflections reveal the assumption that the teacher trainer is an inner outsider. One of the respondents expresses the idea that the teacher trainer is very much in the team because the trainer is present in almost all lessons and takes part in planning and discussions. The conclusions have implications for the proposal to focus on a slightly different aspect, namely that when regarding team teaching in the context of teaching practice, the teacher trainer is also a part of the team. In many ways, it

then means passing the boundaries of pair teaching and putting a stronger emphasis on the element of cooperation.

In terms of the teaching practice, as it appears from the data, team teaching is extremely helpful. Pre-teaching collaboration, actually means regularly to plan lessons, collect or discuss materials and brainstorm ideas, Preparing together with a partner is a very useful experience. Even though it can take more time and effort to coordinate arrangements and all the work, it opens up possibilities in sharing ideas and problem solving. This opinion actually points at one important factor-interaction that takes place when working together. Apart from collaboration before, the process also requires in-class work. An obvious remark to make when undertaking a detailed processing of the data is that when turning to collaboration in the classroom, it seems regarded as an important area. Whether considered the most important area or not, the findings show that much weight is placed on it. In addition, another fact which clearly emerges is that partners can facilitate each other not only before but also during the lesson, which promotes confidence and security. The major concerns are over the ease of conducting the lesson with a partner. The important ways of reducing the workload in the lesson, among many others are seen as follows: distribution of tasks, execution of activities, giving instructions, writing on the blackboard and also dealing with the tape recorder. In this respect it promotes effective classroom management. Furthermore, on the issue of post-teaching collaboration, the positive remarks indicate how useful it is that more people work together. This fact shows the real value of giving trainees the chance to raise awareness when reflections and ideas are developed in the course of regular discussions, which is a unique opportunity when working in a team.

The findings of the study reveal the beneficial aspects of pair or team teaching in teacher education but also shed light on possible difficulties that may arise. Difficulties of any kind might be perceived as a natural outcome of the complex interactions in the process. Clearly, much depends on the partners' compatibility and personality. A team is not able to function if it consists of uncooperative individuals but cooperation can be improved. Another problem is related to the students. Shared lessons can be a disadvantage from the students' point of view. The purpose of referring to the difficulties is twofold: to identify areas where action needs to be taken and to paint a more realistic picture of this form of teaching.

With respect to collaboration between trainees, a number of important issues emerge from the above analysis. First, although sometimes time-consuming and difficult, pre-teaching collaboration is a very useful experience for the trainees. Second, in-class collaboration, as far as team taught lessons are concerned, is very helpful for the student teachers. Third, post-lesson collaboration is crucially important and promotes trainees' professional development. In addition, attention is given to the importance of sharing a common task and responsibilities. The evidence from the results of the study raises the question of how collaborative relationship and interpersonal skills are essential.

Cooperation in the teaching process

As indicated so far, establishing a system of cooperative teaching in the context of teaching practice holds a lot of benefits and advantages regarding the development of trainees. Above all is the fact that trainees learn a very important lesson for their future work: the ability to cooperate. The data about this main topic come from the interview with Respondent 2. The findings, which have been derived from the interview, indicate that there is a need for collaboration not only in teacher preparation but in the teaching profession as well. According to the respondent, teachers do not like working in close partnership and they usually work alone, as they feel that it is an extra problem to discuss issues with the other teachers. However, it is worthwhile making the effort since cooperation holds real values for the ones who are willing to share difficulties and success. It is extremely valuable that during school competitions and examination periods teachers can work together to design and correct test papers. Combined experiences give a sense of achievement and ensure a better opportunity for problem solving, organising ideas and offering help and advice. The responses show that mutual work is particularly helpful outside the classroom as for school events, programs and summer camps when teachers work together; teachers and students work together and also students work together. As for the respondent's generally positive impression of cooperation, the following issues deserve consideration: cooperation is vital as it helps self-evaluation, gives reinforcement and feedback; it makes processes more memorable, helps both cognitively and interpersonally, enables division of responsibilities; develops problem-solving skills as well as creative thinking and communication. In this respect the results of the second interview supports the idea that cooperation is a means of development, which is a basic component of working as a teacher. Comments made such as 'If I consider myself successful, it is because I can feel that I am developing', seem to indicate that professional development is the key to success.

The data reveal that apart from advantages, cooperation with others may possess drawbacks. In a mutual work situation certain problems may arise. Some problems are relatively easy to solve, requiring a minimum of patient rational discussion. Others are more deep-rooted and may lead to a difficult confrontation. The findings of the research imply that relationships with others may pose problems because people are always different and sometimes more dominant. Surveying the participant's responses with the purpose to identify important factors which interact in cooperation in the teaching profession, the most important factors are being addressed as follows. First, combined experiences demand members to possess particular qualities. In order to cooperate they should possess abilities to take responsibility. In addition, one has to be able to listen, discuss and negotiate. To emphasise the qualities even more, it is necessary to tolerate the other person's ideas and compromise. Personally, this indicates that it will be worthwhile to find more about the necessary qualities teachers need in order to cooperate smoothly and efficiently. Second, a potential valuable factor within cooperation is the interaction among people of diverse abilities and characters. Establishing patterns of interaction is seen important in order to cooperate successfully. Third,

mutual work depends on the common task and its nature. Moreover, it is essential that members possess similar attitude to the task. Fourth, working together requires development of problem solving strategies. Clearly, relationships are not without problems, but if there are any, an attempt should be made to solve them. Fifth, it is important to develop effective strategies and skills for cooperation. From the findings listed above we can conclude that if we wish to cooperate well, we need to have a better understanding of how it is done.

Conclusion

The pilot study, broadly speaking, was found to be extremely useful and effective. The main advantage of the current study was that it proved that continued research would be feasible and the directions taken plausible. The instrument for the research was reliable and could be used for further study in this area. The outcomes increase awareness of the crucial role of cooperation and highlight the effective processes in teacher education. The findings of this research call for a more careful consideration of the question of pre-service education and in-service development often regarded as separate areas. The study also opens prospects for research on the relations between pre-service education and in-service development. It increases awareness of the pre-service — in-service continuum of career-long professional development. The findings reveal that it would be useful to consider the individual practices in in-service education as collective ones, which would certainly be relevant to the teaching profession.

A further issue with the pilot study was that some problems did arise. The literature I explored showed that it was extremely difficult to define the notion of cooperation. It may be said that cooperation is a wide area that is most attractive but difficult to cover, therefore, the major problem is associated with the broad and complex nature of the actual field of research. Its broadness is certainly an attraction but can also be considered as a problem; therefore, it was necessary to achieve a clearer focus of the research. A further important challenge was to use triangulation and find appropriate means to put more weight on it. My main concern was about the issue of time as I had to cope with the two spheres that of the research and teaching at the same time. Nevertheless, the process was extremely useful as it helped to improve the link between the researcher and practitioner. Data collection with the semi-structured interviews proved to be time-consuming; however, the process was very rewarding as it allowed great opportunities to get insights into my own work as a teacher trainer and extended my knowledge about mentoring.
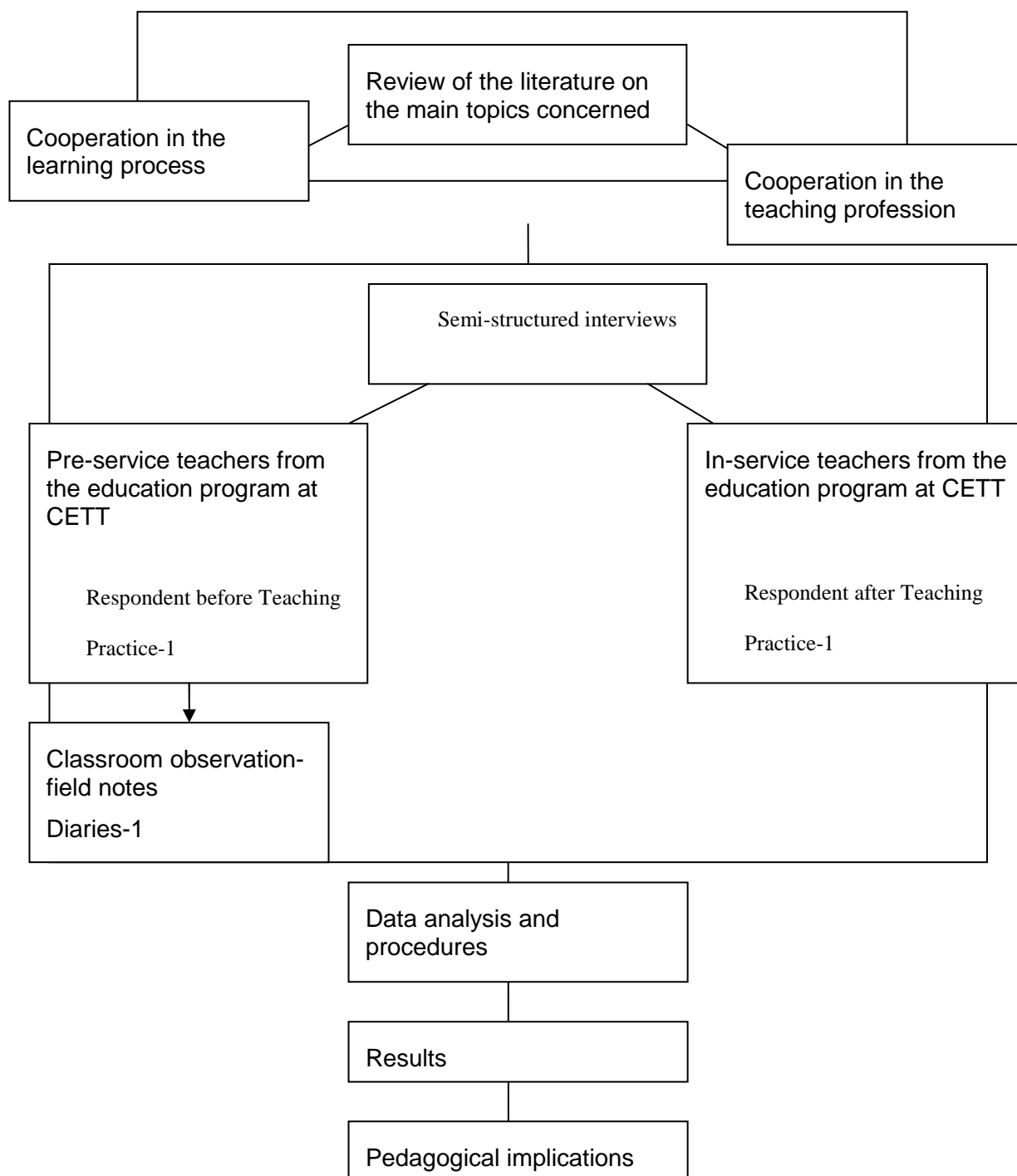
References

Alderson, C. J., & Banerjee, J. (1996). *How might impact study instruments be validated?* Unpublished manuscript commissioned by UCLES.

Barócsi, S. (1998). *The positive effects of team teaching in the training of Hungarian teachers of English*. Unpublished master's dissertation, University of Manchester, Manchester, UK.

Barócsi, S. (2005). *Validation of a long qualitative interview schedule*. Unpublished seminar paper, Eötvös Loránd University, Budapest, Hungary.

Block, D. (1998). Exploring the interpretation of questionnaire items. *System, 26* (1), 403-425.

Bodóczky, C., & Malderez, A. (1994). Talking shop: Pre-service teaching experience and the training of supervisors, *ELTJ, 18* (1), *66-73.*

Brandes, D., & Ginnis, P. (1992). *The student-centred school*. Basil Blackwell: Simon and Schuster and Education.

Brown, J. D. (2001). *Using surveys in language programs*. Cambridge: Cambridge University Press.

Claxton, G. (1989). *Being a teacher: A positive approach to change and stress*. London: Cassell.

De Capua, A., & Wintergerst, A. (2005). Assessing and validating a learning styles instrument, *System, 33* (1), 1-16.

Edge, J. (1992a). Co-operative development. *ELTJ, 46* (1), 62-70*.*

Edge, J. (1992b). *Co-operative development*. London: Longman.

Freeman, D. (1989). Teacher training development and decision making: A model of teaching and related strategies for language teacher education. *TESOL Quarterly 23* (1), 14-26*.*

Goddard, R., Hoy, W., & Woolfolk Hoy, A. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher, 33* (3), 3-13.

Gower, R., & Walters, S. (1983). *Teaching practice handbook*. Oxford: Heinemann.

Gwyn-Paquette, C., & Tochon, F.V. (2002). The role of reflective conversation and feedback in helping preservice teachers learn to use cooperative activities in their second language classroom. *Modern Language Journal, 86* (2), 204-226.

Hornby, A., (1994). *Oxford advanced learner's encyclopedic dictionary* (2nd ed.)*.* Oxford: Oxford University Press.

Johnson, R.B., & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33* (7), 14-26.

Kagan, D. M. (1992). Professional growth among pre-service and beginning teachers. *Review of Educational Research, 62* (29), 129-169.

Kennedy, J. (1993). Meeting the needs of teacher trainees on teaching practice. *ELTJ 47* (2), 157-165.

Kerry, T., & Mayes, A. (1995). *Issues in mentoring.* London: The Open University Press.

Knezevic, A., & Scholl, M. (1996). Learning to teach together: teaching to learn together. In D. Freeman & J. C. Richards (Eds.), *Teacher learning in language teaching* (pp. 79-96). Cambridge: Cambridge University Press.

Kormos, J. Hegybíró-Kontra, E., & Csölle, A. (2002). Language wants of English majors in a non-native context. *System, 30* (1), 517-525.

Lazaraton, A. (2003). Evaluative criteria for qualitative research in applied linguistics: Whose criteria and whose research? *Modern Language Journal, 87* (1), 1-12.

Liang, X, (2004). Cooperative learning as a sociocultural practice. *Canadian Modern Language Review, 60* (5), 637-668.

Maykut, P., & Morehouse, R. (1994). *Beginning qualitative research: A philosophic and practical guide*. London: The Falmer Press.

Mc Cracken, G. (1988). *The long interview*. London: SAGE Publications.

McGroarty, M., & Zhu, W. (1997). Triangulation in classroom research: A study of peer revision. *Language Learning, 47* (1), 1-43.

Medgyes, P. (1994a). *The non-native teacher*. London: Macmillan Publishers.

Medgyes, P. (1994b). 'Natives' and 'non-natives': Much ado about something, *IATEFL Newsletter, 126, 8-9.*

Medgyes, P., & Malderez, A. (1996). *Changing perspectives in teacher education*. Oxford: Heinemann.

Medgyes, P., & Nyilasi, E. (1997). Pair teaching in pre-service teacher education. *Foreign Language Annals, 30* (3), 352-368.

Nunan, D. (1992). *Collaborative language learning and teaching*. Cambridge: Cambridge University Press.

Petric, B., & Czárl, B. (2003). Validation of a writing strategy questionnaire. *System, 31* (1), 187-215.

Slavin, R. (1985). An introduction to cooperative learning research. In R. Slavin, S. Sharan, S. Kagán & R. Lazarovitz (Eds.), *Learning to cooperate, cooperating to learn* (pp.5-14). New York: Plenum Press.

Slavin, R., Sharan, S., Kagan, S., & Lazarovitz, R. (1985). *Learning to cooperate, cooperating to learn.* New York: Plenum Press.

Swan, J. (1993). Metaphor in action: The observation schedule in a reflective approach to teacher education. *ELTJ 47*(3), 243-249.

Underwood, M. (1987). *Effective class management*. London: Longman.

Wallace, M.J. (1991). *Training foreign language teachers.* Cambridge: Cambridge University Press.

Wilhelm, K.H. (1997). Sometimes kicking and sometimes screaming: Language teachers-in-training react to a collaborative learning model. *Modern Language Journal, 81* (*4*), 527-541.

Appendix A

Design for the pilot study

```
                    ┌─────────────────────────────┐
        ┌───────────│ Review of the literature on │──────────┐
        │           │ the main topics concerned   │          │
┌───────────────┐   └─────────────────────────────┘   ┌──────────────────┐
│ Cooperation in │                                     │ Cooperation in the│
│ the            │─────────────────────────────────────│ teaching profession│
│ learning process│                                    └──────────────────┘
└───────────────┘

                    ┌─────────────────────────────┐
                    │   Semi-structured interviews │
                    └─────────────────────────────┘

┌─────────────────────────┐              ┌─────────────────────────┐
│ Pre-service teachers from│             │ In-service teachers from │
│ the education program at │             │ the education program at  │
│ CETT                    │             │ CETT                     │
│                         │              │                          │
│ Respondent before Teaching│           │ Respondent after Teaching │
│                         │              │                          │
│ Practice-1              │             │ Practice-1               │
└─────────────────────────┘             └─────────────────────────┘

┌─────────────────────────┐
│ Classroom observation-   │
│ field notes             │
│ Diaries-1               │
└─────────────────────────┘

            ┌─────────────────────────┐
            │ Data analysis and        │
            │ procedures              │
            └─────────────────────────┘

            ┌─────────────────────────┐
            │ Results                 │
            └─────────────────────────┘

            ┌─────────────────────────┐
            │ Pedagogical implications │
            └─────────────────────────┘
```

Appendix B

Questions for the semi-structured interview before teaching practice-
pilot study September 16, 2005

Background data
1. What is your name?
2. Are you male or female?
3. How old are you?
4. What qualifications do you have? (diploma or degree)
5. Do you have any other degrees?

Participants as learners in general, in languages
1. Can you recall any cooperative tasks from your own school experience as a learner? (think of your school years or when you were a student at university)
2. When and how did your language teachers use cooperative techniques? (think of a story, involvement in projects, pair work or group work in class, so you as a learner)
3. How did you relate to them? (recall your attitude: How did you feel? Were you aware of what you were doing? Was it a meaningful experience then?)

Questions on CETT education
1. When did you begin your studies at the Centre for English Teacher Training in Budapest (CETT)?
2. What programme(s) are you in?
3. How long will your Teaching Practice last at CETT?
4. Which school will you do your Teaching Practice at?
5. Who will you do your Teaching Practice with?

Questions on pair or team teaching
1. How can you define pair or team teaching?
2. What do you think pair or team teaching will involve? (planning, conducting the lesson, pre-post- lesson discussions, work with teacher trainer)
3. When did you first hear about team teaching?
4. Where did you hear about it?
5. Have you ever had the opportunity to observe pair or team teaching? When? Where?
6. What impression did team teaching make on you then?
7. Have you ever had the opportunity to experience pair/team teaching?
8. How long are you going to conduct the lessons together?
9. How do you think the shared lessons will be different from the lessons conducted individually?
10. What do you think the students' attitude towards pair/team teaching will be?

11. What do you think the advantages of pair/team teaching are?
12. What do you think the disadvantages of team teaching are?
13. How can you avoid the disadvantages?

Questions relating cooperation
1. How do you expect to cooperate with your partner and your teacher trainer this year?
2. How do you expect to benefit from the cooperation with your partner and your teacher trainer?
3. Do you consider your teacher trainer as a member of the team?
4. How many people can successfully cooperate in your view?
5. Can two persons cooperate best? (Is that a team? How many should work together?)
6. In what situations can you cooperate easily? (What do you consider vital for cooperation to work?)
7. How are you going to contribute to the work of your team? Would you make efforts? Would you offer your help?
8. Do you think cooperation is necessary in a career? If yes, where?
9. Would you stimulate cooperation? Why? How?
10. What are the advantages of cooperation in your experience?
11. What are the disadvantages of cooperation in your experience? (domineering personality, creating a sense of direction)
12. How can you overcome them?
13. Is there anything else that you would like to add?

Appendix C

Questions for the semi-structured interview-pilot study September 20, 2005

Background data
1. What is your name?
2. Are you male or female?
3. How old are you?
4. What qualifications do you have?(diploma or degree)
5. Do you have any other degrees?

Participants as learners in general, in languages
1. Can you recall any cooperative task from your own school experience as a learner? (Think of your school years or when you were a student at university)
2. When and how did your language teachers use cooperative techniques? (think of a story, involvement in projects, pair work or group work in class, so you as a learner)
3. How did you relate to them? (recall your attitude: How did you feel? Were you aware of what you were doing? Was it a meaningful experience then?)

Questions on work experience
1. Where do you work?
2. How long have you been working for this company/school?
3. Do you have a full-time or a part- time job? (reasons)
4. Did you have any other jobs before your current job?
5. Does the job meet your requirements? What do like about it?
6. What were your previous jobs like? (Why did you leave this job? What did you like about it? What didn't you like about it?)
7. To what extent do you consider yourself successful in your career? Why? (weaknesses and strengths)
8. What is your private tutoring experience? (Who? Where? Why? Advantages?)

Questions on CETT education

1. When did you attend the Centre for English Teacher Training in Budapest (CETT)?
2. What programme(s) did you graduate from:
3. How long did your Teaching Practice last at CETT?
4. Which school did you do your Teaching Practice at?
5. Who did you do your Teaching Practice with?
6. How long did you experience team/pair teaching?
7. How can you define pair/team teaching?
8. What did team/pair teaching involve? (planning, conducting the lesson, pre-post- lesson discussions, work with teacher trainer)
9. Looking back, what is your opinion of team/pair teaching? (positive sides, negative sides, benefits, drawbacks)

Questions relating cooperation

1. Describe how you generally cooperate in your current job?
2. Describe one particular cooperation in your current job?
3. How do you need to cooperate with others in your current job? (In what specific ways do you cooperate with others?)
4. Why do you need to cooperate with others in your current job?
5. What skills are needed in order to cooperate with other(s)? (share knowledge, negotiate, listening skills, working together to accomplish a task)
6. Can you give an example when you felt that cooperation was really beneficial to you? (think of a story)
7. Can you recall a situation when you failed in cooperation? Why? (think of a story)
8. In what situations do you think cooperation is useful? Why?
9. Who can be involved in cooperation? (I mean colleague-colleague, teacher-teacher, EFL students cooperation?)
10. How many people can successfully cooperate in your view?
11. Can two persons cooperate best? (Is that a team? How many should work together?)
12. In what situations can you cooperate easily? (What do you consider vital for cooperation to work?)
13. What are the advantages of cooperation in your experience?
14. What are the disadvantages of cooperation in your experience? (domineering personality, creating a sense of direction)
15. How can you overcome them?

Is there anything else that you would like to add?

The Importance of the Group: A Case Study of a University First-Year Academic Skills Class

Francis J. Prescott

Eötvös Lóránd University, Budapest, Hungary

pompas@mail.tvnet.hu

Introduction

As part of the gradually increasing focus in language pedagogy research put on the importance of the teacher's role and what actually goes on in the classroom since the 1980s (Clarke, 1994), interest in group dynamics has grown considerably over the last decade. Group dynamics as a scientific approach began in the 1940s with the work of the social psychologist Kurt Lewin and was applied in many areas of sociology and psychology (Dörnyei & Malderez, 1999). It is only recently that educationalists have taken an interest in this area, partly as a result of the attention given to communicative language teaching and co-operative learning (Dörnyei & Malderez, 1997; Dörnyei & Murphey, 2003;).

The importance of group dynamics for language learning stems from the view of the classroom as a site of social interaction. Prabhu (1992) points out that in addition to being a unit of the curriculum and the implementation of a method of teaching, a classroom lesson is also a social event and an arena of human interaction. This means that, due to the potential conflict of these four aspects, it is necessary to consider not just the pedagogic dimensions but also the social ones in order for effective teaching to take place. The teacher's role is obviously central to this, not least because her or his personality plays a crucial part: "the teacher's own personality is a major factor in the interplay of forces, and conflict resolution will necessarily have to vary from one teacher to another" (Prabhu, 1992, p.231). This also means that research aimed at discovering the most effective teaching method is mistaken and researchers should rather concentrate on the complex interactions of actual classroom practice.

Dörnyei and Malderez (1997) maintain that an approach based on group theory can yield a much greater understanding of what actually takes place in classrooms and has great relevance to current educational practice such as co-operative learning. Dörnyei and Murphey (2003) also underline the importance for teachers of spending time on group building in order to prevent crises developing and thus to reduce teacher stress and possible burnout in addition to promoting learning: "Learning about group dynamics and organising well-functioning groups will go a long way toward facilitating smooth classroom management and enhancing student performance" (p. 11).

This paper explores how the use of group building techniques in one particular class can positively affect the classroom experience of both students and teacher. It also attempts to discover how group building can contribute to effective learning. The paper presents a case study of a university first-year academic skills course.

Aim and rationale of the study

The present study grew out of a larger research project exploring the academic skills preparation course for first-year students entering the School of English and American Studies (SEAS) at the Eötvös Loránd University in Budapest. In an initial phase of this project two classes were observed over five weeks and the teachers interviewed. One of these classes proved to be of especial interest because of the emphasis the teacher put on the value of group building in order to promote more effective learning. There were several other elements of the teacher's approach which were interesting and unusual in the context of this course but these were all related to the teacher's belief in the importance of creating the right atmosphere for learning and in making the students feel secure and confident.

Because of the special nature of this class, it seemed worthwhile to develop the research using the existing data and supplementing it with some further data (see data collection below) into an individual case study. The aim of this case study was to explore what was being achieved through group building and to attempt to uncover what the benefits might be in terms of both the students' learning experience as well as the teacher's experience of teaching. It is hoped that, although obviously this is a single unique case, it will add to the existing body of work on group dynamics (there is very little research describing group processes in actual L2 classrooms, Senior, 1997) and may be of practical use to other teachers who recognise similarities with their own teaching contexts.

Background to the study

The class which forms the basis of the study was part of a preparatory academic skills course designed to help first-year students reach an acceptable level of competence for their studies with a particular emphasis on writing. This is a once weekly orientation course over two semesters with each class lasting 90 minutes (in practical terms this gives just 42 contact hours over the whole course). The course is taught in English by both Hungarian and native English speaking teachers – the teacher in the case study was a native teacher.

The university itself is scattered all over the city which means that students are often in a hurry to get to other departments and tend not to spend very long in one place. This leads in turn to a lack of strong identification with the university or any particular part of it and probably has consequences for the way students feel in the classroom as well. Moreover, because the majority of students have different timetables, they tend to be in classes where very few, if any people know each

other, and the only time they meet is in the class. Teachers also tend not to form close bonds with students who they only see for a few hours a week.

This background seemingly makes the forming of a strong group identity an uphill task to say the least. It is not surprising that many teachers do little to challenge the status quo and settle for conducting classes in which there are only superficial relationships and often members of the class, and sometimes even the teacher, do not know each other's names. Indeed, several of the students in the study identified this lack of intimacy in other classes as being one reason why they liked the academic skills class:

> I think it was very good to come each morning because it was kind of a different class, not the same as going to literature or linguistics because here the atmosphere was really good and we know each other quite well. (Student Interview 3 (SI3))

> This was the only class I liked going to. I think I need at least a few classes like that. That I just like going to and enjoy doing. It was down to the good atmosphere basically. And the group, basically. (SI 4)

Methodology

A case study was considered to be the best approach since the aim of the research was to examine a complex but delimited social situation in order to understand what was taking place and what it meant for the participants. This places the research solidly within the qualitative domain and more specifically within the tradition of case study research, which, as Creswell (1998, p. 61) notes, involves "an exploration of a "bounded system" or a case (or multiple cases) over time through detailed, in-depth data collection involving multiple sources of information rich in context". In this particular instance the bounded system was the classroom and its participants and their interactions. Because the motivation for the study was one of interest in the particular case itself, rather than in a general phenomenon or in the typicality of this case as being just one of many similar cases, this meant that it was an *intrinsic case study* (Stake 1995), worthy of study in its own right.

Data collection

In order to obtain the required depth and richness of data, multiple sources were used over a sustained time span. These are listed below:

- 5 weeks of classroom observations
- informal discussions with the teacher after each lesson observation
- a long interview with the teacher at the end of the first semester

- brief department feedback sheets from the students at the end of the first semester
- short interviews with 5 students at the end of the course.

The classroom observations were begun in an unstructured way to allow features of the setting and patterns of interaction to emerge clearly. Of course, as a teacher of the academic skills course for seven years myself, it was impossible to avoid having some preconceptions, at the very least on a subconscious level. However, by using multiple data sources and by presenting the participants' own words in the study, it was intended that their voices would emerge sufficiently clearly for the reader to be able to discriminate and judge between the various perspectives of participants and researcher. In qualitative research, subjectivity, traditionally frowned upon as threatening validity and reliability, is not seen as a problem but rather as something that has to be recognised as an integral part of the human condition. Maykut and Morehouse (1994) address this issue through the term 'perspectival':

> we have chosen to use the word perspectival instead of subjective to refer to the way qualitative researchers see the world. Perspectival has the added advantage of being inclusive of differing perspectives, including but not limited to the researchers' perspective. (p.20)

The classroom observations began as loose descriptive notes focusing on events and activities as well as the behaviour of the participants. A note of the timing of events was also made. Immediately following observations additional notes of impressions and thoughts were made while they were still fresh. Following the first couple of observations a preliminary protocol for recording observations was constructed using the method suggested by Creswell (1994). Each page was divided into three columns in which were recorded, from left to right, the time at significant points (such as the start of an activity), what was actually taking place (details of the setting, events, dialogue and descriptions of participants – Creswell's 'descriptive notes' (p.152), and my own thoughts and ideas (Creswell's 'reflective notes' (p.152)). In addition, in the first, narrow column an initial coding system (developed from the themes emerging in the initial observations) was used to help in the later interpretation of the notes.

Shortly after each lesson observation, informal discussions took place with the teacher. These lasted from 20 minutes to almost an hour on one occasion and I was permitted to make brief notes of the main points. The agenda for these discussions was not fixed but always involved matters of interest to one or the other party – usually the teacher would begin by asking me what I thought of the lesson or whether I had any questions. My comments or queries would then trigger other points or topics in a free ranging, natural way.

At the end of the semester, the teacher agreed to do a more formal long interview which was based on an interview schedule (see Appendix A). This schedule had been developed beforehand using a process based on McCracken's four-step model (1988) for designing and implementing a long qualitative interview. The

schedule had been validated using expert respondents and some further additions were made to it as a result of issues arising during the observations (for instance, the section on teaching materials). The open-ended nature of the questions allowed respondents some freedom in answering whilst the semi-structured nature of the schedule enabled the interview to be focused on key issues. The interview was transcribed and submitted to the teacher for checking.

The teacher also allowed me to see the students' feedback slips which were filled in at the end of the first semester. These are standard department forms which are given to the students to fill in anonymously at the end of every semester. The teacher has to leave the classroom while the students fill in the forms and then one of them returns them to the department office, this way ensuring that the students feel free to express their true feelings. For the study only the three open questions on the form were used. These asked the students to say what things they liked about the course, what things they did not like and how the course could be improved.

At the end of the course I attended the final meeting which took place in the teacher's flat and asked the students to come and see me at a later date for a short discussion of their experiences and feelings about the course. For this purpose I used a short interview guide consisting of 11 questions (see Appendix B). An interview guide, in contrast to a more elaborated interview schedule, consists of a series of topics or broad questions which can be freely explored and probed with the interviewee (Maykut & Morehouse, 1994). Five out of 12 students actually participated and the interviews lasted from 20 to 30 minutes (the first interview was with two students). These were transcribed and sent to the interviewees for checking.

Data analysis

The basic approach to analysing the data was to use the constant comparative method as described by Strauss and Corbin (1990) and Maykut and Morehouse (1994). Cohen, Manion and Morrison (2000) also give a seven step sequence for data analysis which is essentially the same approach. This method involves coding the raw data into categories and then comparing and refining the categories as salient points emerge from the data. The basic coding system for the classroom observations was derived in this way, as has already been described above. In developing more refined categories from the observations and the post lesson discussions both single instances of behaviour and repetition of phenomena were considered to be important (Stake, 1995). The long interview and the student interviews, after having been transcribed, were treated in a similar way. First the data was unitised and coded and then correspondences were looked for with the observation data. Based on these correspondences more refined categories or domains including data from all phases of data collection were created and described. For the purposes of the present paper only categories relating to the group have been selected.

Description and discussion of the case

The following description, based on the notes from the third lesson observation, describes the beginning of one of the teacher's classes. It is typical of the classes which I observed.

It is 8.30 on a drizzly November morning. In the small, rather gloomy classroom the tables are pushed back against the walls and a circle of chairs has been formed in front of the blackboard. Seventies pop music is playing on the CD player ("Teenage Kicks" by The Undertones). On the floor in the middle of the circle of chairs are newspaper articles and obituaries about John Peel, a well known British DJ who died a few days ago. In the very centre a small candle flickers next to a couple of Remembrance poppies.

"Join us in the circle?" the teacher asks a student sitting apart.

A couple more students arrive as the teacher begins giving the students slips of paper with quotes about John Peel. He asks them to write down any new words or collocations and then pass their slip to the person on their left. The students begin writing and passing the slips around in silence.

The teacher asks a student what he has written down.

"To be deprived. Are you deprived of anything at the moment? A raincoat?"

The teacher asks the students to shout out the words they have written down and he begins putting them on the board, giving example sentences as he does so.

"Your generation were deprived of being an úttörő." (pioneer. The pioneers were a Communist era children's organisation similar to the boy scouts or girl guides.)

"Genuine – igazi – igazi túró rudi." (a popular sweet with many inferior imitators)

A student asks him how he likes it.

"How do I like it? It's great – we'll have a tasting, a túró rudi tasting session."

The teacher now starts talking about the funeral of John Peel which will be on Friday.

"It will be the the biggest collection of rock stars ever ... This was his favourite song, he said when I die I want this on my tombstone: "Teenage kicks are hard to beat"."

He tells the students what John Peel's death means to him and then starts going through the words on the board, explaining and occasionally asking them for the Hungarian equivalents. He asks them what to say when somebody dies.

This description gives a snapshot of what this teacher's classes were like, but to have a clearer understanding of why they were like this and what the effect on learning was, I will look at certain key aspects which emerged during the study. In what follows I will discuss the main categories relating to the group which developed during the data analysis. Inevitably, some of these categories overlap and their identification is dependent on the researcher's perspective.

The group atmosphere and relationships

In the end of term feedback forms five of the students explicitly commented on the atmosphere as being one of the things they liked about the course. Three others mentioned forming a group as something they liked, two mentioned "openness", one each said they liked the course because it was "interactive" or because of "the teacher's support", and the last student's only comment was "motivating". In addition, as already mentioned (see section 3 above), several students mentioned the positive atmosphere during the student interviews at the end of the course and all of them said they enjoyed the classes.

It seems reasonable to suppose that this positive feedback concerning the students' experience of the lessons was a direct result of the efforts of the teacher to build a good atmosphere and close relationships within the class. The literature on group dynamics places great emphasis on group formation and group development and Dörnyei and Malderez (1997) cite Tuckman and Jensen's (1977) five stage development sequence of forming, storming, norming, performing and adjourning. It is likely that, since the observations began in the middle of the first semester, this class had already arrived at the performing stage, and in retrospect it would have been helpful to see the first lessons as well. However, many clues were provided in the data as to how group development took place.

The classes always began in a similar way. The classroom was open at least 10 minutes before the lesson started and there was always music or some sort of recording being played to make the room interesting and inviting. The music usually had some connection to the lesson for that day (on one occasion an audiobook of *Ulysses* was being played because the homework from the previous week had been to write a short essay on James Joyce). The teacher himself would usually be busy setting up the lesson but would chat to the students who arrived early. The lessons took place with the students seated in a circle and the desks pushed to the sides of the room, with materials for the lesson arranged in the middle of the circle. In the interview the teacher mentioned that one of the reasons for this was to bring students closer together and counteract the effect of the classroom itself:

> I like people to be able to see each other so everyone can see into each other's eyes. So that you organise the classroom in that way, so you don't get people sitting in rows. That's important. And I like a circle, and it's possible to do that with a group of 14 people. It's a bit grim in that room, in 330 – it's better in the morning than it is in the afternoon. (TI, p. 15)

In the student interviews, when asked about the seating arrangement, all the students were positive (although one student said that she found it 'a bit difficult' to write with her file on her lap). Several students felt it was important in forming a good atmosphere:

> It was quite interesting because there were no tables between us and we were in a quite small circle. And in retrospect I think it was very, very

useful in creating a good atmosphere, because we were close to every-body, we could see each other's eyes and the physical closeness sort of contributed to the openness of everyone. So I think the seating ar-rangement is very important. (SI 4)

The teacher also made frequent use of humour in the lessons and usually the hu-mour was used whilst making a serious point. Often it was used to illustrate vo-cabulary in an amusing way, such as in the following example from a discussion on plagiarism:

Teacher (reading from the article): "To beef up your rather scant bibliography."
"A nice expression with scant is 'scantily clad' – a scantily clad woman is a woman with – *not – many – clothes – on*"
Teacher carries on reading.
(Classroom Observation 1)

Building closer relationships with and amongst the students was also fostered through e-mail communication:

They'll write to me and then they'll say 'Well can you have a look at this?' or 'I wasn't sure about something'. But I encourage that and I en-courage, I like to have a dialogue with them, for those that want it. (TI, p. 13)

Students were encouraged to communicate with the group as a whole, for instance when organising a film trip which served as the basis for one of the essays, and the teacher asked students to share anything interesting they came across by using the group mailing list. On one occasion this resulted in a list of formal linking expres-sions which was used to supplement the material the teacher provided.

Teaching philosophy

The teacher's approach to teaching was clearly driven by his personal teaching philosophy which he often mentioned both after the observations and during the end of semester interview (for the importance of teacher beliefs to classroom prac-tice see Pajares, 1992, and Woods, 1996, among many others). This philosophy was manifested in a number of ways but his explanations of these aspects were con-sistent throughout the study. A strong group identity was, of course, central to his approach:

... the group thing is important to me. I'm a very social being myself. And the belief that nothing much is going to be learnt if the atmosphere in the group's not very good. So I try lots of techniques to make it work. (TI, p. 5)

Part of the reason for having good group relationships was so that the students would enjoy working together and felt secure enough to share their ideas knowing that their contributions would be valued by the others (TI, p 4). The issue of student confidence was identified as being central both to the course and to the broader notion of a teacher's duty:

> I think it's about confidence. It's about giving them, yeah, confidence, so that they, when they approach their assignments in linguistics, or their first assignments in literature and in linguistics, that they feel more able to do them and they've had some help or they've been, we've had discussions about different ways of writing things. So that they can refer to the academic skills course when they do that. (TI, p. 2)

> ... they're first-year students and I think it is our responsibility to see ourselves not just as subject teachers, but people who can be there to help them get through their university career and make sense of it ...maybe 20, 30 per cent of the way I see teaching is what you've got in that question there. (TI, p. 15)

Making students feel comfortable working with others was just part of this.

Another aspect of his teaching that reflected this concern with confidence was the teacher's attitude to marking. After the first lesson observation the teacher stated that no marks would be given for written work until the second semester. He did not give marks because he saw the course "very much as a developmental, non-judgemental course, and I find the marks get in the way". In the case of problematic essays he preferred to write comments such as "'this is a bit disorganised because of that', or 'your grammar, really, is all over the place', or 'you made a lot of really careless mistakes with this'" and draw "a little smiley face and an exclamation mark and [my initial] at the end" (TI, pp.5-6). The students who were interviewed clearly appreciated this and none of them said they would have preferred to receive marks. They were quite happy with the feedback he gave them and could tell "whether what I wrote was good or not. I could figure it out" (SI1).

The teacher's main approach to teaching the course and to ELT in general was to promote learning through discovery: "I like to encourage a thinking-for-themselves attitude. I don't like spoonfeeding them. I don't like – my views of education are to do with discovery, self, with finding out" (TI, p.3). For this to work, the teacher was aware that developing a feeling of safety in the group was important so that students could share and discuss their ideas and work together, rather than working as individuals. He said that he believed that "learning in a classroom is a social thing and not just a group of individuals who are there processing the learning content individually" (TI, p.4). He was also aware that some students might want to be spoonfed (Ehrman & Oxford (1990) and Nelson (1995) discuss learning styles and the problems that may arise when students and teacher have differing assumptions of how learning occurs) but he believed that the value of the approach was more important:

There might be those people who are a little less secure who might want more mark feedback type thing. But that – you know, you have to make a decision. You have certain approaches to teaching and you can't please everybody. If you have an approach which involves more them thinking for themselves, then those people who are more teacher dependent, will be those people who will be more upset by that. But that's something I'm prepared to put up with, you know, for the sake of the approach. It might be right, it might be wrong. Don't know. (TI, p. 11)

Promoting student interest

One thing that was particularly noticeable about the classes during the observations was the almost unflagging interest and involvement of the students in each lesson. The students' enjoyment of the course has already been mentioned but this was not merely a matter of successful group development. In the discussion after the second observation the teacher said that he felt there had to be a reason for students to come to the class, something to interest them, particularly since the class began early in the morning (8.30) and they needed time to wake up. He usually spent time thinking about next week's lesson at the weekend. He always began the lesson with vocabulary warmers which were based on current topics of interest in their own right, such as newspaper headlines about the commemoration of the 1956 Hungarian uprising and the titles of Hungarian newspapers compared to those of British newspapers, or quotes from the obituaries of a famous British DJ who had just died. He always tried to ensure there was a link between this part of the lesson and academic writing, such as the importance of titles.

This use of up-to-date authentic materials was characteristic of the teacher's lessons and the students seemed to be more stimulated by tasks based on extracts from real sources, such as a comprehension text on plagiarism from a humorous guide to student life or a paraphrase of a short biography of Noam Chomsky taken from the front of one of his books. Several of the students specifically mentioned learning about English culture and the 'intriguing topics brought up' as being things they liked on the feedback forms. In the interview one student said "he brought us interesting topics, not the usual ones that you can learn all the time in courses, but connected to everyday life" (SI3).

Teacher personality

The importance of the teacher's personality in influencing the 'interplay of forces' in the class has already been mentioned (Prabhu, 1992, p. 231). Undoubtedly it is a crucial factor in the dynamics of a classroom group and one which is not fully acknowledged by the current literature. This teacher obviously had a very strong personality but it was one the students seemed to appreciate. He clearly enjoyed teaching the course and as he himself said, part of the reason it was important to have a good group was because he was a social person himself.

The teacher's social and open personality was reflected in the way he was ready to share his interests and feelings with the students in the form of music and books and topics which had a personal motivation, such as the death of someone who had meant a lot to him when he was a student. He was also very open about his own life and experiences, for instance he used his own school reports for a discussion on grading.

He was very energetic in his teaching. His lessons had a good pace – they never flagged and although he often stepped back from the circle or sat and listened to what students were saying he was very much the leader and prime mover. This was the one area where he showed signs of discomfort in post observation discussions – he did not like to see himself as taking control all the time but was conscious of the need to maintain a good pace. However, the students did not seem to have any problem with his personality. They spoke of him as being dynamic and fun (one girl said that he might be a bit too dynamic for some students but not for her) and were keen to praise him as a teacher.

Students' feelings about their learning

Although there was no direct evidence in the study to demonstrate the effectiveness of the learning taking place in the class, there were a lot of indications which were suggestive of it. The fact that all the students felt positive about the group was one indication, since, as Dörnyei and Murphey point out, "the class group can have a significant impact on the effectiveness of learning" (2003, p. 3).

The students also expressed positive feelings about their own development. The response of one of the weakest students (possibly the weakest according to the teacher) serves as a good example:

I: Did you feel that the classes helped you develop your language?
S4: Yes. Vocabulary for example and I felt from [the teacher] that he encourages us. I was not the type of speaking person all the time but still in this position I felt more confident to speak up and he showed me to be more confident. (SI 3)

The opportunity to talk in classes was mentioned as being helpful by more than one student.

There was also evidence indicating that the classes had made the students much more aware of the importance of teaching methodology. Several students talked about the teacher's methods as having been motivating for them. Students showed that they were aware of the purpose behind many of the techniques that the teacher used and in at least one case this had led to reflection on teaching and the student's own future practice as a teacher. It has been suggested by Wenden (1987) that critical reflection on the learner's own learning context is necessary for the learner to become autonomous.

Conclusion

It seems reasonable to suppose that the good classroom dynamics in this class had a strong positive effect on the learners' attitudes to the course and to their own learning, as well as making the experience of all participants more enjoyable. The study shows that the possibilities for creating good dynamics, while requiring considerable effort on the part of the teacher, exist even when the circumstances or context is apparently unpromising: even in a university class that meets once a week for an hour and a half, the members of which do not meet as a group in any other class, it is possible to forge a strong positive group dynamic that encourages a questioning attitude in the members to their own learning development and their needs and expectations as learners. While it may not be possible or even desirable for all teachers to handle their classes in such a way (it is clear from this study that teacher personality was a very important factor in this class, and almost certainly would be in any class), there certainly seems to be grounds for more attention to be given to ways of developing positive group feelings as an aid to more effective learning, even in situations where this may not seem worthwhile.

References

Clarke, M. A. (1994). The dysfunctions of the theory/practice discourse. *TESOL Quarterly*, *28* (1), 9-26.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London, New York: Routledge Falmer.

Creswell, J. W. (1994). *Research design – Qualitative and quantitative approaches*. Thousand Oaks: Sage Publications.

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks: Sage Publications.

Dörnyei, Z., & Malderez, A. (1997). Group dynamics and foreign language teaching. *System*, *25* (1), 65-81.

Dörnyei, Z., & Malderez, A. (1999). The role of group dynamics in foreign language learning and teaching. In J. Arnold (Ed.), *Affect in language learning* (pp. 155-169). Cambridge: Cambridge University Press.

Dörnyei, Z., & Murphey, T. (2003). *Group dynamics in the classroom.* Cambridge: Cambridge University Press.

Ehrman, M. E., & Oxford, R. (1990). Adult language learning styles and strategies in an intensive training setting. *Modern Language Journal*, *74* (3), 311-327.

Maykut, P., & Morehouse, R. (1994). *Beginning qualitative research: A philosophic and practical guide*. UK: Routledge Farmer.

McCracken, G. (1988). *The long interview*. Newbury Park, CA: Sage Publications.

Nelson, G. L., 1995. Cultural differences in learning styles. In J. M. Reid (Ed.), *Learning styles in the ESL/EFL classroom* (pp. 3-18). New York: Heinle & Heinle.

Pajares, M. F. (1992). Teachers' beliefs and educational research: cleaning up a messy construct. *Review of Educational Research*, *62* (3), 307-332.

Prabhu, N. S. (1992). The dynamics of the language lesson. *TESOL Quarterly, 26* (2), 225-241.

Senior, R. (1997). Transforming language classes into bonded groups. *ELT Journal, 51* (1), 3-11.

Stake, R. E. (1995). *The art of case study research.* Thousand Oaks: Sage Publications.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* CA.: Newbury House.

Tuckman, B. & Jensen, M. (1977). Stages of small group development. *Group and Organization Studies, 2* (4), 419-27.

Wenden, A. (1987). Conceptual background and utility. In A. Wenden, & J. Rubin (Eds.), *Learner strategies in language learning* (pp. 3-13). UK: Prentice Hall.

Woods, D. (1996). *Teacher cognition in language teaching.* Cambridge: Cambridge University Press.

Appendix A: Interview schedule for long interview (shortened version without prompts)

A. Background questions
A1. How long have you been teaching the course?
A2. Did you have any previous experience of teaching academic writing or academic skills before teaching this course?

B. What the course is for
B1. How do you feel about the course?
B2. What do you think the main purpose of the course is?
B3. How do you explain the purpose of the course to the students? What do you say to provide a rationale for the course?
B4. At the end of the course what do you want your students to understand about academic writing and academic study?

C. How you teach the course
C1. How would you describe your approach to the course?
C2. What elements in the course (in both semesters) do you give most prominence to and why? By elements I mean both the given course contents and the methodological aspects such as reworking drafts, peer reading, analysing sample papers and so on.
C3. Are there any elements that you give less prominence to or that you consider to be of less importance?
C4. What feedback techniques do you use in the course?
C5. What use do you make of different kinds of feedback?
C6. How do you handle your own feedback to the students?
C7. Are there any special tricks or techniques that you use in teaching the course (to make it more interesting or easier to understand)?

D. Marking and correction
D1. Do you have any strong feelings about marking and correction of student work?
D2. Could you briefly describe your correction and marking policy for me?
D3. Does the way you mark or correct work change at all during the course (across both semesters)? If so, how?
D4. How do you handle written feedback on students' work?
D5. How do you decide the students' final grade for each semester?

E. Course materials
E1. What do you think about the importance of materials for language teaching?
E2. What sort of materials do you use in the course and how do you use them?

F. Dealing with problems

F1. In your experience what are the students' biggest problems in coping with the requirements of the course?

F2. Have you any idea why students might experience difficulties (if they do) ?

F3. What do you do if you think students don't understand something?

F4. Have you ever encountered any resistance or dissatisfaction from the students to do with the course and if so how do you deal with it?

F5. Do you as a teacher find the course difficult or problematic in any way?

F6. This course is based on Anglo-Saxon ideas of academic writing but it is taught in a Hungarian university to non-native students – almost all having Hungarian as their first language. How does the context of the course affect your teaching of it?

G. Classroom dynamics and student confidence

G1. How important do you think classroom atmosphere and relationships within the classroom are?

G2. How do you handle the social aspects of the course and how important do you think this is to the success of the course?

G3. These are first year students at the beginning of their university career and most of them have come from a very different academic environment in secondary school. How far do you think we are responsible as teachers for fostering the students' confidence and how does this affect how you teach the course?

G4. Do you feel that the classroom itself can have an effect on the teaching and learning taking place within it?

H. Feelings about the effectiveness of the course

H1. How successful do you feel the course is?

H2. If you could improve or change anything about the course what would it be?

I. Your own learning experience

I1. How did you yourself learn to write in English for academic purposes?

I2. How do you think your own experience of learning academic skills has affected your views of what academic writing is and how it should be taught?

Appendix B: Guide for student interviews

- How do you feel about the classes?
- Was there anything you didn't like about them?
- How did you feel about sitting in a circle?
- Did it make you comfortable?
- Did it give you the chance to write?
- How much motivation did the classes give you?
- To what extent did the classes give you confidence in the 199 exam (first year proficiency exam)?
- What effect do you feel the classes had on your language development?
- Did you feel your accuracy improved (in writing)?
- How satisfied were you with the feedback you received on your work?
- How do you feel about the marking of your work? (Would you have preferred to have a mark for every piece of work?)
- What do you think about the way your mark for the whole course was given in the first semester?
- What changes or improvements would you make to the course?
- Is there anything else you would like to say?

Explanations and Instructions in Teaching English as a Foreign Language

Sanja Čurković Kalebić

Faculty of Philosophy, University of Split, Croatia

sanja.kalebic@st.t-com.hr

Introduction

Explaining is a feature of teacher talk which is usually considered the epitome of teachers' role. There is some indication in research that learners see the ability to explain things as one of the most important qualities of a good teacher (Wragg & Wood, 1984 cited in Ur, 1996, p. 16).

Different terms are used when referring to teacher explanatory behaviour and kinds of this behaviour. Sinclair and Coulthard (1975) code explanations as "informative" acts. Chaudron (1988) distinguishes explanations of concept and procedural explanations. Explanations of concept refer to the teacher's information about grammatical rules, meanings of words, uses of expressions and so on. Procedural explanations refer to many types of structuring lesson activities. Ur (1996) uses the term instruction when referring to one particular kind of explanation which comprises the directions that are given to introduce a learning task which entails some measure of independent student activity. Todd (1997) makes a clear distinction between explanations and instructions. He identifies different purposes served by explanations. They are: to introduce new language, to introduce context, aids, content, to introduce objectives, goals, rationale, to clarify, to deal with students' questions, to treat errors (Todd, ibid). Instructions comprise a series of directives, possibly mixed with explanations, questions, and so on, which aim to get the students to do something (Todd, ibid).

Aim

The aim of this paper is to get better insight into the nature of teachers' verbal behaviour in the English as a foreign language (EFL) classroom with respect to the use of explanations and instructions in teacher talk.

The paper attempts to answer the following questions:

- What is the presence of explanations and instructions in teacher talk?
- How is the content of explanations and instructions distributed in teacher talk in general and at different levels of language learning in particular?
- What is the presence of the mother tongue in the teachers' explanations and instructions at different levels of language learning?
- Does the use of the mother tongue in the teachers' explanations and instructions depend on the level of learner language knowledge?

Database

The database consists of transcribed samples of tape-recorded EFL classroom talk in Croatian schools. The recordings were made in Split and Zadar between November 2004 and October 2005. The database is organised in two corpora.

Corpus I consists of 13 tape-recorded samples of classroom talk at the elementary level of learning EFL in elementary school (grades 1-3). The total length of recordings is 295 minutes. The recorded samples cover various types of communicative activities with four teachers.

Corpus II consists of 24 tape-recorded samples of classroom talk at pre-intermediate and intermediate level of language learning (grades 6-7 in elementary school, and grades 1-2 in secondary school). The length of all recordings is 367 minutes. The recorded samples cover various types of communicative tasks conducted by 16 teachers.

Procedures

An observation system that was developed for the purpose of qualitative and quantitative analysis of EFL classroom talk (Čurković Kalebić, 1997) was used in this paper. The system enables the researcher to discover the functional distribution of classroom talk. It consists of eight main categories; six of them are broken down in 26 subcategories. The main categories are elicititative, responsive, evaluative, corrective, conductive, explanatory, affective reactions and silence. Explanatory function is broken in two subcategories: explanations of linguistic content and explanations of extralinguistic content. Conductive behaviour is broken in instructions, nomination and discipline.

The unit of analysis in this observation system is an utterance or part of an utterance that has a different meaning from the previous one. By means of this type of segmentation of discourse we were able to identify the type of verbal behaviour whenever it occurred.

For the purpose of analysis of content of instructions we used Todd's classification of the types of content that may be included in a set of instructions (Todd, 1997). According to this classification the types of content of teacher instructions

are: goals and rationale, class organisation, roles of students, materials, proce-
dures, language aspects, time, and other aspects. Out of eight types of content
seven were discovered in our corpora.

Results and discussion

The results show that explanations and instructions are important aspects of
teacher verbal behaviour in the foreign language (FL) classroom (see Table 1 and
Table 2). They cover more than one-third (38%) of teacher verbal behaviour in
elementary teaching programmes. One-fifth (20%) of teacher talk at pre-interme-
diate and intermediate level of language teaching is realised by instructions and
explanations.

Table 1: General data – Corpus I

Total teacher turns	3,421
Teacher turns with instructions	938 (27% of all turns)
Teacher turns with explanations	383 (11% of all turns)

Table 2: General data - Corpus II

Total teacher turns	2,877
Teacher turns with instructions	359 (12% of all turns)
Teacher turns with explanations	241 (8% of all turns)

The observed features of teacher talk are not equally present in our database. In-
structions are much more frequent than explanations, particularly in elementary
teaching programmes. There are three times as many instructions as explanations
in Corpus I. These results are similar to the results obtained by Ramirez et al.
(1986, cited in Chaudron, 1988, p. 86) who found that procedural explanations (i.e.
instructions) comprised almost two-thirds of the teachers' explanations with con-
cept explanations being used in about one-third of explaining utterances in teacher
talk in elementary bilingual education programmes.

Types of explanation are not equally present in the corpora (see Table 3). At
elementary level explanations of extralinguistic content are about twice of the
amount of explanations of linguistic content. At pre-intermediate and intermedi-
ate level it is the opposite, explanations of linguistic content are about twice as
often found as explanations of extralinguistic content. The reasons for the different
distribution of the types of explanation in the corpora can be seen in different lev-
els of learners' knowledge, different cognitive abilities and different teaching
strategies that are used at the observed levels of language instruction. In teaching
young learners teachers provide learners not only with language knowledge, i.e.
with explanations of the meaning of words but with the meaning of many extral-
inguistic concepts that learners are not familiar with yet. At this age learners are
not able to understand grammar rules, i. e. explanations about the structure of the

target language system. When introducing new language items to young learners, teachers often make learners acquire new language forms and structures mechanically, i. e. by mere repetition. No grammar terminology is used at that level of language learning. At pre-intermediate and intermediate level of language learning learners need less extralinguistic explanations than learners at elementary level. They are able to understand grammar rules and grammar terminology. Teaching strategies at pre-intermediate and intermediate level significantly differ from the strategies that teachers use at elementary level. The approach to teaching is mostly cognitive and requires more creativity in learner language production (e.g. "T: We miss him. On nam nedostaje. Idete obrnuto i dobijete traženu formu. Kako ćeš reći: On nam nedostaje.? . L: I miss him.", "T: ... Znači kada je ispred ipsilona suglasnik onda se taj *wai /y/* mijenja, *Wai* /y/ changes into *ai* /i/ and *i:* /e/ *es* /s/ is added. O.K.?... Maja, how will you spell *studies*?"). That is why teachers give much more linguistic explanations at these levels of language learning.

Table 3: Distribution of explanations in teacher talk

Type of explanation	Corpus I	Corpus II	Total
Explanation of linguistic content	119 (31%)	53 (63%)	172 (33%)
Explanation of extralinguistic content	264 (69%)	88 (37%)	352 (67%)

Areas of content of teacher instructions are not equally distributed in the observed samples of classroom talk (see Table 4). Procedures are the most frequent. High frequency of procedures can be explained by the nature of communicative tasks in today's language teaching. Learners are often supposed to follow some procedures in order to accomplish a communicative task (e.g. T: Now, read the text once again, silently this time, then you are going to make a story). Instructions that refer to the use of materials are the second in frequency. A relatively high amount of this area of instructions is not surprising, since doing exercises in the coursebook is an important part of FL teaching process. Class organisation and roles of students are the third in frequency. Other areas of content are much less frequent.

Table 4: Distribution of areas of content in teacher instructing

Areas of content	Corpus I	Corpus II	Total
Goals and rationale	5%	0%	2.5%
Class organisation	15%	8%	11.5%
Roles of students	8%	15%	11.5%
Materials	20%	32%	26%
Procedures	50%	33%	41.5%
Language aspects	1%	12%	6.5%
Time	1%	0%	0.5%
Total	100%	100%	100%

In Corpus I all areas of content of instructions are present in teacher talk but they are not equally distributed. Half of the instructions covers the area of procedures. The second in frequency is instructing about the use of materials, whereas instructions about class organisation are the third in frequency. Other areas of content are not significantly present.

It is interesting to point out that *goals and rationale* and *time* are not present in teacher instructive behaviour in Corpus II. Procedures and materials are most frequent. The third and the fourth in frequency are roles of students and language aspects. Instructions about language aspect are much more frequent in this corpus than in Corpus 1. This is not surprising because at pre-intermediate and intermediate level of FL learning learners have more explicit knowledge about a FL grammar than they have at lower levels.

The mother tongue in teacher explanations is differently distributed in the observed samples (see Table 5). Croatian is slightly more frequent than English in teacher explanatory verbal behaviour at elementary level. More than two-thirds of teacher explanations at pre-intermediate and intermediate level are given in English. Unequal presence of the mother tongue in the observed teaching strategy can be explained by the unequal level of target language knowledge.

Table 5: The use of mother tongue and target language in teacher explanatory behaviour

Type of explanation	Corpus I		Corpus II	
	English	Croatian	English	Croatian
Explanations of linguistic content	59%	41%	55%	45%
Explanations of extralinguistic content	36%	64%	89%	11%
Total	47%	53%	72%	28%

Teachers mostly use Croatian when giving explanations of extralinguistic content to young learners (e.g. T: I onda su oni već bili na sigurnom u svojoj mišjoj rupi i onda su mu se rugali.). With higher graders it is the opposite, the majority of extralinguistic explanations are in English. The use of the mother tongue with explanations of linguistic content is almost equal in both corpora. This can lead us to the conclusion that the teachers' use of the mother tongue depends not only on the level of learners' language knowledge but on the type of explanation as well.

The percentages of the use of the mother tongue in teacher instructions are similar to the percentages of the mother tongue in teacher explanations in both corpora (see Table 6). In teaching young learners teachers mostly use Croatian when they explain procedures and roles of students in a communicative activity. English is most frequent when they instruct about the use of materials (e.g., T: Open your books again. Page sixty-two. OK. Now. Listen. Number three. Complete the puzzle). At pre-intermediate and intermediate level teachers give the majority of instructions about procedures in English. Also, English is widely used

in giving instructions about roles of students and class organization. The lowest use of English is in the instructions about the use of language. This finding confirms the above conclusion that it is not only the learners' language level but the type of explanation and the content of instruction that influence the teachers' choice of code. The inspection of the transcribed recordings shows that teachers very often give linguistic explanations or instructions about language aspects only in Croatian or they give them in English and then, immediately translate them into Croatian. The reasons for such behaviour might be seen in the teachers' belief that the use of the mother tongue is necessary in metacommunication.

Table 6: The use of mother tongue and target language in teacher instructive behaviour

Areas of content	Corpus I		Corpus II	
	English	Croatian	English	Croatian
Goals and rationale	40%	60%	0%	0%
Class organisation	66%	34%	75%	25%
Roles of students	30%	70%	75%	25%
Materials	86%	4%	67%	33%
Procedures	27%	73%	79%	21%
Language aspects	50%	50%	46%	54%
Time	50%	50%	0%	0%
Total	50%	50%	68%	32%

Conclusion

Several conclusions can be made on the basis of the results of the above analyses. First, the presence of instructions and explanations in teacher talk in the EFL classroom depends, among others, on the level of learners' language knowledge; at elementary level teachers give more instructions and explanations than at pre-intermediate and intermediate level. Second, instructions are more frequent than explanations at all the observed levels of language learning. Third, the distribution of the types of content of teachers' explanations depends on the level of learners' language knowledge. For example, teachers give much more extralinguistic explanations at elementary level than at other observed levels. The distribution of the areas of content of instructions does not depend on the learners' language knowledge. Some areas of content (procedures, materials) are much more frequent than other areas of content at all the levels. Fourth, the use of the mother tongue in teacher talk is highly dependent upon the level of learners' language knowledge; at elementary level teachers use much more Croatian when they explain and instruct than at pre-intermediate and intermediate level. Fifth, the use of the mother tongue in the teachers' explanations depends not only on the level of learners' language knowledge but on the type of explanations as well. For example, lin-

guistic explanations are usually given in the mother tongue at all the levels of language teaching.

Further research

The results of the above analyses and the inspection of the transcribed recordings of classroom talk show that explanations and instructions are complex types of teacher verbal behaviour which require further investigation. When giving explanations teachers use different techniques, for example, they repeat the utterance, use paraphrases, say an explanation or instruction in English and then translate it into Croatian and vice versa. The influence of these techniques upon learner language learning should be examined in order to define the patterns of teacher explanatory and instructive behaviour that enable better communication between the teacher and learners and thus, promote better language learning.

References

Chaudron, C. (1988). *Second language classroom research on teaching and learning*. Cambridge: Cambridge University Press.

Čurković Kalebić, S. (1997). On the functional distribution of talk in the EFL classroom. *SRAZ, XLII*, 95-104.

Sinclair, J., & Coulthard, R. M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford: Oxford University Press.

Todd, R.W. (1997) *Classroom teaching strategies*. Prentice Hall. International Book Distributors Ltd.

Ur, P. (1996). *A course in language teaching: Practice and theory*. Cambridge: Cambridge University Press.

Teacher Development — How To Do It?

Gordon Dobson

Eötvös Loránd University, Budapest, Hungary

gdobson@ludens.elte.hu

Introduction

This paper seeks to demonstrate a number of the possible ways in which language teachers can develop. This will be done by offering a possible definition of what teacher development is, by examining the literature pertaining to teacher development and language classroom research, and by describing what the author has done to facilitate his own development. Finally, the paper draws a number of conclusions and outlines further, and alternative, approaches to teacher development.

Background

This paper has been inspired by a general lack of teacher development in the institutions in which I have worked over the last seventeen years, with particular emphasis on the public tertiary sector. As a teacher in higher education, I am able to work, unsupervised and autonomously, in a climate of (closely guarded) academic freedom. While very much enjoying and greatly appreciating this freedom and the possibilities it offers, I am left feeling that teachers need to concentrate not only on their subject, but also on themselves as developing professionals (Crookes, 2003). For a variety of reasons, all beyond the scope of this paper, this appears rarely to be the case.

Formal teacher development, that is teacher development initiated by the institution, has generally been restricted to a very small number of compulsory and routine classroom observations, usually during the first semester at the institution concerned, and used more for purposes of quality control and assessment than teacher development. Likewise in-service teacher development has thus far been restricted to a small number of days immediately prior to the beginning of the academic year, or the very occasional short session during semester, and has emphasised the transmission of knowledge or the organisation of teaching and learning. Despite being aware of the professional and personal commitments of teachers, and the pressures of the socio-economic climate and its effects on teaching professionals, I remain curious regarding the apparent absence of teacher development in current institutional practice and discourse in the places where I work, or have worked.

Given my circumstances I came to the conclusion that the only viable option would be to take responsibility for my own development. However, what is meant by *teacher development*? The next section attempts to provide a definition.

Teacher development – What is it?

According to Richards, Platt and Platt (1992) teacher development can be defined as a process which

> looks beyond initial training and deals with the on-going professional development of teachers, particularly in IN-SERVICE EDUCATION programmes. This includes a focus on teacher self-evaluation, investigation of different dimensions of teaching by the teacher…, and examination of the teacher's approach to teaching. (pp. 374 – 375, capitals in original)

In contrast to this general definition, Head and Taylor (1997) emphasise the personal nature of teacher development, seeing it as a 'way of learning which is complementary to *training*, and which is motivated by teachers' own questioning of who they are and what they do, rather than by any external training agenda' (p. 1, italics in the original). Underhill (1986) meanwhile is very clear in viewing teacher development as 'the process of becoming 'the best kind of teacher that I personally can be' (p.1). The significance of the preceding quotations is that teachers take responsibility for their own development. As Mann (2005) maintains, 'teacher development is more inclusive of personal and moral dimensions' (p. 104). In contrast, *continuing professional development* is an institution-level preference, while *professional development* is more 'career orientated and has a narrower, more instrumental and utilitarian remit' (p. 104).

Having decided to take responsibility for my own development, I needed to clarify what I was to do and how to do it. The next two sections, a review of the literature and a subsequent description of the methods used to embark upon my own development, will position the current study and clarify how the examination was carried out.

Review of the literature

The review will focus on two main areas: teacher development and language classroom research. In regard to the first, teacher development, views abound and for various reasons the term may be inappropriately used, specifically in connection with the term *training* (Mann, 2005). As a result the distinction between training and development will be drawn at the outset of this review. Following this, the overview will examine the field of *language classroom research*. In order to do this a number of areas will be discussed: *reflective practice* as a means for teachers to contemplate their work; *action research* as a method for teachers to instigate change; and *exploratory practice* as a way for teachers to understand what they do.

The review will demonstrate why exploratory practice is a more appropriate means of teacher development for the author.

Teacher development

As Head and Taylor (1997) point out there would appear to be a difference between teacher training and teacher development. Kennedy (1993) refers to 'the distinction between training and development' describing training as 'a view of teaching as a skill which has finite components which can be learnt'(p. 162) while development 'focuses much more on the individual teacher' own development of a 'theory' through personal reflection, examination and intelligent analysis' (p. 162). Richards et al. (1992) talk of teacher education and define it as 'the field of study which deals with the preparation and professional development of teachers' (p. 374), within which a distinction is sometimes made between teacher training and teacher development (pp. 374 - 375). Allwright (1999a, p. 18) refers to a three-way split: to 'technical competence' or training, to 'information' or education, and to 'understanding' or development. However, such distinctions aside, Ur (1998) advocates caution with regard to the teacher training versus teacher development dichotomy:

> the distinction today has outlived its usefulness. It has certainly been helpful in sharpening our thinking about how teachers learn best…
> But the issue today is not the difference between the two, but rather their integration. We need to evolve a model which combines the best of both in order to design optimally effective professional courses, both initial and continuing. (Ur, 1998, p.21)

Nevertheless, while the two may be integrated, they are not identical. Wallace (1991) concludes that 'training or education is something that can be presented or managed by *others*; whereas development is something that can be done only by and for *oneself*' (p. 3, italics in original). As will be demonstrated subsequently, this is the defining characteristic of teacher development for the purposes of the present study.

Building on the personal nature of teacher development as highlighted by Mann (2005) and Wallace (1991), Underhill (1992) stresses that 'teacher development takes many forms, has different meanings in different contexts, operates from a variety of implicit and explicit beliefs and value bases and is manifested in different forms of action' (p. 80). Edge (1992) uses the terms rather loosely referring to 'autonomous development' (p.62) as well as stressing a more personal use of 'the term development to mean self-development' (p.62) and that 'the purpose of development is action' (p. 64) whereas Allwright (1999b) sees action as only one of a number of choices. Despite this Edge (1992) makes clear that there is 'extra effort involved' in development (p. 70). Development, as Gebhard and Oprandy (1999) emphasise, is indeed personal and can be neither externally imposed nor undertaken with any lack of conviction. Underhill (1992) views development as

being 'no different from personal development, and as such can only be self-initiated, self-directed, and self-evaluated. No one else can do it for us…' (p. 79).

However, it is important to be aware that for many teachers, and for many of their managers, teacher development is still often equated with performance reviews, including classroom observations, as well as career planning/counselling and even disciplinary and dismissal procedures. White, Martin, Stimson and Hodge (1991) refer to 'staff development' in such a vein (pp. 61 – 96). As a consequence, teacher development is all too often done under the guise of quality control, not primarily for the benefit of teachers but for that of the institution, its management and (ostensibly) its students. This is completely at odds with the view of teacher development presented in this paper.

Similarly, teacher development may also, falsely, be seen as little more than receiving 'positive evaluations' (Lamb, 1995, p. 73) from classroom inspectors, mentors or students, and/or keeping abreast of current developments in EFL/ESL, and the collecting of further academic qualifications from educational institutions. Initially this reflected my interpretation of development, or perhaps more accurately, given my ignorance of the notion of teacher development, this was what I thought I ought to be doing: attending seminars and workshops, acquiring certificates, diplomas and academic degrees and doing a not inconsiderable amount of reading of professional literature. The culmination of this approach was enrolment in a PhD programme.

Language classroom research

Stereotypically, pedagogical insights come from academe, and classroom teachers are expected to accept them as sound educational practice. However, the wealth of currently available literature about teachers researching their own lives and activities reveals that this situation no longer pertains universally, and displays a confidence and fund of knowledge on the part of practising classroom teachers (Freeman, 1998, Johnson & Golombek, 2002; Senior, 2006). Widdowson (1984), while emphasising teaching's inherent contradictions – '[n]o matter how concerned teachers may be with the immediate practicalities of the classroom, their techniques are based on some principle or other which is accountable to theory' (p. 87) - admits that there is the need for 'a recognition that what is at the heart of teaching is intellectual enquiry and experimentation, operational research which uses various techniques to test out principles explicitly spelled out' (p. 88). This operational research and its various techniques are now widely accessible via the literature, and it is no longer 'the teacher's role to apply other people's models' (Edge, 2001, p. 5). As Freeman (1998) contends 'many people outside the classroom try to define what teaching should or shouldn't be' (p. 13). This view is the antithesis of the observation that '[t]eachers are not consumers of research, but researchers in their own right' (Widdowson, 1984, p. 90). As Ramani (1987) contends 'teachers often see themselves as 'practitioners' who have little or nothing to do with theory' (p. 3), including the theory of what teacher development could be. Bowen and Marks (1994) underscore this: '...if the *only* channels of enquiry into

professional development which teachers use lead *away* from them, towards outside authorities of various kinds, the result can be an unnecessary and debilitating degree of disempowerment' (p. 5, italics in the original).

Allwright and Bailey (1991) offer the view that classroom research, or '[c]lassroom-centred research is just what it says it is – research *centred* on the classroom' (p.2, italics in the original). For Field (1997) '[c]lassroom research in ELT is not a hobby: it is a professional imperative' (p. 192). Moreover, it is only by undertaking classroom research that teachers will be able to extend their knowledge of the impact of their teaching on the learning processes of their students. In examining our everyday practice we may have findings which are of general application or specific to a particular course, group or individual. As Chaudron (2000) reports, language classroom research has been taken up by teachers who have 'dramatically expanded the scope of their research to address critical areas of practices and problems in language acquisition and use in classrooms' (p. 1). The following sections discuss three possible approaches to language classroom research: reflective practice; action research; and exploratory practice.

Reflective practice

Kullman (1998) regards reflective practice as today's 'most widely promoted model in English Language Teacher Education and Development' (p. 471). For Allwright (1999b, p. 7) reflective practice is 'the real-world exemplar of contemplation for understanding', while Gebhard and Oprandy (1999, p. xi) argue that teaching is a cerebral as well as an active profession. Richards (1998, p. 21) sees reflection, that is critical reflection, as 'an activity or process in which an experience is recalled, considered, and evaluated, usually in relation to a broader purpose'. Bartlett (1990) goes further maintaining that critical reflection implies that teachers 'transcend the technicalities of teaching and think beyond the need to improve [their] instructional techniques. This effectively means [that they] have to move away from the 'how to' questions, which have a limited utilitarian value, to the 'what' and 'why' questions, which regard instructional and managerial techniques not as ends in themselves but as part of broader educational purposes' (p. 205). Richards and Nunan (1990) point out 'that experience alone is insufficient for professional growth, and that experience coupled with reflection is a much more powerful impetus for development' (p. 201). Richards and Lockhart (1996) emphasise that teachers develop routines and strategies for handling the recurring dimensions of teaching. However, research suggests that for many experienced teachers classroom routines and strategies are applied almost automatically and do not involve much conscious thought or reflection. Experience is the starting point for teacher development, but in order for experience to play a productive role, it is necessary to examine such experience systematically (p. 4).

Teacher education, used in a broad, non-technical sense, has two principle components according to Wallace (1991): 'received knowledge' and 'experiential knowledge' (p.17). The former refers to knowledge won via scholarly and scientific work, the latter to practising teachers' ongoing experience. However, this ex-

perience amounts to nothing if it is not supplemented by the 'critical examination of our motivation, thinking and practice' (Bailey, Curtis & Nunan, 2001, p. 39). Such examination (reflection) can be immediate and automatic (p. 41) of the 'reflection-in-action' type (Schön, 1983), or it can be long- term and ongoing in nature, more akin to 'reflection-on-action' type (Schön, 1983) 'informed by public academic theories' (Bailey et al., 2001, p. 42). However, the theories should not be informed by the academic debate alone; classroom teachers have their own contributions to make.

Action research

Action research, as its name would imply, concerns itself with taking action. According to Burns (2005), action research 'takes an explicitly interventionist and subjective approach' (p. 60). Nunan (1989), meanwhile, refers to it as a form of 'teacher-research… distinguished from other forms of research by its practical focus. Teacher-research should grow out of the problems and issues which confront teachers in their daily work' (pp. 15-16). Wallace (1998) stresses 'the systematic collection and analysis of data relating to the improvement of some aspect of professional practice' (p. 1). Field (1997) considers action research as 'a small-scale investigation undertaken by a class teacher' (p. 192). However, Wallace (1991) warns of the dangers inherent in so narrow and strict a definition of an action research cycle, namely that it 'be addressed to practical problems and should have practical outcomes' (p. 56).

Edge (1999) sensibly draws our attention to the fact that '[o]ne of the wonders of action research is that it can help you realise that some of your so-called problems are not problems at all, not in the sense that they have potential solutions' (p. 7). For this author this real-world potentiality is be more akin to reflective or exploratory practice than to action research (see Allwright, 1999b).

Exploratory practice

Exploratory practice seeks to avoid teachers being trapped in contemplation (reflective practice) or leaping headlong from the identification of a problem to its solution (action research). This means that understanding per se can be seen as sufficient: there is no compulsion to act as a result of achieving understanding. For Allwright (1999b) reflective practice, exploratory practice and action research can all bring about change. Whereas only reflective and exploratory practice can lead to an understanding of a situation, action research merely formulates a proposal for change, and, where circumstances allow, implements such change. While both reflective and exploratory practice can result in the understanding of a situation which leads to no call for change, they can also lead to protest should change not be possible. However, it is exploratory practice alone which begins with data generation in order to achieve an understanding of the situation under consideration before deciding whether change, protest or no action is necessary.

Given the above, exploratory practice, rather than reflective practice alone, or a concentration on implementing change as offered by an action research approach, was felt to lend itself better to exploring my own practice and development as a teacher. Gebhard (1996) presents teacher exploration as a four-stage cycle:

1. the collection of teaching samples
2. the analysis of these samples
3. the appraisal of the teaching in the samples
4. the decisions on any changes in teaching behaviour (p. 22)

Using this cycle, development can involve the likes of self-observation, peer observation, writing a teaching diary or talking about teaching with other teachers. Unlike action research, exploratory practice is considerably less intrusive. Allwright (1999b) is very clear on how this is to be achieved: exploratory practice is the 'the deliberate exploitation of standard classroom language learning and teaching activities as the means for collecting data on what happens in the classroom'. In doing so these activities should simultaneously make 'a direct contribution to the learning, and all for the explicit purpose of developing *understanding* of what is happening in the classroom' (pp. 6 - 7, italics in original).

Examination of the literature described above has led to the decision to pursue professional development by exploring my own practice since this combines the roles of teacher and researcher and currently offers a workable way forward.

Research methods and data collection — How to do teacher development

The following section contains a number of possible options for the classroom practitioner interested in, or concerned about, their own development. It does not claim to be definitive and further suggestions can be found in the literature. For a most useful and usable overview, from which some of the suggestions are taken, see Richards and Farrel (2005).

Given that the investigation described here adopted exploratory practice as its approach, and so encompasses the notion that the research be as non-intrusive as possible, the principle means of data collection were the keeping of a teaching diary and recordings of classroom teaching. It is immediately apparent that such data sources alone are insufficient. Although it lies outside the scope of the present paper, as far as research methodology is concerned there is a clear need to triangulate data sources (Cohen, Manion & Morrison, 2000, pp. 112 – 115). To that end I also used peer observation, both being observed by colleagues and in turn observing them. In addition, I made use of conversations with fellow teachers, frequently but not exclusively following observations. A final illuminating source of data was student feedback, more usually written, but on occasion spoken also. These methods are now described briefly. The Appendices contain examples of

data collected. However, all references to individuals, dates and the like have been removed to maintain anonymity.

Diary writing

This study sees the diary as 'a place to record our observations of what goes on in our own and other teachers' classrooms, write about our discussions, consider teaching ideas and reflect on our teaching' (Gebhard, 1996, p. 36). An illuminating example of the diary as tool for teacher development is amply demonstrated by Appel (1995) as he records the highs, the lows, and all stages in between, in the life of a language teacher.

The literature points to diary writing as being prone to two major problems: the question of finding the time and quiet to write as soon as possible after the event; and the need to write in a way which is not just descriptive, but also does not succumb to the dual temptations of overly praising or criticising one's own practice. There are no hard and fast rules for the keeping of diaries. Entries can be complete cohesive texts, lists or bullet-points, sketches, transcripts or even visuals. The prime concern is that writing takes place at regular intervals, ideally while the experience is still fresh, and that entries be reread, reviewed, and analysed for 're-curring patterns or salient features' (Bailey, 1990, p. 215). A critical, non-judg-mental distance to one's own practice can be achieved via such reworking, re-thinking and reinterpreting. The nature of diary writing means that there is not just a need to document evidence, but also to constantly revisit the themes re-vealed via the writing in the light of further teaching in order to ascertain their relevance for development and further examination. Appendix A contains an example of a diary entry made immediately following teaching.

The time-to-write factor has so far been solved by liberal use of much self-discipline, that is forcing myself to write as soon as possible at set times after classes. This has often involved removing myself from the physical environment of my office to the library where the peace and quiet means I can generally write undisturbed. However, when classes are taught back-to-back this can mean a time lapse of a minimum of 90 minutes, if not more. In order to overcome this difficulty and that involved with being unable, for various reasons, to write at scheduled times, I have taken to writing field notes, that is 'brief descriptions in note form of key events that occurred throughout the lesson, including interpretations of incidents where relevant' (Richards & Farrell, 2005, p. 89) as a means of overcoming the natural phenomenon of forgetting. In addition, I have also made use of a number of other techniques: lesson reports which try to 'record what actually happened during the lesson… as a written narrative… or using a checklist or questionnaire (pp. 38-39). These additional tools, are somewhat time-consuming in their preparation, for example, a lesson report, being the opposite of a lesson plan, presumes the existence of the latter. As many practising teachers admit, full-blown lesson plans are not characteristic of their everyday work. Nevertheless, while not reverting to the production of copious lesson planning documents, I have been

forced to pay more and careful attention to the planning of classes, to the input and output, to the contents, objectives and aims.

It soon became apparent that the diary was not only a collection of observations on what happened during teaching, but also a means of commenting and reflecting upon the processes of exploration and development themselves. As Borg (2001) suggests there is potential for validating one's classroom research as one conducts it. He proposes taking the process of diary writing a stage further; examining the diary as a tool of professional development rather than just a data collection instrument. This can add an additional layer to an investigation, a layer of data atop the data initially sought via the teaching diary, and the diary as a professional development tool can thus be used to validate the diary as a research tool. Given the need for qualitative research, that is research of the type described here, to establish its credentials and show its workings (Holliday, 2002), the *research journal* as proposed by Borg (2001) would appear a most appropriate vehicle.

Recording teaching

Audio recording can be used to preserve the actual language used, is naturalistic, constitutes an objective record of the events recorded and can be reanalysed after the event (Nunan, 1992, p. 153). According to Richards and Farrell (2005) '[t]he purpose of making a recording of a lesson is to identify aspects of one's teaching that can only be identified through real-time recordings' (p. 42). The exigencies of everyday teaching reality mean that teachers 'do not often have the opportunity to hear or see themselves teach' (p. 42) and benefit from the, all too frequent sobering and revealing, results.

However, there are the clear limitations to the recording of teaching. Nunan (1992) refers to data overload, the time consuming nature of audio transcription, the need for clear contextual references and a link to another form of data collection, technical problems, the potential hit-and-miss nature of the collection procedure, and, despite what has gone before and a clear wish not to intrude, the clearly intrusive nature of a recording device (p. 25). As Richards and Farrell, (2005) emphasise the presence of a recording device and the knowledge that the lesson is being recorded will influence the dynamics of the lesson, resulting in a lesson that is not typical or representative (p. 43).

In addition, the logistics of making audio recording are complex. The primary decision concerns the focus of any recording; in the case of this discussion, it is the teacher. As a result the recording device is set up to record the teacher only. This is done either using a lapel microphone, or by placing the recorder close to the teacher. The latter has the attendant problem of the teacher needing to stay within close proximity of the device. Tape length is a significant factor, and so a 90 minute lesson requires a tape of the same length and a method devised to facilitate turning over such a tape. The presence of the recording device, if already forgotten about by classroom participants, is made manifest once the end of one side of the tape has been reached. Dealing with this can range from the open and public

turning over of the tape, to a more surreptitious, covert attempt so as not to interrupt the flow of the lesson. Both are extremes and allow for various permutations depending on classroom events. The phenomenon of forgetting to turn over the tape is as simple as it is laughable and seemingly avoidable: it does happen with an alarming and all too sad regularity. The same can be said of batteries running out. Given these very real technical limitations, the use of a digital recording device is now advocated for the future.

In terms of potential difficulties in regard to securing the audio data, it was found to be advisable to review the audio recordings as soon as possible after the lesson. This is done with the aid of a) the diary entry for the class, and b) observation schedules, where appropriate. However, despite the pressing need to review the audio data, it was imperative to complete the diary entry before listening to the tapes. Otherwise the diary data would have been contaminated with data gleaned from the tapes. Given the enormous volume of work involved in audio transcription, it was decided not to transcribe every recorded lesson. Rather than literal word-for-word transcripts in every case another form was devised, the listening report. Appendix B contains an example of the type of brief notes made during the first listening.

Besides the technical aspects already listed students are aware that the lesson is being recorded. Despite raising the issue during the first session and emphasising the personal nature of my undertaking, there remains the potential that student participants may object to the presence of a recording device. While students have usually readily agreed to my use of an audio recording device, thus far, and generally (appear to) have forgotten about the recording, to date one student has mentioned feeling inhibited by its presence. Despite the fact that the technology of the device and the acoustic environment in which the classes take place means that the recordings generally fail to record any student input, this is a genuine concern warranting consideration and discussion. However, since this is the only example of student criticism, this may be discounted as a serious impediment. On the contrary, numerous students have expressed an interest in what I am collecting via audio tapes, what I (intend to) do with the data, and most refreshingly, have commented on how positive they find it that a teacher is interested and concerned enough in their own practice to make such an effort.

Observation

While a number of published (Somogyi-Tóth, 2000) and homemade observation schedules were used, observers were often asked to watch for one or more aspects of practice only, or given free rein to concentrate on what they saw fit. This allowed for observations to be more focussed on the one hand, while freeing both observer and observee from prearranged constraints on the other. This seeming lack of focus meant that the observation could reveal aspect of teaching which neither side had previously contemplated. Appendix C contains such an example of peer feedback, as received after the class. When used, such schedules addressed

various aspects of classroom practice (for example, the use of voice, instruction giving, wait-time or teacher questioning techniques).

However, peer observation has its limitations. As well as being notoriously difficult to organise beyond the one-off occurrence there is also the observer's paradox: just as with the presence of a recording device above, the observer's presence influences classroom events. In addition, the observation procedure needs to be protocolised and standardised. Time needs to be set aside to conduct feedback, often mutual feedback, and these data also need to be recorded. These considerations and practicalities do impose real constraints.

While Fanselow (1987) contends that we learn from watching others, I would posit an extension to this, that we learn a considerable amount from what others say about watching us. Peer observation, is, thus, a most valuable data collection tool. As well as contributing to and aiding triangulation, it can facilitate all too infrequent teacher-teacher communication.

Feedback

It is acknowledged that, as regular classroom participants, students can take on the role of observer (Borg, 2004). However, the parasitic nature of research, as discussed by Allwright (2003), still needs to be addressed. So far this has been done by integrating student observation as far as is possible into the course. This has proved successful on occasion, for example asking students for short feedback on the course a number of weeks into the semester and using this feedback as the basis of the agenda for a meeting simulation in a business English class. Alternatively, ongoing feedback can be taken and used to implement a negotiated syllabus. Appendix D is an example of student feedback collected six weeks into a semester.

Nevertheless, such methods are not suitable for all course types and student groups. The same holds for the feedback itself: some students are able to see that their feedback serves a purpose, and one with direct relevance to them, others are less aware of this, or see no use or relevance of such feedback at all.

The results of using the above are discussed in the next section.

The results — What can exploratory practice bring to teacher development?

Given the preceding sections, teacher development may appear both daunting and expensive. Daunting in as much as the potential task appears colossal when set against practising teachers' (apparently) myriad and more pressing everyday concerns; expensive because time is a most valuable commodity for teachers and there never appears to be enough of it even for the most basic, necessary and core professional tasks, let alone for any of the numerous alternatives and extras clamouring for this precious currency.

The next section lists a number of the benefits to have accrued to the author in the course of research into his own professional teacher development. However, at this juncture it is important to note the following:

1. all of the benefits listed have not occurred , and do not occur, at the same time;
2. the listed benefits likewise have not made, and do not make, themselves manifest immediately – some take considerable time to emerge and some are the result of previously appearing phenomena;
3. some benefits are particular to one activity, and so are more detailed, others are more general in scope and so offer less detail
4. the list is far from being exhaustive.
5. such benefits may not be the result of such exploratory practice alone.

The benefits

There have been a number of benefits of pursuing my own teacher development using the methods outlined above.

For example, one beneficial concomitant of diary writing has been that it does not just explore practice to seek understanding, but can lead to change, and potential improvement, by making us more aware of what we do in the classroom, as opposed to what we think we do. A concrete example is a keenly felt awareness of the necessity of the lesson planning stage, and the usefulness of the lesson plan per se, as illustrated by the comments in Appendix A.

Another unexpected but equally beneficial result of diary writing emerged via rereading and reassessing the diary entries. The diary was not only a collection of observations on what happened during teaching, but the diary became a means of commenting and reflecting upon the processes of exploration and development themselves. This has led to a heightened awareness of what others see us doing, since both colleagues and students can help clarify our understanding of what we do. An awareness of what happens in our classrooms enables us to react to changing classroom circumstances

Mutual observation and discussion of practice can engender a spirit of collegiality and such collegiality facilitates discussion amongst teachers and increases teacher-to-teacher communication, thus reducing the all too common feelings of isolation.

Of equal importance is the inclusion of one's students in the exploratory process as this can serve to break down barriers within the classroom. This means, for example, that existing classroom activities can be used to examine our practice and that the results of such examination can be presented to our students and used in the classroom. Student feedback can thus become a regular feature of classroom practice, no longer restricted to the one-off summative type, but ongoing and formative and so of immediate use. As a result of knowing more about our classrooms, course assessment does not have to be based solely on formal tests – alternative forms can be used to complement standard approaches.

And finally, students appreciate teachers who make such efforts to become more aware of their work.

The drawbacks

Nevertheless, the various teacher development options are not without their attendant problems. The following, listed below, are those that the present author has encountered in the course of exploring his own practice. More are likely to be found as the work continues, both in the field and in the library.

1. exploring one's own practice is extremely time-consuming
2. finding out what goes on in one's classrooms can be a shocking and unnerving experience
3. have colleagues observe one's teaching is potentially face-threatening
4. including students in research into the classroom can be an enterprise fraught with risk
5. such research can likewise be disruptive
6. contrary to claims, here and elsewhere, classroom research practices are still intrusive and potentially parasitic.

Conclusion

The practical rationale behind the suggested activities listed above is that they can be part of a teacher's everyday work and so detract less from their core responsibilities and duties. Add to this Allwright's (2003) caveat concerning the parasitic nature of much academic research, research to which teachers are generally subject without being party, and the attractions become clear, despite the numerous drawbacks. Similarly, teacher research needs to be sustainable (Allwright, 1997) and teachers using their own classroom can readily achieve this. Exploring one's own practice allows the teacher to set the agenda, in their own classrooms, and on their own terms: the benefits outweigh the drawbacks.

The main point to be borne in mind here is not to dwell on the negative aspects revealed by this study. Rather than burden ourselves with such accompanying characteristics of exploratory practice, they should be viewed as challenges to be overcome and important lessons to be learnt in the process of teacher development.

As mentioned at the beginning of this paper, these are not the only avenues open to those wishing to explore their own practice. In addition to the collection of formal teaching qualifications, teachers can also use courses, be they long or short, conferences, workshops and seminars to develop. They can also collect their work in portfolio form as an alternative to standard paper qualifications. Similarly, they can establish teacher support groups, peer teach, mentor other teachers, examine case studies or critical incidents from teaching, or embark on action research. Fi-

nally, they can publish their own work in order to receive constructively critical feedback from their peers.

In sum, although the exploration of practice does not set out to solve problems, the above benefits and drawbacks clearly illustrate the slightly modified adage that for teachers a problem shared can be a problem solved.

References

Allwright, D. (1997). Quality and sustainability in teacher-research. *TESOL Quarterly, 31,* 368-370.

Allwright, D. (1999a). Am I now, have I ever been, and could I ever be - a 'developer'? *NovELTy, 6* (3), 4-19.

Allwright, D. (1999b). *Three major processes of teacher development and the appropriate design criteria for developing and using them.* Paper presented at the Voices from the Field International Conference on Teacher Education, University of Minnesota, Minneapolis..

Allwright, D. (2003). Exploratory practice: Rethinking practitioner research in language teaching. *Language Teaching Research, 7,* 113-141.

Allwright, D., & Bailey, K. M. (1991). *Focus on the language classroom.* Cambridge: Cambridge University Press.

Appel, J. (1995). *Diary of a language teacher.* Oxford: Heinemann.

Bailey, K. M. (1990). The use of diary studies in teacher education programs. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 215–226). Cambridge: Cambridge University Press.

Bailey, K. M., Curtis, A., & Nunan, D. (2001). *Pursuing professional development: The self as source.* Boston: Heinle & Heinle.

Bartlett, L. (1990). Teacher development through reflective teaching. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 202–214). Cambridge: Cambridge University Press.

Borg, M. (2004). Key concepts in ELT: The apprenticeship of observation. *ELT Journal, 55,* 274 – 276.

Borg, S. (2001). The research journal: A tool for promoting and understanding researcher development. *Language Teaching Research,* 5, 156-177.

Bowen, T., & Marks, T. (1994). *Inside teaching.* Oxford: Heinemann.

Burns, A. (2005). Action research: An evolving paradigm. *Language Teaching, 38,* 57– 74.

Chaudron, C. (2000). Contrasting approaches to classroom research: Qualitative and quantitative analysis of language use and learning. *Second Language Studies, 19,* 1-56.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London: RoutledgeFalmer.

Crookes, G. (2003). *A practicum in TESOL: Professional development through teaching practice.* Cambridge: Cambridge University Press.

Edge, J. (1992). Co-operative development. *ELT Journal, 46,* 62–70.

Edge, J. (1999). The freelance action researcher. *BESIG Newsletter, 3,* 6-7.

Edge, J. (2001). Search and re-searcher. *English Teaching Professional, July*, 5-7.

Fanselow, J. F. (1987). *Breaking rules: Generating and exploring alternatives in language teaching.* New York: Longman.

Field, J. (1997). Key concepts in ELT: Classroom research. *ELT Journal, 51,* 192-193.

Freeman, D. (1998). *Doing teacher research: From inquiry to understanding.* Boston, Mass.: Heinle & Heinle.

Gebhard, J. G. (1996). *Teaching English as a foreign or second language.* Michigan: University of Michigan Press.

Gebhard, J. G., & Oprandy, R. (1999). *Language teaching awareness: A guide to exploring beliefs and practices.* Cambridge: Cambridge University Press.

Head, K., & Taylor, P. (1997). *Readings in teacher development.* Oxford: Heinemann.

Holliday, A. (2002). *Doing and writing qualitative research.* London: Sage Publications.

Johnson, K. E., & Golombek, P. R. (Eds.). (2002). *Teachers' narrative inquiry as professional development.* Cambridge: Cambridge University Press.

Kennedy, J. (1993). Meeting the needs of teacher trainees on teaching practice. *ELT Journal, 47,* 157-165.

Kullman, J. (1998). Mentoring and the development of reflective practice: Concepts and context. *System, 26,* 471–484.

Lamb, M. (1995). The consequences of INSET. *ELT Journal, 49,* 72-80.

Mann, S. (2005). The language teacher's development. *Language Teaching, 38,* 103–118.

Nunan, D. (1989). *Understanding language classrooms: A guide for teacher-initiated action.* Hemel Hempstead: Prentice Hall International (UK).

Nunan, D. (1992). *Research methods in language learning.* Cambridge: Cambridge University Press.

Ramani, E. (1987). Theorizing form the classroom. *ELT Journal, 41,* 3–11.

Richards, J. C. (1998). *Beyond training.* Cambridge: Cambridge University Press.

Richards, J. C., & Farrell, T. S. C. (2005). *Professional development for language teachers.* Cambridge: Cambridge University Press.

Richards, J. C., & Lockhart, C. (1996). *Reflective teaching in second language classrooms.* Cambridge: Cambridge University Press.

Richards, J. C., & Nunan, D. (Eds.). (1990). *Second language teacher education.* Cambridge: Cambridge University Press.

Richards, J. C., Platt, J. & Platt, H. (1992). *Longman dictionary of language teaching & applied linguistics.* Harlow: Longman.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action.* Aldershot: Ashgate Publishing Limited.

Senior, R. (2006). *Experience of language teaching.* Cambridge: Cambridge University Press.

Somogyi-Tóth, K. (2000). *Observation tasks: A workbook for student teachers.* Piliscsaba: Pázmány Péter Catholic University.

Underhill, A. (1986). Editorial. *Teacher Development Newsletter, 1*, p. 3.

Underhill, A. (1992). The role of groups in developing teacher self-awareness. *ELT Journal, 46,* 71-80.

Ur, P. (1998, July). Distinctions and dichotomies: Teacher training, teacher development. *English Teaching Professional*, p. 21.

Wallace, M. J. (1991). *Training foreign language teachers: A reflective approach.* Cambridge: Cambridge University Press.

Wallace, M. J. (1998). *Action research for language teachers.* Cambridge: Cambridge University Press.

White, R., Martin, M., Stimson, M., & Hodge, R. (1991) *Management in English language teaching.* Cambridge: Cambridge University Press.

Widdowson, H. G. (1984). The incentive value of theory in teacher education. *ELT Journal, 38*, 86-90.

Appendix A – Diary entry

Aside form the now usual assembly line preparation process, today's lesson was planned with a number of considerations at the front (back?) of my mind:

Avoid repetition of activity types
Use experience won on previous courses
Focus on the participants (their needs, experiences and enthusiasm)

However, despite a knowledge that I'm in control and I know what I want and where I'm going (basically) with the course I still have the nagging feeling that I'm nevertheless in danger of losing this control and/or this direction.
I ought to have read last week's feedback sheet prior to today's lesson (rather than forget them on my desk).
Input my questionnaire – I always am concerned as to the clarity, flow (logic) and relevance of such tasks.
Before this, to bridge the time until all are present, I ask around re the research outline, sources and the like. As expected library-based sources are scarce, while www sources are often too many either to process or ascertain the validity/relevance/accuracy/credence.
This goes better than last week's attempt but gives me the slight feeling that I'm locking the stable door after the horse has bolted.
I didn't do what I wanted to last week re MI (multiple intelligences) and NLP (neuro-linguistic programming) - ought to do it today? I decide against launching straight into authentic materials. Providing the option of pair/small group/group work does not meet massive enthusiasm – do they not wish to make their own decisions re the classroom..? Or am I neglecting my duty in that it is I who should tell them what to do and how to do it. I set a time limit.

Notes taken while teaching:

I planned and planned… and planned…
Too much TTT
No movement in the class
All me/my/mine
Discussion, but…
They need to teach something – a praktikum?
Student involvement was better but still limited

Appendix B – Listening report

administration re HLT articles for next session – lots of time is being wasted (?)

- tedious presentation
- suggestions for how to do presentations in the future (???)
- reflecting in action vs. reflecting on action
- students need to ask themselves the WIIFM question: What's in it for me?
- administer course evaluation-cum-survey

Appendix C – Peer observation

activity	things I especially liked	things I noticed	things I asked myself
Schedule for the day	I think it's useful, and it's good that you are organised enough to be able to keep your schedule		
Names - ball		Students looked a bit embarrassed, maybe because they were not always able to catch the ball or juts they are not used to playing these sorts of games. Some students seemed to have enjoyed it.	
			What shall we do with latecomers???
HW check	•You double-checked all the words. It was really reassuring because sometimes it was difficult to understand what the students were saying. • Note on how to learn words.	"binomial": students did not understand the concept. The blackboard might have been useful?	
Company groups		• *Instruction:* I thought it was clear, but since you had handed out the sheets before giving the instructions some students did	

		not pay attention to you. When you started checking and said: "Look at your notes!" , some students asked "Notes? What notes?"
		• Because you gave two copies to the two bigger groups each, rather than working as a group, students worked in 2 threes separately. They did not share their ideas with the others.
		• People spoke Hungarian in their groups.
	• When nobody volunteered you I liked the way you tricked them with the question about the currency in Japan.	
	• You "forced" your students to talk to each other not you. – an evergreen problem.	
	• I appreciated that you took notes on post-its and reflected on what they said.	
Presentation	•Good vocabulary. Do other students all know these expressions? Maybe he could have written them on the blackboard or could have provided a handout.	The presentation is good for the presenter but how much do the others benefit from it? Would it be an idea to give them some task related to the presentation to focus their attention?
	• "Are you happy with the expression …?" Students look confused. They might not know this expression?	
Is war good for business?	• When students had finished writing on the blackboard. You sat down among them but turned you back to several students. They looked quite confused because, I guess, could not keep eye contact with you.	

News items	• students looked really motivated	• It's always the same people who take the initiative in the groups.	How can we make everybody contribute approximately equally?
		• They seemed to have some problems with the vocabulary of the articles.	
		• As we have already talked about it, you monitor less than I normally do. If they know what to do and have no problems, less monitoring is better. Less embarrassing. But if the have questions or problems, the presence of the teacher might come in handy. They might fear to lose face by asking you to come over and help.	
HW		• I was surprised that they undertook to read the articles at home. I guess this is because they think it's useful.	If you want to work on these next time, I wonder if you will have your copies back. What do you do about this? Do you always come with extra copies? I often have difficulties with that.

Appendix D – Student feedback

1 What did you like and enjoy about the course, and why?

Group and teamwork
C-tests
Talking about citations and references
(the diverse) exercise sheets
That it is thorough enough
Interactive students form what the course is about
Interesting topics
Challenging
"Serves as an additional to the language practice class."
Vocabulary exercises
Free discussion
exam practice
involving activities from different fields
learning from others
atmosphere

2 What did you NOT like and enjoy about the course, and why?

Difficult to understand teacher
More C-tests
Very difficult to fall asleep
Too early
The topic/material of the course
Democratic approach
3 Could the course be improved, and how?

Everything can be improved, but how?
More teamwork
More texts
"Activating some of the sleepers."
More interactivity
More "academic-style practice.
Rescheduling courses for the afternoon

Shaping Teachers' Beliefs Through Narratives

Réka Lugossy

University of Pécs, Hungary

lugossy.reka@gmail.com

Introduction

What prompted me to look at teachers' beliefs about using narratives in the EFL class was something I often hear from in-service teachers: there is no time for stories and there are other, more pressing things to do. We tend to allow the time for a story only when we have finished with the "more serious" things, and thus ignore the potential of stories as tools for learning. Therefore, in a qualitative study involving 32 English teachers and their students, I tried to find answers to why stories fall on the educational margin in Hungarian EFL classrooms, and identify latent teaching and learning philosophies that underlie this phenomenon.

The study reports on the impact of narratives on students and teachers from a double perspective: (1) as teaching materials in the EFL class, and (2) as frames within which teachers reflected on their own practice and attempted to link apparently isolated classroom events in coherent patterns in order to create meaning.

The power of narratives

Framing experience

Anthropologists claim that the story form is a cultural universal: it reflects a fundamental structure of our minds and is therefore one of the earliest and most basic and powerful forms in which we organize knowledge and make sense of the world (Lévi-Strauss, 1966). Cognitive psychologists Schank and Abelson (1995) talk about the power of narratives in framing experience, and claim that "all human knowledge is based on stories constructed around past experiences" as "new experiences are interpreted in terms of old stories" (p. 1). Therefore, stories function as schemata on the basis of which we organize the world and make sense of it. This means that if we expose children to stories we provide them with more opportunities to interpret new information and gradually develop "disembedded", more abstract ways of thinking. In Margaret Donaldson's (1987) view, this is the prime function of education.

Bettelheim (1991) argues for the importance of fairy tales in the emotional development of children because they help children make sense of life and cope with their anxieties. However, it is important to realize that it is not simply the fairy tale genre that contributes to the development of personhood, but narrative thinking in general. Bruner (1986, 1996) talks about the role of "narrative as a mode of

thought" in framing self-identity and cultural identity in the sense that a sufficient story about oneself on the one hand, and awareness of myths and histories of one's own culture on the other, will help individuals in their effort after meaning. It is only in the narrative mode, claims Bruner (1996), that one can construct an identity.

Studies of life narratives (Horsdal, 2005, 2006) show that telling their stories helps people view past events as part of a coherent and meaningful unity, out of which they gain a more profound understanding of themselves and their decisions. In other words, meaning emerges through narrative patterns (Ricoeur, 1981, 1984).

The personal stories lived and told are understood as parts of specific cultures which have their own stories. In the case of teaching, shared cultural assumptions about teaching and learning are part of what Bruner (1996) calls "folk psychology," and are among the basic stories of the local educational culture. However, these shared psychological and pedagogical beliefs and the set of practices they inspire, often go unexamined as part of the "habitus" (Bourdieu, 1990) of school life, while the original underlying theories and social and political imperatives behind them get lost (Millard, 1997). Teachers' awareness of the culturally constructed assumptions about teaching and learning is crucial in understanding practices which now appear to them as "natural" behaviour, and in shaping their identities as teachers.

Language and literacy development

Along their role in the development of personhood, narratives have benefits for language and literacy development. Studies support that exposure to stories is a significant source of incidental vocabulary learning (Elley, 1989), it promotes the development of the language style used in stories (Rosen, 1988), as well as children's background knowledge of a variety of topics (Purcell-Gates, 1988, 1989 cited by Bialystok, 2001, p. 158).

Similarly to L1 development, in foreign language education stories create socially and linguistically meaningful contexts for learning. While this is an important factor for learners of all ages, it is crucial in the case of children, as they make sense of language by making sense of the situation (Donaldson, 1987). Most children are familiar from their mother tongue with the story telling/reading frame and with the kind of discourse stories involve, and when listening to a story told in English they can build on knowledge and skills acquired during story telling sessions in their mother tongue. Also, story sessions create an environment in which learners feel secure and, as the affective filter is low (Krashen, 1985), learning occurs in a natural way. Finally, through stories the target language words and structures are not presented in an isolated way, but embedded in a relevant context.

Building on narratives in the curriculum

Considering the benefits of telling and listening to stories and 'thinking in stories' it would go without saying that a prime scheme of schools should be to create and cultivate narrative sensibility (Bruner, 1996). However, narrative as a mode of thought is valued differently in different cultures. Bruner (1986, 1996) distinguishes between mathematical and narrative thinking, and points out that most western cultures privilege the logical scientific approach in organizing knowledge of the world. Consequently, the convention in most schools is to regard narrative as decoration instead of exploiting its potential as a means for cognitive development. Working in the same stream of thought as Bruner, Egan (1989) suggests that the model that dominates educational programs draws on a part of children's capacities only, namely logical-mathematical thinking, and it ignores the power of imagination as a tool for learning.

The context of research

The background to this study was a methodology seminar for in-service teachers, within the framework of a post-graduate teacher training course at the University of Pécs. The seminar had a double focus. On the one hand, it aimed to provide teachers with techniques and with theories relevant to their practice, and to support participants' reflection on their classroom decisions. On the other hand, it was implicitly meant to encourage story-telling and story-reading in the EFL class on an everyday basis.

As part of the course-requirements, participants were asked to conduct a classroom research project on using stories with a chosen age-group. As during the seminar we had worked extensively with authentic picture books, several teachers decided to work with these, but basically they were free to experiment with any other stories they found appropriate.

Participants were to observe their own teaching and their learners' contribution, and then to analyse and interpret the collected data first in a brief presentation given in front of the whole group, then in a paper handed in at the end of the semester. The last session was dedicated to discussing the findings and reflecting on participants' papers. I considered it crucial that teachers repeatedly talk about and write down their findings, as I hoped that by retelling their stories, they would gain a more profound understanding of classroom events. As shown above, narrative psychology argues for the central role of narrative as a founding principle which gives coherence and integrity to the representation of past events, and it is, as such, a prerequisite of creating meaning (Horsdal, 2005, 2006; László, 2003; Ricoeur, 1981, 1984).

Repeated reflection on their experiences was also meant to make teachers link isolated bits of practice into a coherent methodology, and thus hopefully acknowledge experiential knowledge as a valuable opportunity for professional development, and their classrooms as legitimate learning sites for teacher development (Johnson, 2006, p. 242).

The study

Research questions

I expected to gain insights into how in-service teachers related to using narratives in their English classes. I also assumed that the exploratory project, the writing process and the discussion of the findings would reveal teachers' beliefs about themselves, about their learners and about the teaching-learning process. Therefore, they would be able to regard their practice more consciously, embedded in a wider spectrum of social-cultural and educational traditions. Finally, I hoped that by gaining more articulate knowledge of their own practice, teachers would be able to provide more conscious support for their learners.

I expected to find answers to the following research questions:

- How does using stories influence teachers' and pupils' motivation?
- How do participant teachers relate to the use of narratives, in particular authentic stories, in the EFL class before the project?
- How do these views change as a result of the project?

Participants

There were 32 participants (30 females and 2 males) in the study, all of them in-service teachers coming from various primary, secondary, and language schools in Hungary. The participants, aged 27-45, had received their teaching degrees either from teacher training colleges or from primary teacher training colleges in Hungary. Out of the 32 teachers 15 teach in primary schools, 12 work in secondary schools and 5 teach adults at various language schools.

Data collection methods

In collecting data I relied on

- a questionnaire referring to participants' professional practice and reading habits;
- teachers' written assignments; and
- notes taken during and after the follow-up session and informal discussions with participants.

The questionnaire consisted of five open questions. The first question inquired into teachers' attitudes towards the use of narratives in their EFL lessons, while the second question focused on the use of authentic texts. In both cases I was also interested in the reasons teachers give for using or not using (authentic) stories. The third question inquired into participants' reading habits. With the fourth question I wanted to find out the extent to which teachers rely on their pleasure readings in finding teaching materials. Finally, I was interested in what partici-

pants expected from the course, so that I could adapt the syllabus to what they perceived as their professional needs.

Data were also gained from teachers' research projects, where teachers had the opportunity to carefully articulate their impressions and opinions. Thus, by means of the written assignments I gained insights both into events that took place in teachers' classrooms and into teachers' beliefs about these events.

Finally, the feedback discussion in the last session provided a rich spectrum of data concerning teachers' beliefs about teaching and learning. It also illustrated how teachers, while discussing professional issues, built on one another's ideas and constructed knowledge in their small community (Vygotsky, 1978).

Procedures

The research was carried out in the spring semester of 2003-2004. On the first session a questionnaire was administered to participants, before any input was provided in terms of methodology and applied linguistics. At the same time teachers were informed about course requirements, which included (1) participating in seminar activities and discussions, (2) reading the assigned articles, and (3) doing an exploratory project, where teachers had to observe their own story-teaching.

In the second phase participants experimented with using narratives in their EFL lessons and carried out self-observation. Participants were free to choose the stories they found suitable for their projects. The use of authentic materials was encouraged, but not compulsory.

In the meanwhile the seminar sessions attempted to provide teaching tips (as, according to teachers' answers to initial inquiries about the expectations of the seminar, this was what all participants were looking forward to), and also to involve teachers in discussions in order to highlight the rationale of the activities-in-question.

In the third phase the outlines of the written assignments were discussed. I considered this important partly due to the uncertainty I sensed when the task was assigned. In spite of the written guidelines the teachers received, as well as my initial explanations, the teachers apparently felt the need to come back to the topic of how to structure and write down their observations on every seminar. Discussing the outlines primarily scaffolded teachers' planning process, but it also helped them raise further issues for research in their collected data.

In the fourth phase the written assignments were handed in, and data were analysed along with the data provided by the questionnaires.

The fifth phase consisted of a follow-up discussion of the issues that emerged from the questionnaires and the research projects. This gave participants the opportunity to repeatedly reflect on, articulate and share their understandings of classroom practice with other members of the discourse community. It also helped their learning by relating new information to already existing schemata, which is considered crucial by current learning theories (Vygotsky, 1978).

Finally, in the last phase of the research, data gained from the follow-up discussion were analysed together with the findings from the questionnaires and teachers' projects.

Results

Learners' motivation

All participants claimed that using stories was rewarding in that it increased their learners' as well as their own motivation. As opposed to the initial opinions expressed in the questionnaires (e.g., *"I don't think my students would like reading stories. They're not motivated."*), or skeptical statements made in an informal context right after the first session (e.g., *"Te nem tudod kik járnak hozzánk!"["You have no idea about the students who attend our school."]*), several teachers claimed that students who were otherwise considered "shy," "lazy," and "passive" enjoyed reading and listening to stories and were motivated to participate in follow-up activities.

Stories proved to be a good choice with all the age groups involved. Fifteen teachers tried stories with children, and reported that young learners participated enthusiastically in the shared reading of picture books and in the related activities. Some teachers also reported that the use of authentic picture books in the lesson encouraged some of the children to bring in their favourite Hungarian story-books focusing on similar topics (e.g., animals), while other children asked their parents to buy them books like the ones they had seen in the English lesson. This also sounds like good news concerning the benefits of story-telling and reading for literacy development in the long run. It can be hoped that exposing children to narratives in their foreign language lessons, especially in the early years, creates positive attitudes not only towards the foreign language and towards reading in the foreign language, but towards reading in general. This becomes especially important in contexts where children come from socio-economically and culturally disadvantaged backgrounds, and therefore are at a disadvantage regarding access to print and books.

Much to their surprise, seventeen teachers identified gains in motivation in the case of secondary school students and adults. In spite of initial fears that stories would not attract the attention of these age-groups, teachers experienced the opposite: teenagers and adults enjoyed the stories and the follow-up activities, and asked their teachers to *"bring some more of this next time."* Also, adults learning in a language school in the evening felt that *"this* [listening to stories] *was more or less the only thing they felt like doing at this time of the day"* and *"they didn't even realize time went by and the lesson was over."* Apparently, these age-groups were not used to dealing with stories in the English class.

Learners' language use

Two teachers quote their students, saying that the story-telling frame and the meaning-focused talk that followed the story allowed students to feel more confident about using the language (*"…they told me they at last dared to say the things they were thinking about - in English…"*). This is an important point to remember, mostly because research suggests that confidence and risk-taking are important variables in successful second- and foreign language acquisition (Ellis, 1997).

Finally, several participants note that using stories gave the learners the chance to show their competence in other skills than the ones they were usually expected to display. One of the teachers claimed that the story sessions challenged her views of "good students" both in terms of performance and of behaviour: *"Some students were a complete surprise to me because they were very good at story telling, much better than in any kind of other activity in connection with language learning. Those who are often fed up with the exercises and seem to be bored or would misbehave, sometimes turned out to be the best in this particular area of learning."*

I need to note that despite the enthusiasm expressed by secondary school students and adult learners, there were also worries related to the contribution of stories to their learning. Some learners (as well as some teachers) made it explicit that, although they found stories engaging both cognitively and emotionally, and therefore they preferred them to other activities, they also believed that they could benefit more from direct instruction. This paradox can partly be explained by the socio-cultural background and educational traditions of the teaching and learning community. I will provide a more detailed discussion of the interaction of the cultural context and beliefs about teaching and learning in the discussion part.

Teachers' motivation

Both the written assignments and the follow-up discussion made it clear that teachers found experimenting with stories enjoyable and worth the effort. There is, however, one notable exception in this sense: a teacher who, although considered his lesson successful (*"more successful than I had expected"*), and gradually grew to enjoy it (*"I felt more tense at the beginning, but then I got more relaxed and started to enjoy telling the story."*), concluded that *"story-telling was not my cup of tea"*. I will come back to this teacher's beliefs about teaching and learning, about his learners and about himself in the discussion part. For the rest of the participants, though, the project was motivating because they experienced a feeling of success by attempting to do something most of them had not tried before, and because they evaluated this experience in positive terms.

Teachers' professional development

According to teachers' analyses of their own teaching, another factor that influenced their motivation, besides successful story-teaching, was the realization that by planning and thinking about their teaching, as well as by challenging and revisiting the ideas related to their practice, (e. g., *"I never thought it would work in my classes."*), they developed their expertise.

All teachers, including the above-mentioned exception, found their involvement in the project challenging for their professional development in the sense that they had to (1) consult sources for finding appropriate stories and read articles on classroom research, (2) think about intrinsically motivating tasks to relate to stories, (3) try something new in terms of practice, and (4) articulate their ideas.

It is interesting to note that in their projects many of the participant teachers talk about their initial worries in connection with carrying out an experiment which was in several cases unusual both for themselves and their learners. At the same time teachers express a sense of success in being able to accommodate these worries and taking the risk of the story. As one participant claims: *"Although in the beginning I loved the idea of trying stories in my teaching, I was afraid it might end up in total failure, and I wished I could find a way out of it. But now I am glad I took the risk."*

This sounds like good news in the sense that the development of a risk-taking attitude is crucial for maintaining teachers' motivation in the long run. As teachers encounter the same kinds of situations over the years, they are likely to construct routinized patterns of action to deal with classroom situations, and will suffer from inevitable routinization and professional fatigue (Huberman, 1989 cited by Rudduck, 1992, p. 89). In our case, it seems that the project has brought excitement and induced a risk-taking attitude into teachers' professional practice, and thus enhanced motivation.

I need to add that, as mentioned in the follow-up discussion, it was not only the actual story-teaching experiment that teachers found rewarding (mostly because they turned out to be success-stories), but also the thinking and writing process, that is the process of testing their practice against their beliefs. Studies also suggest that teachers' ongoing professional thinking and development have a great role in sustaining their motivation in the long run (Gebhart, 1999; Gebhard & Duncan, 1992; Spada & Massey, 1992). In this sense the project supports Freeman's (1992) definition of teaching as a process which involves both thinking and doing.

Finally, along with trying new activities and problematising classroom events, reading the relevant literature was also mentioned by participants as an event that contributed to their professional development. The fact that participants explicitly found them useful in planning their teaching and writing shows the need for relevant theoretical input and reflection on it in the process of teacher-development.

Teachers' attitudes to using narratives

The second research question investigated the extent to which teachers use narratives, in particular authentic stories in their classes, and looked at the reasons why they apply or avoid using these reading materials.

The answers to the questions revealed that thirteen teachers out of the 32 who completed the questionnaire sometimes use stories as complementary materials, while nineteen teachers do not. Both yes and no answers list several reasons for their choice.

Teachers who use stories do so because they believe that stories

- increase students' motivation (11)
- provide a good context for new words and facilitate vocabulary acquisition (8)
- have a positive effect on classroom atmosphere (7)
- are helpful in providing visual support (1).

Reasons for not using stories include

1. lack of time (13)
2. learners' low language competence (9)
3. language / school-leaving exams (5)
4. incompatibility of available materials with age-group-in-question (3)
5. students' lack of motivation (1)
6. students' cognitive abilities (1)
7. students' lack of literacy skills in L1 and in the foreign language (1).

Those teachers who use stories identify their role in creating motivation and a stress-free classroom atmosphere. They also mention that narratives facilitate language acquisition by providing context-embedded, situationally-linked language, occasionally supported by pictures. What is surprising, though, is the relatively small number of teachers who identify these benefits. Taking into account that in-service teachers had already received education in terms of psychology and methodology before this postgraduate course, there are few of them who rely on the theoretical knowledge previously acquired in their first degree programmes. What is even more disheartening, however, is that they do not seem to rely on their common sense and experience gained as educators and even parents.

As for the reasons for which teachers claim to neglect stories, it is interesting to note that they are mostly attributed to external factors (e. g., time, language exams as well as to students' perceived levels of language competence and motivation). None of these reasons refer the respondent teachers' personalities or professional preferences or the expertise they believe to have.

Teachers using authentic stories

It also turned out from the questionnaires that out of 32 teachers 20 claim not to use authentic stories, because they find them too difficult for the learners' level, and because they consider it hard to find materials which are appropriate to their level. However, four teachers sometimes use extracts from *Newsweek*, five use nursery rhymes, one teacher uses *Winnie the Witch* (Korky & Thomas, 1997) along with the activity book, and two of them mention songs. Twenty teachers claim never to use authentic materials in their English class.

Two out of the five teachers who use nursery rhymes and the two teachers who use songs, mention that rhymes and songs may have clear story-lines and can therefore be considered narratives in their own right. One of the arguments for using traditional nursery rhymes is that they are *"the real thing."* Teachers claim that *"it's natural to use them . … this is what we would use in Hungarian with children."*

All the reasons for which teachers do not use authentic reading materials relate to the level of difficulty of these materials compared to students' level of language skills. However, studies focusing on using authentic stories in the young learner classroom (Jones Mourao, 2006; Williams, 1995) show that learners from an early age find the adequate strategies to cope with authentic materials which are not meant for pedagogical purposes, and therefore do not offer language specifically tailored to the learners' level.

Change in teachers' views about using narratives: teacher learning

Teachers' written assignments and the follow-up discussion suggest that teachers have to a certain degree re-evaluated their beliefs about the role of narratives in their language classroom. It seems that most teachers no longer reject them ab ovo, (e.g., "*I teach in secondary school,"*), but acknowledge stories as motivating learning experiences. I previously mentioned that participant teachers identified benefits of stories in terms of learners' and their own motivation and development, as well as classroom atmosphere.

However, a deeper analysis suggests that the change in teachers' attitudes to stories is rather on the level of expressed beliefs than a fundamental change of beliefs which consistently shapes teachers' behaviour. For, although teachers explicitly talk about the benefits of integrating narratives in their teaching, the belief that stories may not bring about the desired learning outcome is also repeatedly articulated. This apparent contradiction can be explained by Schön's (1983) distinction between teachers' "espoused theories," that is the externally imposed knowledge and their "theories-in-action:" the tacit knowledge which determines the ways teachers spontaneously act in the classroom, and which is rooted in their personal and professional biographies, as well as the social, political, economic and cultural histories of contexts where teachers learn and teach (Johnson, 2006).

In writing and talking about their research the teachers documented what turned out to be their covert understandings of their teaching situations. The ac-

tual writing of their research papers and the follow-up discussion were meant to give teachers the opportunity to think about their practice, articulate their beliefs and confront them with the received knowledge, that is, their "espoused theories" by means of telling their own stories. In this sense, Elbaz (1991 cited by Johnson, 2006, p. 236) refers to the use of narrative as an important means of understanding and documenting teachers' knowledge. Johnson (2006) also points out that narratives "connect phenomena, infuse them with interpretation, and thus uncover teachers' interpretations of the activities they engage in" (Johnson, 2006, p. 242). Therefore they reveal teachers' individual meaning-making process and the understandings of their particular context.

Making tacit knowledge explicit by reflection is a necessary step in order for fundamental change to occur in teacher behaviour. The research papers document that some change in teachers' practice has already begun. This is supported by teachers' enthusiasm, their recognition of the benefits of using stories and of the project, and also their claim that they would *"try stories again,"* or *"try the same story with other groups as well."* Promising as they sound, these statements also imply that teachers do not yet think of using stories on a principled and regular basis, but much rather in terms of occasional experiences. This also suggests that the process of change has only just begun: although newly acquired theories are there on teachers' palette, they still greatly rely on what Schön (1983) identifies as "theories-in-action."

The fact that teachers' implied theories about teaching and learning coexist with newly acquired ones is a normal stage in the process of development. Theories that define human learning as a dynamic social activity situated in physical and social contexts (Rogoff, 2003; Vygotsky, 1978) argue that participation and context are crucial in human cognition. It is therefore natural that teachers build on their implicit understanding of their lived practice when making sense of new information. Also, ecosocial theory and research carried out in this tradition suggest that short timescales do not allow for fundamental changes in attitude and habits of reasoning (Lemke, 2002). Although short-term events (such as teachers' exploratory projects) contribute to such changes, it is only if they are "reinforced by subsequent events which make for the kind of persistent change we really mean by 'learning'" (Lemke, 2002, p. 75).

Culturally constructed beliefs about teaching and learning

As shown above, all participants found using stories rewarding in that it increased their learners' as well as their own motivation; it improved teacher-student rapport, and learners' language use. Most teachers mention, though, that they were "*surprised*" to find that stories were well received in their classrooms, and give various reasons why using stories in the EFL class is problematic: (1) Teachers find there is no time for stories, and fear that by integrating stories in their teaching they might take away time from other, more effective and therefore more important tasks. (2) Stories are considered by some participants as mere entertainment with no real learning potential, and which may therefore harm teachers' prestige.

(3) Stories are uninteresting for secondary-school students or older learners. (4) Pupils' level of language and literacy skills does not allow for using stories, in particular authentic texts.

Seen in their social-educational context, these ideas convey messages related to teachers' beliefs *about* teaching and learning, as well as assumptions *about* their learners and *about* themselves and their role as teachers. These sets of beliefs are interrelated, and go deeper than what happens in the classrooms. Indeed, they are based on shared and often tacit cultural understandings of what teaching and learning are. Therefore, when we think about why teachers use or do not use stories, we must approach the question from a wider social context than that of the classroom only, one that involves values transmitted by the whole community. In the following I will analyse data along the main issues explicitly identified by teachers and try to uncover the underlying assumptions, which may occasionally go deeper than teachers' individual histories.

Wasting time with stories

In spite of the overall positive feedback, on the basis of the classroom research project it appears that some teachers consider that stories involve merrymaking much rather than learning, and therefore are rather used when they have finished with the more "serious" tasks.

In the data gained from the questionnaires thirteen teachers explicitly claim that they have no time for stories. The worry that by doing stories in the English class one might take away time from other, more important tasks appears in three classroom projects where teachers try stories with students a few weeks before school-leaving exam or a language exam. One of the participant teachers claims: *"[Learners] have to take the language exam. In this case direct instruction is more effective."* From the papers it also turns out that it is not only the teachers, but the learners themselves who are worried about losing time with apparently irrelevant activities. One of the teachers notes: *"They seemed a bit concerned because they thought that we were wasting a lot of time with this exercise, but in the end they said: Hát ez tök jó volt! [It was cool.]"* In another instance an 18-year-old is quoted to have said to her fairly young teacher: *"Erika, we really like your stories, but we think this year we ought to be doing the topics for the final exam."*

Teachers' background

Both teachers' and learners' fear lest stories may take away time from what seems to be more focused activities is partly due to the washback effect of exams. At the time of the study, the only accredited language exam and the final examination system in Hungary were strongly focused on grammar and form, with minimal or no emphasis on communication. Johnson (2006) makes the telling point that most second-language teachers work in institutions in which "they, their students, and their instructional practices are constructed by the positivistic paradigm that de-

fines good teaching in terms of student performance on standardized tests..." (Johnson, 2006, p. 247). The fact that teachers are held accountable for student learning which is measured by achievements on standardized assessment instruments, makes it understandable that they would much rather teach content involved in these tests than something that is not directly tested, even if the latter may offer more satisfaction and gains for both teachers and students. Johnson (2006) also notes that most language teachers are the product of the above-mentioned positivistic paradigm, as they have been socialized into normative ways of thinking about language teaching and learning.

Therefore, the view that stories are largely pointless activities, as opposed to more serious language development work, reflects underlying assumptions about teaching and learning that teachers and learners share, due to the educational traditions long existing in Hungary. Most teachers participating in the study belong to the generation taught with the Grammar Translation and with the Audio-lingual Methods. Therefore, even when they use communicative teaching materials, they enact these materials in non-communicative ways. Research conducted by Nikolov (1999, 2002) shows that teachers rely on typical traditional classroom activities, and adopt these traditional techniques in spite of the fact that they may go against current language acquisition theories that the teachers study and are expected to know about and apply in practice. That is to say that even though teachers are familiar with the theories, they fail to link knowledge gained from theoretical training to their everyday practice. In other words, they fail to develop what Wood (1988) calls "expertise."

Similarly, when it came to developing follow-up activities to the stories they used, some teachers preferred activities with a grammatical focus to meaning focused tasks (for example, certain grammatical structures from the story were reinforced with follow-up drills, so that learners would surely remember them). Paradoxically, at the same time they expressed their confidence that by stories learners' communicative competence would develop. As it turns out from the research papers, after reading or telling the stories, most teachers considered that the best way to check comprehension was by (1) true or false questions, (2) comprehension questions, or (3) rendering the summary of the text in Hungarian. While it is true that these techniques can be useful for this purpose, they are not what some teachers believe them to be: meaning-focused tasks for developing communication in English.

The discrepancy between what teachers know in terms of theories and their concrete choices in terms of methods and techniques suggests that the teachers participating in the study believe in a bottom-up, more form- than meaning-focused, approach to teaching and learning a language. Also, a product- rather than process-oriented view is detectable with several teachers. I will quote only one suggestive comment in this sense, referring to the fact that after the first reading of the story, learners did not understand all the words. Their teacher then concludes: *"Maybe we should have practised the words before the story, so that they could recognize them while they were reading."*

When I anonymously came back to the above statement during the follow-up discussion, teachers appeared to predominantly agree with the importance of re-

lying on the context as a basic reading comprehension strategy. They claimed that a text that matches learners' cognitive and linguistic level and is relevant to their interests involves them in guessing the meaning of the words they may not know. Interestingly, their research papers often document quite the opposite: while teaching their lessons, eleven teachers admittedly resorted to pre-teaching the new words before they actually dealt with the story.

Learners' expectations

Interestingly, messages about "right" teaching also come from learners, who are instilled with the beliefs about how learning occurs transmitted by the socio-cultural community (e.g., parents, peers, and their own experiences). Therefore, when we talk about what underlies teachers' classroom decisions in terms of methods, techniques and materials, we need to realize that it is not only teachers and their professional background that influence these decisions, but also learners' and indirectly, parents' ideas and expectations about what should be going on in a classroom for the sake of what they believe to be the best outcome.

Out of 32, nine classroom research projects touch upon the point that while most learners visibly enjoyed the story-telling and story-reading experience and the follow-up activities and discussions, they expressed the fear that no real learning takes place. One of the teachers quotes the example of an upper primary school child, who asked for more exercises from *Project English* as homework, saying that: *"We haven't worked that much today anyway." [Ma úgysem dolgoztunk annyit].* The teacher also confessed that while at the beginning of the project she thought she was being innovative, and therefore would be more appreciated by the children and their parents, after this remark she was worried lest parents would object to the idea of her using stories. In an informal comment she remarked, *"I wonder what they think I might be doing in the lesson…"[Kíváncsi vagyok, mit gondolnak, mit csinálok én órán…].*

That teachers often tend to evaluate and adapt their teaching according to what they believe to be parents' expectations is also highlighted by the following teacher-comment (made during an extensive reading project currently run by secondary school teachers among their students): *"I wonder when they* (i.e. the students) *will get one of the parents call and complain that the only thing I do is keep reading in the English class." [Már várom, mikor telefonáltatnak be egy szülőt, hogy én csak olvasgatok angolórán].* The fear that students and their parents might disapprove of the innovative practices used by this teacher is an understandable cause of worry, as school administration is also likely to react in unfavourable ways to whatever may go against "regular" ways of teaching, mostly if this might displease parents.

Educating learners

As "wasting time" with stories proved to be a major concern, in particular for secondary school learners, it seems that one of the tasks that teachers face in this context is raising learners' awareness about the impact of certain activities in the lesson. This is important partly because, being aware of the aim of certain tasks in terms of language development, may bring about learners' conscious and active participation in the learning process, and therefore enhance their motivation during the lesson. On the other hand, making learners familiar with the long-term implications of these classroom processes is the beginning of creating the foundations for learner autonomy. Classroom-based studies show that involving learners in classroom decision-making and self-assessment empower learners and create responsibility in the long run (Nikolov, 2000; Serrano-Sampedro, 2000). In the case of our project, regularly involving students into story-telling and story-reading activities may have far-reaching benefits for students' personal development, their foreign language learning skills and motivation to read in both a foreign language and in the mother tongue.

This is entirely in keeping with a constructivist approach to education (van Glasersfeld, 1995 cited by Williams & Burden, 1997, p. 49). In this framework of ideas, education is essentially seen as a political enterprise the purpose of which is to empower learners to think for themselves. This can be best accomplished by giving learners "the reasons why particular ways of acting and thinking are considered desirable" (van Glasersfeld, 1995 cited by Williams & Burden, 1997, p. 49). Therefore, making learners aware of the long-term choices they make should be part of the process of education that teachers undertake. This is obviously difficult when teachers themselves are not able to clearly articulate why they are doing what they are doing in the classroom and are not convinced of their own "exploratory study."

Johnson (2006) points out that one of the largely unexplored areas in current research on teacher cognition is the relationship between teacher learning and student learning. She also notes that, although teaching does not necessarily result in student learning, therefore the relationship is not a causal one, there is, however, "a relationship of influence" between what teachers learn, how they organize their classroom activities, and what students learn from engaging in those activities (Freeman & Johnson, 2005 cited by Johnson, 2006, p. 249). This view of the teaching-learning equation suggests that, even if indirectly, learners are influenced by the changes that occur in teachers' thinking and actions.

Stories are meant for young learners

Besides the idea that stories take away important teaching time, another widespread belief among the participants of the study was that from a certain age upwards learners are no longer interested in stories. The data gained from the questionnaires suggest that teachers associate story-telling and story-reading in the classroom with young learners, and believe that they are no longer of interest to

teenagers or adults. Some of the teachers who (almost) never use stories in their classrooms, apart from those in the course book, argue their choice by claiming: *"Secondary-school pupils are not that keen on stories."* Others only imply the lack of interest in narrative by simply indicating the age-group they teach: *"I teach in a secondary-school."*

Inappropriate texts

One of the questions that would naturally arise is why some of the teachers believe that narratives are suitable exclusively or in particular for young learners. Hardy (1978) talks about the importance of children's fiction in that it responds to the *universal human need for narrative*. Besides, the narrative pattern is believed to be a most basic form in which humans across cultures and age-groups (Lévi-Strauss, 1966) organize experience and relate it to their lives (Schank & Abelson, 1995).

Also, if we look at teachers' answers to the question whether they consider themselves good readers, it turns out that they value reading (most of them wish they had more time to read). This suggests that it is not *reading* that they find uninteresting for teenagers and young adults, but, as it appears from the case of the teacher presented below, rather the kind of texts associated with classroom use. As it repeatedly turned out in seminars, teachers were scarcely familiar with books that are both appropriate for students' linguistic level, and at the same time are relevant to their lives and make human sense to them. Therefore, when it came to using stories in their classes, with few exceptions teachers opted for simplified versions of classical fairy tales, or borrowed books we used during the seminars, such as *The three little wolves and the big bad pig* (Trivizas & Oxenbury, 1995) and *Revolting rhymes* (Dahl, 1984).

Teachers' prestige

There is one assignment in particular which is worth dealing with in more detail, as it clearly articulates several of the underlying assumptions that teachers appeared to have in connection with stories. The author of this assignment is a teacher of History and English in a group where he is also form-master. However, he chose to teach a story to a group of 15-year-olds whom he did not know. In the introduction to his research paper he argues that his status as a form-master and History teacher in his regular group was incompatible with "story-telling activities" in the English class. He claims, *"…I must admit that I didn't know how I would motivate my students to take a story-telling lesson seriously…"*

It appears that one of his worries is that he might lose face with using stories, which he implicitly believes to be a light and frivolous occupation with no real learning outcome, and which would therefore not contribute to learners' language development and thus, do harm to the teacher's perception as a professional. The image he has about himself as a teacher is defined by the perceived social value of

a subject (History being regarded as a prestigious subject in Hungarian schools), and by his status of form-master. His face-saving manoeuvre also indicates his beliefs about himself and his role as a teacher, which is incompatible with using unaccepted teaching materials: stories. Part of the paradox is that the hero of this story is also a teacher of History, which, along with literature, best cultivates "narrative thinking" (Bruner, 1996) out of all school subjects.

Although the teacher himself was pleased and surprised by the fact that pupils enjoyed the lesson, and even confessed having "enjoyed the task" himself, he summarizes his experiences by saying: *"Leaving the class I had conflicting feelings. On the one hand, I had to admit that the lesson was more successful than I had expected. I was surprised that the pupils were so enthusiastic and that they were keen on listening to a fairy tale."*

After the successful outcome and the apparently positive teacher reaction, he concluded his paper by making a rather disillusioning statement: *'On the other hand, it was clear for me that … story telling was not my cup of tea. … Moreover, I am sure that with secondary school students this sort of game can be played once or twice but in the long run they would find it odd and boring."*

Quite obviously, this teacher remained unconvinced of the benefits of exposing learners to narratives on a regular basis, in particular of the potential of stories as a tool for learning. From what I can conclude from his research paper, his views about learning as a step-by-step process based on structured input provided by course books remained unchallenged and unchanged. It seems that he implemented his research project and went back to his old practice without considering the value of his experience.

Pupils' level of language skills

In the answers provided in the questionnaires ten teachers argue that they do not use narratives because of their students' low language and literacy skills. Some of the emblematic answers in this sense are the following: *"I teach teenagers from a disadvantaged background. They don't even read well in Hungarian;" "I would like to use stories, but they don't know enough language yet;" "It's hard to find appropriate stories for teenagers who are still beginners;"* and *"I teach children. Stories are too difficult for their language level."*

When it came to the question of using authentic stories in the language classroom, the participant teachers tended to resist the idea on grounds that authentic stories, as opposed to texts designed for language learning purposes, are too demanding for their students. When asked whether they use authentic texts (picture books, short stories, poems, comics etc) four teachers out of 32 claimed that they sometimes use extracts from *Newsweek*, five teachers use nursery rhymes, one uses *Winnie the Witch* (Korky & Thomas, 1997) along with the activity book, and 2 of them mention songs. Two teachers use songs, while 20 teachers claimed that they never use authentic materials.

Two observations are in order here. While most teachers agree that their learners are below the linguistic level that they suppose to allow them to cope with authentic stories, four teachers claim to use extracts from *Newsweek* as "authentic

stories" in their lessons. Apart from the fact that texts from *Newsweek* may not be among the prototypical examples of narratives, it is also interesting that these teachers do not mention using other, perhaps more accessible authentic readings. So it is rather hard to imagine that these articles would make appropriate readings for students otherwise considered unable to cope with authentic stories. It may well be that by alluding to authentic materials which are considered difficult and therefore of high prestige, the teachers wanted to convey what they thought would be a positive image about themselves as informed and demanding teachers. This would probably be in tune with their underlying assumptions about good teachers.

The other observation relates to the use of a picture book, *Winnie the Witch* (Korky & Thomas, 1997), in an edition tailored for pedagogical purposes. While there is no doubt that, due to the quality of the pictures and the text authentic picture books assist children in becoming readers and also in acquiring first or foreign language (Arizpe, 2006, Baddeley & Eddershaw, 1994, 1998; Graham, 1990, 1998, Marriott, 1998, Parkes, 1998, Stephens, 1992), there is an inherent danger in using them as if they were meant for language *instruction*, and thus overexploiting them. The poet Michael Rosen (2005) is of the opinion that in schools books for children are being turned into worksheets where children are supposed to spot grammatical forms. This happens in order to support more structured literacy development, and also to make assessment of reading skills easier. Rosen (2005) claims that this is the very process which undermines literature and the enjoyment which authentic children's books are meant to generate with readers. By quoting this view, I do not mean to imply that using *Winnie-the-Witch* and the related workbook are bad options in the young learner classroom. On the contrary, out of all the authentic materials, picture books are probably the most suitable option to make in the young learner EFL classroom. But it needs to be added that using pictures books which are not published with pedagogical aims, and exposing learners to stories without the "necessary" follow-up language-focused tasks can also be very useful in terms of language development, and at the same time offer a satisfying aesthetic experience.

Teachers' beliefs about authenticity

It appears that the arguments by which teachers explain why they avoid using authentic materials, are similar to the reasons why teachers avoid using stories in general, namely that they believe them to be too difficult for the level of the learners. This shared assumption most teachers have regarding narratives in general, and authentic texts in particular, may have to do with their schemata of children's literature acquired through education and consisting largely of great classics ranging from the *Household tales* of the Brothers Grimm to *Alice in Wonderland*. This culturally acquired concept of children's literature is reinforced by the fact that the only choice the Hungarian book market offers in terms of children's literature in English contains examples of canonized literature. Artistic, multi-layered picture books, touching on issues which are emotionally involving and cognitively en-

gaging for older learners as well, such as Anthony Browne's *Zoo* (Browne, 1992) and *Piggybook* (Browne, 1986) are mostly unknown to the Hungarian audience. Thus, the concept of authenticity evokes the notion of linguistically demanding texts associated with native speakers' proficiency.

The 32 teachers involved in the study seemed to agree when one of the primary-school teachers said: "I had never tried real English books before. I thought they would be too difficult for the children because they might not know the words…. I always liked to keep it simple for them: to match their level." This statement is particularly interesting in terms of what it reveals about teachers' underlying beliefs about how children learn a foreign language and the value of authentic materials in this process.

One of the assumptions implied by the quoted passage is that in order to understand a story one has to understand all the words, or else one might not be able to make sense of the text (*"They might not know the words…"*). However, both experiences we have from our own efforts of meaning making and cognitive models about how we organize and interpret knowledge about the world (Schank & Abelson, 1995) seem to contradict this fear: familiarity with the story-telling frame, the prototypical story structure and certain features of the text, as well as the pictures in the case of picture books may compensate for at least some of the words that the learners might not know.

In the comprehensible input hypothesis, Krashen (1985) argues for the importance of roughly tuned input, containing linguistic information which is one level beyond the learner's actual linguistic level. Krashen claims that by exposing learners to comprehensible input, language development occurs. In the same frame of ideas, Smith notes that we do not learn to read first and then read, but we learn to read by reading (Smith, 1985).

Another belief that is made explicit in the statement quoted above is that the "real" story or the "original version" is always more difficult to understand than a text specifically designed to match learners' level, and is therefore "kept simple" (*"Keep it simple … to match their level"*)**.** Cameron (1994) argues that the natural redundancy of authentic texts (as opposed to simplified versions) may actually assist learners in the meaning making process. Removing certain text features (e.g., changing simple past, which is basic to narratives, into simple present) affects not only the style of the narrative, but may also deprive readers of valuable clues in understanding the story. I am not assuming that authentic texts are all easy to comprehend or comprehension is simply guaranteed by authenticity, but we have reasons to assume that careful selection of authentic stories and sensible scaffolding on the part of the teacher can assist learners in the meaning-making process.

Finally, there is one more point I wish to make in connection with participants' use of authentic materials. From the questionnaires it turns out that out of the ten teachers who sometimes resort to authentic materials seven are primary school teachers, and only three teach on the secondary level. One would expect that, due to the importance attributed to the perceived language level of the learners, it is rather secondary school teachers who would think their learners to be more able to cope with authentic discourse, and therefore would be more willing to experiment. The answers also reveal the fact that primary school teachers tend to use

rhymes and songs, which offer more here-and-now input, and make shorter texts and readings than stories would for secondary school pupils. This leaves us again with the often inarticulate lack of knowledge concerning accessible authentic materials for learners from upper primary upwards.

Teachers' beliefs about themselves

One of the issues mentioned in four of the 32 assignments was related to teachers' beliefs about themselves as non-native speakers. Three teachers considered that stories are easier to use for native teachers than for non-natives, while one teacher noted that working with a story was "a difficult job" for her, as she felt threatened by unexpected situations and was afraid that the adults she taught would "immediately notice every mistake" she made.

Teachers' fear of making mistakes and being criticised by their students reflects an emphatic concern with grammar and accuracy. By this I do not mean to under-evaluate the importance of language competence. Quite the contrary. It seems that having a good command of English is important not only because it guarantees quality input for learners, but also because it has implications for teachers' self-concept as teachers, in particular their confidence. Therefore, it is by all means advisable for teachers to pursue "perfection" in this sense.

However, there is an inherent danger in too much concern for perfection, as soon as it hinders experimenting and risk-taking in teaching and in teacher learning. In a classroom environment which does not allow for mistakes, students' performance is also likely to be perceived in a product- rather than a process-oriented form. In such instances learners' results are also seen as being largely a matter of teachers' responsibility.

Conclusion

The findings of the study suggest that both teachers and their students found story-telling and reading a motivating experience in their EFL lessons. Despite teachers' initial worries that older learners may find dealing with stories odd and boring, narratives turned out to be intrinsically motivating for all age groups, including young learners, secondary school students and adults. Even students who were otherwise found passive and reluctant to participate appeared to enjoy the story sessions and performed well in new activity types.

However, to most teachers this came as a surprise: despite the theoretical and practical input provided in terms of the contribution of stories to language and literacy development, and their own enthusiasm while participating in story-based activities during seminars, teachers did not expect their students to enjoy and cope with stories.

Along the positive feedback, their written assignments also reflect sceptical attitudes as to the role of narratives in the EFL lessons, as compared to the value of direct instruction. As I pointed out earlier, this attitude is largely due to the edu-

cational culture, in particular to teachers' ways of knowing about teaching and learning and the stereotypes about teachers and learners into which they are socialized.

Still, teachers found their involvement in the project challenging for their professional development in the sense that by trying something they had rarely or never done before, they developed a more risk-taking attitude. Also, problematising apparently minor classroom events as issues for research involved teachers into thinking about their practice more objectively. This is an important result in that, as Freeman (1992) mentions, the idea of change does not imply that teachers must necessarily discard their current practices and do something differently. The focus in the concept of change is rather on raising teachers' awareness of their existing constructs by encouraging them to reflect on their practice in a critical way. It seems to me that this process has started.

The study leaves us with a number of messages for teacher education and further research. It first of all suggests the need to provide would-be teachers with relevant theoretical perspectives, and to help them link these theories with their existing practice or classroom experience. In the case of already practising teachers, it seems crucial to raise awareness of the value of their classrooms, where they spend a great amount of their lives, as sites for teacher learning. Finally, teacher education should take into consideration the local context, in particular educational traditions and shared cultural assumptions about education, as a complete story about oneself is crucial in meaning making.

References

Arizpe, E. (2006). Young interpreters: Children's response to pictures. In J. Enever, & G. Schmid-Schönbein (Eds.), *Picture books and young learners of English* (pp. 35-48). München: Langenscheidt.

Baddeley, P., & Eddershaw, C. (1994). *Not-so-simple picture books: Developing responses to literature with 4-12 year olds.* Stoke-on-Trent: Trentham Books.

Baddeley, P. & Eddershaw, C. (1998). Linking books to develop older children's response to literature. In J. Evans (Ed.), *What's in the picture? Responding to illustrations in picture books* (pp. 58-78). London: Paul Chapman Publishing Ltd.

Bettelheim, B. (1991). *The uses of enchantment: The meaning and importance of fairy tales.* Fourth edition. London: Penguin.

Bialystok, E. (2001). *Bilingualism in development: Language, literacy and cognition.* Cambridge: Cambridge University Press

Bourdieu, P. (1990). *In other words.* Cambridge: Polity Press in association with Blackwell Publishing.

Browne, A. (1986). *Piggybook.* London: Julia MacRae Books.

Browne, A. (1992). *Zoo.* London: Julia MacRae Books.

Bruner, J. (1966). *Towards a theory of instruction.* Cambridge, Mass.: Harvard University Press.

Bruner, J. (1986). *Actual minds, possible worlds.* Cambridge: Cambridge University Press.

Bruner, J. (1996). *The culture of education.* Harvard University Press: Cambridge, Massachusetts.

Cameron, L. (1994). Organizing the world: Children's concepts and categories, and implications for the teaching of English. *ELT Journal, 48* (1), 28-39.

Dahl, R. (1984). *Revolting rhymes.* Harmondsworth: Puffin Books.

Donaldson, M. (1987). *Children's minds.* London: Fontana Press.

Egan, K. (1989). *Teaching as story telling: An alternative approach to teaching and curriculum in the elementary school.* Chicago: University of Chicago Press.

Elley, W. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24* (2). 174-189.

Ellis, R. (1997). *Second language acquisition.* Oxford: Oxford University Press.

Freeman, D. (1992). Language teacher education, emerging discourse, and change in classroom practice. In J. Flowerdew, M, Brock, & S. Hsia (Eds.), *Perspectives on second language teacher education* (pp. 1-21). Hong Kong: City Polytechnic of Hong Kong.

Gebhard, J.G., & Duncan, B. (1992). EFL teacher education curriculum development as inquiry: The orosz szakos tanárok átképzési programja. In J. Flowerdew, M. Brock, & S. Hsia (Eds.), *Second language teacher education* (pp. 319-335). Hong Kong: City Polytechnic of Hong Kong.

Gebhard, J.G., & Oprandy, R. (1999) *Language teaching awareness: A guide to exploring beliefs and practices.* Cambridge: Cambridge University Press.

Graham, J. (1990). *Pictures on the page.* Sheffield: The National Association for the Teaching of English.

Graham, J. (1998). Turning the visual into the verbal: children reading wordless books. In J. Evans (Ed.), *What's in the picture? Responding to illustrations in picture books* (pp. 25-43). London: Paul Chapman Publishing Ltd.

Hardy, B. (1978). Narrative as the primary act of mind. In M. Meek, A. Warlow, & G. Barton (Eds.), *The cool web: The pattern of children's reading* (pp. 12-33). New York: Atheneum.

Horsdal, M. (2005, December). Életút-történetek elemzése és értelmezése [Analysing and understanding life-story narratives]. Retrieved from: http: civil.nepfoiskola.hu3/3.1.3htm on 8 June 2006.

Horsdal, M. (2006, August). Life-story narratives in voluntary organisations. Retrieved from: www.lua.it./esrea/papers/Horsdal.doc on 9 August 2006.

Johnson, K. E. (2006). The sociocultural turn and its challenges for second language teacher education. *TESOL Quarterly, 40* (1), 235-257.

Jones Mourao, S. (2006). Understanding and sharing: English storybook borrowing in Portuguese pre-schools. In J. Enever, & G. Schmid-Schönbein (Eds.), *Picture books and young learners of English* (pp. 49-58). München: Langenscheidt.

Korky, P., & Thomas, V. (1997) *Winnie the witch.* Oxford: Oxford University Press.

Krashen, S. (1985). *The input hypothesis: Issues and implications.* London: Longman.

László, J. (2003). A második kognitív forradalom: A narratív pszichológia és a kognitív pszichológia viszonya. In J. László, J. Kállai, & T. Bereczei (Eds.), *A reprezentáció szintjei* (pp. 66-87). Budapest: Gondolat-Osiris.

Lemke, J. L. (2002). Language development and identity: Multiple timescales in the social exology of learning. In C. Kramsch (Ed.), *Language acquisition and language socialization* (pp. 68-87). London: Continuum.

Lévi-Strauss, C. (1966). *The savage mind.* Chicago: University of Chicago Press.

Marriott, S. (1998). Picture books and the moral imperative. In J. Evans (Ed.), *What's in the picture? Responding to illustrations in picture books* (pp.1-24). London: Paul Chapman Publishing Ltd.

Millard, E. (1997.) *Differently literate: Boys, girls and the schooling of literacy.* London: Routledge Falmer Press.

Nikolov, M. (1999). 'Natural born speakers of English': Code switching in pair- and group-work in Hungarian primary classrooms. In S. Rixon (Ed.), *Young learners of English: Some research perspectives* (pp. 72-89). London: Longman.

Nikolov, M. (2000). 'We do what we like': Negotiated classroom work with Hungarian children. In M. Breen, & A. Littlejohn (Eds.), *Classroom decision-making: Negotiation and process syllabuses in practice* (pp. 83-93). Cambridge: Cambridge University Press.

Nikolov, M. (2002). *Issues in English language education.* Bern: Peter Lang.

Parkes, B. (1998). Nursery chidren using illustrations in shared readings and re-readings. In J. Evans (Ed.), *What's in the picture? Responding to illustrations in picture books* (pp. 44-57). London: Paul Chapman Publishing Ltd.

Ricoeur, P. (1981). Narrative time. In W. J. T. Mitchell (Ed.), *On narrative* (pp. 165-186). London: University of Chicago Press.

Ricoeur, P. (1984). *Time and narrative. Vols. 1, 2, 3.* Chicago: University of Chicago Press.

Rogoff, B. (2003). *The cultural nature of human development.* Oxford: Oxford University Press.

Rosen, M. (2005). Are books for children worth reading? *Books for Keeps, 154*, 4-5.

Rudduck, J. (1992.) Practitioner research and programs of initial teacher education. In T. Russel, & H. Munty (Eds.), *Teachers and teaching: From classroom to reflection* (pp. 87-98). London: The Falmer Press.

Serrano-Sampedro, I. (2000). Refining negotiated classroom work in a Spanish secondary school. In M. Breen, & A. Littlejohn (Eds.), *Classroom decision-making: Negotiation and process syllabuses in practice* (pp. 108-132). Cambridge: Cambridge University Press.

Schank, R. C., & Abelson, R. P. (1995). Knowledge and memory: The real story. In R. S. Jr. Wyer (Ed.), *Advances in social cognition. Volume VIII* (pp. 1-85). Hillsdale: Lawrence Erlbaum Associates.

Schön, A. D. (1983). *The reflective practitioner: How professionals think in action.* London: Maurice Temple Smith Ltd.

Smith, F. (1985). *Reading.* Cambridge: Cambridge University Press.

Spada, N., & Massey, M. (1992). The role of prior pedagogical knowledge in determining the practice of novice ESL teachers. In J. Flowerdew, M. Brock, & S. Hsia (Eds.), *Second language teacher education* (pp. 23-37). Hong Kong: City Polytechnic of Hong Kong.

Stephens, J. (1992). *Language and ideology in children's fiction.* London: Longman.

Trivizas, E., & Oxenbury, H. (1995). *The three little wolves and the big bad pig.* London: Mammoth.

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Williams, L. (1995). Literature-based activities in a foreign-language nursery school. In R. D. Sell (Ed.), *Literature throughout foreign language education: The implications of pragmatics* (pp. 21-35). London: Modern English Publications.

Williams, M., & Burden, R. L. (1997). *Psychology for language teachers: A social constructivist approach.* Cambridge: Cambridge University Press.

Wood, D. (1988). *How children think and learn: The social contexts of cognitive development.* Oxford: Blackwell Publishers Ltd.

How Do Hungarian School Teachers of English Plan?

Éva Szabó

Eötvös Lóránd University, Budapest, Hungary

szab.eva@gmail.com

Introduction

The present article reports on a qualitative investigation into how Hungarian school teachers of English plan their teaching. The study had three main phases out of which results of the second phase, later referred to as the questionnaire survey, will be discussed in detail.

The reason why teacher planning was chosen as the focus of inquiry was that it seemed to provide an explanation to a problem of teaching English in Hungary, which came to be identified from conversations with teachers of English at primary and secondary schools, with mentor teachers and teacher trainers at teacher training colleges, and was also experienced by myself through the example of children in my own environment. What I have observed as a teacher trainer and the parent of a thirteen-year-old girl who has been learning English at the primary school for four years, and what my friends also complained about when their children changed schools is that the number of years spent with learning English at the primary school and the number of course books and teachers of English consumed during the same period of time do not seem to correlate with the amount of knowledge acquired. Very often, when thirteen-year-olds enter six-grade secondary schools they almost start studying English from scratch, as their knowledge is judged to be insufficient to continue at intermediate level. Streaming these children according to their level and placing them into the appropriate groups also seem to be often impossible because of the variety of the knowledge acquired at the different primary schools from different course books with the help of different teachers. This means that teachers have to work with mixed-level groups until the different gaps in the children's knowledge are filled. As some teachers pointed out there was simply no point in inquiring about the length of time seventh-graders had spent learning English at the primary school, as most often this piece of information proved to be meaningless. The same was confirmed by several mentor teachers at secondary schools, who often commented on the fact that the first two years of the course at a secondary school, especially at six-grade secondary schools, had to be devoted to bringing the learners to the same level and consolidating a firm knowledge base that could later be built on.

Another problem, which also appears to be related to the lack of transition between the courses at primary and secondary schools relates to choosing the appropriate course book for the group. As some of the teachers I talked to emphasized, the range of course books used by the learners at primary schools was so

wide that it was almost impossible to find one book that no learner in the group had learnt from previously. All this suggests that children entering secondary school, their families and their teachers invest a considerable amount of time, energy and money into learning English without seeing the benefits that should come out of it, which might lead to frustration, demotivation and an early disillusionment with foreign language learning in general.

My observations through my personal example and conversations with teachers were confirmed by the findings of a classroom observation project (Nikolov, 1999) carried out in order to provide insight into the teaching situation in Hungarian secondary schools before developing the new school leaving examination. The study drew attention to the lack of transition from primary to secondary schools and threw light on a number of problems associated with it. Another study by Nikolov and Nagy (2003) found that the lack of continuity between primary and secondary language learning contributed a great deal to the participants' failure to achieve an acceptable level of knowledge of a foreign language. Nikolov and Nagy pointed out that the participants of the study reported not to have been streamed according to their level of knowledge when starting the secondary school: they were either put into groups where the required level was much higher than their own level and could never really catch up with the others, or their knowledge acquired at the primary school was simply ignored and they had to start the foreign language from scratch. When studying a sample of 100 000 of year 9 secondary school students, Vágó (2005) also revealed that approximately 25 % of the respondents perceived a mismatch between their knowledge of a foreign language and the requirements they had to come up to when starting the secondary school, and the majority found that the foreign language learning requirements were too low.

What started to emerge from my observations and the reported studies was that one of the main reasons that seemed to be partly responsible for the problem described above was a lack of coordinating teaching at primary and secondary schools. The different curricula, such as the *National Core Curriculum* [Nemzeti Alaptanterv] (1995, 2003) and the *Frame Curricula* [Kerettantervek] (2000), which were designed to provide a framework for language teaching all through the primary and the secondary years do not seem to serve as the basis of planning for most teachers. Rather, these curricula determine loose guidelines as to what minimum to teach, but appear to be insufficient for helping secondary schools cope with building on learners' primary experience. This points to a need to examine what creates the base of teachers' planning decisions in their everyday teaching situations, and what considerations influence these decisions. More precisely, research into how teachers plan could give valuable insight into what factors might be responsible for the gap between primary and secondary school English teaching.

Another problem which has fostered the need for researching the planning activity of teachers of English was a perceived difficulty of the teaching of planning in pre-service teacher training courses at the teacher training college where I teach. Beyond providing a number of teaching techniques that help trainees to start planning their lessons, it has proved difficult to throw light on more important

aspects of planning, such as coordinating short-term and long-term planning, concentrating on short-term as well as long-term objectives, and most importantly, to make these aspects of planning relevant to trainees who teach altogether 15 lessons during their pre-service teaching practice. The same problem has been identified by our mentor teachers who often complained that some trainees had brilliant ideas and lots of interesting tasks for the individual lessons, but experienced serious difficulties in linking those lessons because of a lack of understanding of how to plan longer units of teaching.

Finally, I also observed that the problem of relevance in a pre-service course could often be solved by sharing teachers' stories of their work with trainees and analyzing the different aspects of teaching through these stories. Teachers' stories of how they dealt with problems in their own school environment often made seemingly irrelevant information gain relevance in the eyes of trainee teachers, which led me to build my inquiry on teachers' self-reports elicited by a questionnaire and interviews.

A revision of the relevant literature showed that research on teacher planning had three main characteristics, which justified the need to carry out research into the planning activity of teachers of English in Hungary. First of all, most research was carried out in the United States or in Great Britain, in educational contexts which are quite different from the Hungarian one. Second, it was mainly focusing on the planning activity of teachers who teach at the elementary level (Clark & Yinger, 1979; McCutcheon, 1980; Sardo-Brown, 1988; Yinger, 1979, 1982), and very few research programs involved teachers at secondary levels (McCutcheon & Milner, 2002; Milner, 2003; Sardo-Brown, 1993, 1996). It has to be pointed out that elementary teachers spend the whole school-day with one group of children and teach a number of subjects to them; therefore, their planning activity including all their considerations differs drastically from that of teachers teaching one subject to different groups of learners. It leads on to the third important feature of research on planning, which is a lack of research in the field of language teaching. Finally, though certain studies were conducted in the 1990s or later (McCutcheon & Milner, 2002; Milner, 2003; Sardo-Brown, 1993, 1996; Woods, 1996), a significant part of the relevant literature dates from the 1970s and 1980s (Clark & Yinger, 1979, 1987; McCutcheon, 1980; Morine-Dershimer, 1979; Sardo-Brown, 1988, 1990; Yinger, 1979, 1982), which also illustrates the need to devote more attention to this field.

Research design

Aim and participants

The present research was designed to focus on the teaching of English in the Hungarian educational context both at primary and secondary level using qualitative research methods according to guidelines provided by Patton (1990), Maykut and Morehouse (1994), and Cohen, Manion and Morrison (2000).The research, therefore, had an exploratory and descriptive focus and tried to describe teacher plan-

ning relying on the research participants' words and meanings. Its aim was not to be the generalization of results, but a deeper understanding of teacher planning from the perspectives of the teachers involved in the research. Research participants were selected by means of convenience sampling where the principle of maximum variety was followed in terms of the type of school the research participants were teaching at and the amount of teaching experience they had had.

The trustworthiness of the study was attempted to be established by seeking to achieve credibility, transferability, dependability and confirmability (Cohen et al., 2000). Credibility was to be provided through the triangulation of different data sources and the use of different methods (questionnaires, interviews, classroom observation), transferability was to be achieved by supplying thick description of the research situations and the participants, while dependability and confirmability were to be taken care of with the help of peer debriefing. The research can also be characterized by an emergent design and an ongoing process of data collection and analysis, where data collection and analysis of the individual phases showed directions for data collection in the subsequent phases.

Phases of the research

The research had three main phases: (i) a questionnaire survey in which 84 teachers were asked about their planning activity in winter and spring 2004; (ii) interviews on planning with 14 teachers who agreed to talk about planning between 2004 autumn and 2006 winter; and a (iii) observation of one lesson of each teacher who was interviewed in the first phase and a follow-up discussion with the teacher after the lesson between 2004 autumn and 2006 winter. The purpose of the questionnaire phase was to outline a picture of the planning activity of a large number of teachers in the ten most important aspects of planning and to identify the most and the least characteristic features of their planning. The purpose of the interviews was to refine the picture by identifying the specific considerations that led the individual teachers in taking planning decisions in their own teaching contexts, and to see whether any influential factors of planning that were not included by the questionnaire would emerge. It should also be added that though the interviews investigated the relationship of the different levels of planning, the scope was narrowed down to lesson planning as it is the most common form of planning that teachers do almost every day; therefore, it can be more easily traced than any other form of planning.

The purpose of the third phase of the research, the classroom observation phase, was to illustrate through concrete examples how planning occurred and to triangulate findings of the interviews. The follow-up discussions after the observed lessons were intended to identify what particular considerations guided concrete planning decisions and to clarify in what way the individual lessons fit into teachers' longer-term plans.

The first phase of the research: The questionnaire survey

The questionnaire survey had three main parts. Part 1 was intended to elicit background information about the respondent. This was later to be used in the statistical analysis of the findings and also in providing the description of the interview participants. Part 2 was the main body of the questionnaire containing 59 statements on planning, which respondents had to rate on a four-point Likert scale ranging from '1' meaning 'not at all characteristic of me' to '4' meaning 'absolutely characteristic of me'. Part 3 included an open-ended questionnaire item, which invited respondents to share their individual thoughts on planning and to comment on any parts of the questionnaire. Anonymity and confidentiality were guaranteed, but respondents were asked to volunteer for the interview phase and the classroom observation phase.

The following part of the present paper will focus on the main body of the questionnaire, also referred to as Part 2 of the questionnaire by giving a detailed account of its analysis and its most important findings.

The main body of the questionnaire: Part 2

Method of designing the questionnaire items

The questionnaire items of Part 2 were designed on the basis of six preliminary interviews on planning that were conducted with teachers who teach at different primary and secondary schools and whose teaching experience ranged from 5 to 40 years. The interviews had a loose interview guide consisting of four questions used as reminders to encourage the interview participants to speak freely about their own planning activity. The questions were the following:

1. How do you plan your teaching? What comes to your mind when you hear the phrase 'planning teaching'?
2. How do you put together the material that you use in teaching?
3. How do you use the curriculum, or the syllabus?
4. How do the different group characteristics affect your planning?

The interviews were then analyzed with the constant comparative method, and 103 units of meaning were identified in them. Then relationships between the units of meaning were looked for and 24 larger categories were set up. The 24 categories were finally arranged in 10 larger groups, which were seen as representing the ten most important areas of planning. The ten groups developed this way were compared to the most important areas of planning revealed by previous studies (Table 1) and similarities and differences were identified. When constructing the questionnaire, the analysis of the preliminary interviews was used in such a way that each of the 24 categories was addressed by at least one item. To reverse it, and to look at it from the perspective of the ten large areas of planning, each area representing an important aspect of planning came to contain a certain number of statements.

Table 1: The ten most important areas of planning identified by the preliminary interviews in the questionnaire survey and the most important areas of planning identified in previous studies

Areas of planning identified in the present research	Areas of planning identified in previous studies
(1) Levels of planning	(1) Levels of planning
(2) Aspects of lesson planning	(2) Reasons for planning
(3) Written plans and mental planning	(3) Written plans and mental planning
(4) The teacher's affective needs	(4) Teacher knowledge and beliefs
(5) Teaching experience	(5) Teaching experience
(6) Group characteristics	(6) Group characteristics
(7) Documents: curricula and syllabuses	(7) Documents: curricula and syllabuses
(8) Teaching materials	(8) Teaching materials
(9) Exams and marking	(9) Tests, marking
(10) Team membership	(10) Exams
	(11) Team membership

Results of preliminary interviews compared to results of previous studies

As can be seen in Table 1, the most important areas of planning identified by the present research largely overlap with those reported by previous studies, but there are a number of differences, which are listed below. To sum it up, in the present research

- reasons for planning are integrated into the different groups, such as *Aspects of lesson planning* or *The teacher's affective needs*, as they emerged as strongly related to these aspects, and not so much as independent features.
- Teacher knowledge and beliefs were not identified by the preliminary interviews as either smaller units of meaning or larger groups representing important features of planning, and are, therefore, not listed among the most important aspects of planning. The present research rests on the assumption that all teacher thinking and teachers' actions are based on teacher knowledge and beliefs and regards them as underlying all aspects of planning, but it does not intend to reveal them and analyze them.
- Tests and marking appear in the same group as exams, as they emerged from the preliminary interviews as all being strongly interrelated.
- There are two groups (*Aspects of lesson planning* and *The teacher's affective needs*) that appeared to represent such importance to the interviewees that separate groups were set up to include them.

Method of data analysis in Part 2 of the questionnaire

The data of Part 2 of the questionnaire were analyzed in two major phases. First, a picture of each of the ten large areas was provided by carrying out descriptive statistical analysis of the ratings given to the statements included by that group: the mean, the mode, standard deviation and the minimum and maximum values

were studied. Second, statements with the highest means (over 3.4) and those with the lowest means (under 1.6) were collected to identify the most and the least characteristic features of the respondents' planning activity.

Findings

The most interesting findings of the analysis within the ten most important areas of planning from the perspective of the problem that was presented at the beginning of this paper seem to be related to (i) the different levels of planning (Group 1), (ii) the use of the different curricula and syllabuses (Group 7) and to (iii) the use of teaching materials (Group 8).

Different levels of planning

As far as the different levels of planning are concerned, the analysis revealed that teachers plan their teaching at four levels in the following order of frequency: (1) long-term, (2) yearly, (3) unit and (4) weekly planning (lesson planning is not listed here, as it was found to be the most frequently practised form of planning and was, therefore, devoted a separate group within the analysis). Table 2 shows the means and standard deviation calculated for the questionnaire items on the different levels of planning. Out of the four levels of planning, long-term planning appeared to be the most characteristic of all the respondents planning activity, as indicated by the mean of Item 7 (3.66), and the relatively low value of standard deviation (0.47) shows a homogenous set of answers to Item 7.

Table 2: Mean scores and standard deviation calculated for questionnaire items on the different levels of planning

Items on levels of planning	Number of respondents (max.: 84)	Mean (max.: 4)	Standard deviation
1. I think over what I am going to teach during the year at the beginning of the school year.	84	3.3	0.71
7. I bear in mind what the students should achieve by the end of their studies.	83	3.66	0.47
8. Before starting to teach a new unit of the course book, I think over how I will teach it.	84	3.29	0.7
9. I plan what I will teach the following week before I start the week.	84	2.92	0.81

Use of curricula and syllabuses

As to the use of major curricula, such as the local curriculum, the *Frame Curriculum*, and the *National Core Curriculum*, the most often used one was found to be the local curriculum, which was followed by the *Frame Curriculum*, and the *National Core Curriculum*. Table 3 shows the means and the standard deviation calculated for questionnaire items on the use of curricula and syllabuses. As it can be seen from the mean of scores given to Item 5 (3.0), writing a syllabus for an academic year was rated as being mostly characteristic by respondents, but the mean of scores given to Item 6 (2.43) indicate that following the same syllabus during the school year was found to be less characteristic. It should also be pointed out that the values of standard deviation calculated for all items on the use of curricula and syllabuses are quite high, especially the one calculated for Item 5 on writing a syllabus for the academic year (1.24), which shows that the answers were very much varied. This indicates that schools and teachers differ a great deal in this respect and raises the question whether it is individual considerations or different school requirements that are responsible for the distribution of the answers.

Table 3: Mean scores and standard deviation calculated for questionnaire items on the use of curricula and syllabuses

Items on the use of curricula and syllabuses	Number of respondents (max.: 84)	Mean (max: 4)	Standard deviation
2. When I plan the year, I plan in accordance with the guidelines of the National Core Curriculum.	83	2.46	0.97
3. When I plan the year, I plan in accordance with the guidelines of the frame curriculum.	83	2.54	0.99
4. When I plan the year, I plan in accordance with the local curriculum.	81	2.85	1.03
5. I write a year-syllabus at the beginning of the school year.	82	3.0	1.24
6. I follow my year-syllabus during the school year.	81	2.43	0.92

Use of teaching materials

Findings within the third important group, which is the group of teaching materials, show that it is the course book that provides the framework for planning and teaching, but it is adapted to the special needs of the learners and the circumstances and is often supplemented with other materials. The means of scores given to questionnaire items on the use of teaching materials listed in Table 4 show that

the exclusive use of the course book was found to be very little characteristic by the respondents, which is indicated by the low means calculated for Item 25 (1.49), Item 28 (1.63) and Item 29 (1.48). According to the mean of scores given to Item 39 (2.29), more than half of the material used in teaching is taken from the course book, which points to the role of the course book as the basic framework of teaching. The fact that the course book is always supplemented with other materials is highlighted by the high mean of scores given to Item 33 (3. 66) and its low value of standard deviation (0.54), while the mean calculated for Item 41 (3.04) indicates that it is also the teacher's affective need to use materials other than the course book. The means of scores calculated for Items 27 (2.32) and 40 (3.04) show that units of the course book are exploited according to the teacher's plan and are adapted to the specific circumstances.

Table 4: Mean scores and standard deviation calculated for questionnaire items on the use of teaching materials

Items on the use of teaching materials	Number of respondents (max.: 84)	Mean (max: 4)	Standard deviation
25. We only use the course book in the lessons.	83	1.49	0.73
27. When teaching a unit from the course book, I follow the order of the exercises.	83	2.32	0.81
28. I do not supplement good course books with extra teaching materials.	84	1.63	0.72
29. If I teach too many hours a week, I only use the course book.	84	1.48	0.71
33. I supplement the course book with different materials.	84	3.66	0.54
39. At least half of the material that I use in teaching comes from books or sources other than the course book.	84	2.29	0.86
40. When teaching a unit from the course book, I vary the order of the exercises according to my own plan.	81	3.04	0.75
41. I feel bored if I only use the course book.	83	3.04	0.89

Most and least characteristic features of planning

In order to see how all these findings take us closer to understanding what specific reasons lie behind the problem presented earlier, it is important to have a look at the most and the least characteristic features of planning. Out of the features that were identified, three seem to be of primary importance in the continuation of the investigation. The first one confirms what has already been revealed, namely that the form of teacher planning most often practised was long-term planning (see Table 2, mean of Item 7). The second most important feature of planning identified by the study is that one of the most important components of planning is the use of the course book and the selection of supplementary materials; in other words, the selection of teaching materials (see Table 4, mean of Items 25, 28, 29, 33, 39, 41). The third one is that there is no teacher–proof course book. A closer look at Table 5 confirms that no matter what course book teachers teach from, they do think over what and how to teach and it is their affective need to do so. This was suggested by the high mean scores calculated for Items 20 (3.61) and 21 (3.57), and the low value of standard deviation indicating a homogenous set of answers to Items 20 and 21. The low mean of Item 26 (1.53), on the other hand, seems to support that no course book can be used without thinking over what to teach and how to teach it.

Table 5: Questionnaire items on the need of planning and teacher thinking

Items on the need of planning and teacher thinking	Number of respondents (max.: 84)	Mean (max.: 4)	Standard deviation
20. If I can clearly see what I will do, I feel better in the lesson.	84	3.61	0.57
21. If I think over what I will do in the lesson, it makes me feel more confident.	84	3.57	0.64
26. The course book that I use does not require prior thinking about what to teach and how to teach it.	79	1.53	0.67

Implications for subsequent data collection

If these findings are interpreted, a number of questions arise which should be answered by later phases of the investigation and by other research projects. The questions resulting from the analysis also reflect an awareness of the fact that the data were collected from teachers' self-reports. In other words, the data show what teachers say, but they do not reveal in what way teachers' own – perhaps subjective - interpretations of certain concepts influence the results. Therefore, in the following phases of the research teachers' interpretations should also be clarified.

The questions that need to be further investigated are the following:

- What exactly do teachers mean by long-term planning? In other words, what objectives are set in long-term plans and what is the content of long-term planning?
- Could long-term plans be coordinated in such a way as to provide continuity across different schools and school types?
- If teachers do not seem to rely on curricula, then what else can help them when they plan?
- What is wrong with year syllabuses? Why do teachers not follow them?
- Once planning and teaching are based on course books, how could course books provide a framework that helps to establish continuity between primary and secondary levels?

Though the interviews that were conducted in the subsequent phase have not yet been analyzed in detail, certain answers already emerge from them. One possible answer is that long-term planning usually means thinking over the school year at the beginning of the year, but very often these long-term plans are forgotten, and teachers concentrate on short-term achievement. Another problem might be that teachers' concern of how to spend the 45 minutes of a lesson might outweigh other considerations as to what to achieve by the end of the school year. Another suspicion is that one of the main reasons for supplementing the course book is that course books are often judged to be quite insufficient for achieving certain objectives.

Summary

In this paper I presented a study on teacher planning, with special emphasis on the results of a questionnaire survey, which was the first phase of the study involving teachers of English at primary and secondary schools in Hungary. The aim of the questionnaire survey was to outline a picture of the participating teachers' planning activity, which will be further refined in the second and third phases of the study through interviews and the observation of some of the research participants' lessons in order to see what concrete considerations shape planning in specific classroom contexts. The final aim of the study will be to identify in what way teacher planning is responsible for the lack of continuity between primary and secondary school English teaching.

The findings of the questionnaire survey point to a need to conduct further inquiry into teachers' interpretation of the different levels of planning and the way they coordinate these levels, the factors that create the base of teachers' planning activity, as well as the role of course books in planning and teachers' interpretation of the shortcomings of course books. The data collected in the second and third phases will hopefully take us closer to an understanding of how teachers' planning activity is related to the lack of continuity and will also provide clues as

to what could be done in order to provide smooth transition between primary and secondary school English teaching.

References

Clark, C. M., & Yinger, R. J. (1979). *Three studies of teacher planning*. Research Series No. 55. East Lansing: Michigan State University.

Clark, C. M., & Yinger, R. J. (1987). Teacher planning. In J. Calderhead (Ed.), *Exploring teachers' thinking* (pp. 84-103). London: Cassell.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education*. London: Routledge.

Kerettantervek : Az alapfokú nevelés-oktatás kerettantervei. A középfokú nevelés-oktatás kerettantervei (Gimnázium). A középfokú nevelés-oktatás kerettantervei (Szakközépiskola, szakiskola). (2000). Budapest: Oktatási Minisztérium.

Maykut, P., & Morehouse, R. (1994). *Beginning qualitative research*. London: The Falmer Press.

McCutcheon, G. (1980). How do elementary school teachers plan? The nature of planning and influences on it. *The Elementary School Journal, 81* (1), 4-23.

McCutcheon, G. & Milner, H. R. (2002). A contemporary study of teacher planning in a high school English class. *Teachers and Teaching: Theory and practice, 8* (1), 81-94.

Milner, R. (2003). A case study of an African American teacher's cultural comprehensive knowledge and self-reflective planning. *Journal of Curriculum and Supervision, 18* (2), 175-196.

Morine-Dershimer, G. (1979). *Teacher plan and classroom reality: The South Bay study: Part 4.* Research Series No. 60. East Lansing: Michigan State University.

National Core Curriculum [Nemzeti alaptanterv]. (1995). Budapest: Művelődési és Közoktatási Minisztérium.

National Core Curriculum [Nemzeti alaptanterv]. (2003). Budapest: Oktatási Minisztérium.

Nikolov, M. (1999). Classroom observation project. In H. Fekete, É. Major, & M. Nikolov (Eds.), *English language education in Hungary. A baseline study* (pp. 221-245). Budapest: The British Council Hungary.

Nikolov, M., & Nagy, E. (2003). "Sok éve tanulok, de nem jutottam sehova": Felnőttek nyelvtanulási tapasztalatai ['I have been learning for a long time, but I have not achieved anything': Language learning experience of adults]. *Modern Nyelvoktatás, 9* (1), 14-41.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park: Sage Publications.

Sardo-Brown, D. (1988). Twelve middle-school teachers' planning. *The Elementary School Journal, 89* (1), 69-87.

Sardo-Brown, D. (1990). Experienced teachers' planning practices: A US survey. *Journal of Education for Teaching, 16* (1), 57-71.

Sardo-Brown, D. (1993). Description of two novice secondary teachers' planning. *Curriculum Inquiry, 23* (1), 63-84.

Sardo-Brown, D. (1996). A longitudinal study of novice secondary teachers' planning: Year two. *Teaching and Teacher Education, 12* (5), 519-530.

Vágó, I. (2005). Kilencedik évfolyamos diákok idegennyelv-tanulási útjai [Language learning paths of year 9 students]. Unpublished manuscript. Budapest: Országos Közoktatási Intézet Kutatási Központ.

Woods, D. (1996). *Teacher cognition in language teaching*. Cambridge: Cambridge University Press.

Yinger, R. (1979). Routines in teacher planning. *Theory and practice, 18* (3), 163-169.

Yinger, R. J. (1982). A study of teacher planning. In W. Doyle, & T. L. Good (Eds.), *Focus on teaching: Readings from the Elementary School Journal* (pp. 239-259). Chicago: The University of Chicago Press.