DOKTORI DISSZERTÁCIÓ

# CREATIVITY AND ORAL NARRATIVE TASK PERFORMANCE: A STUDY OF FIRST YEAR ENGLISH MAJORS

ALBERT ÁGNES

2008

Eötvös Loránd Tudományegyetem Pedagógiai és
Pszichológiai Kar

DOKTORI DISSZERTÁCIÓ

ALBERT ÁGNES

# CREATIVITY AND ORAL NARRATIVE TASK PERFORMANCE: A STUDY OF FIRST YEAR ENGLISH MAJORS

# SZÓBELI NARRATÍV FELADATOKON NYÚJTOTT TELJESÍTMÉNY ÖSSZEFÜGGÉSEI A KREATIVITÁSSAL ELSŐÉVES ANGOLSZAKOS EGYETEMISTÁK KÖRÉBEN

ELTE PPK Neveléstudományi Doktori Iskola
Iskolavezető: Dr. Bábosik István, DSc., egyetemi tanár
Nyelvpedagógia Doktori Program
Programvezető: Dr. Károly Krisztina, PhD. habil. egyetemi docens

Témavezető: Dr. Dolgosné Dr. Kormos Judit, PhD. habil. egyetemi adjunktus

A bíráló bizottság:
Elnök: Dr. Barkóczi Ilona DSc., professor emeritus
Opponensek: Dr. Csölle Anita PhD.
Dr. Ottó István PhD.
A bizottság tagjai:
Dr. Heltai Pál PhD., egyetemi docens
Dr. Károly Krisztina PhD., egyetemi docens
Dr. Tankó Gyula, PhD., egyetemi adjunktus
Titkár: Dr. Csizér Kata PhD.

Budapest, 2008 február 9.

# Abstract

Changes in language instruction that involve greater reliance on learners' creativity imply that researching creativity as a potentially important individual variable should be imminent. The prominence of tasks in the classroom and in tests suggests that tasks and their decisive features leading to differences in task performance should also be investigated.

The aim of this dissertation was to investigate the relationships between a potentially important individual difference, creativity, and performance on oral narrative tasks. Participants of the study were 41 1st-year English majors studying at ELTE, whose creativity was measured with the help of a standardised test and whose oral narrative task performance was examined on two narrative tasks differing in cognitive complexity. Besides, their language aptitude and level of proficiency was also assessed. Connections were examined between the two individual variables: language aptitude and creativity, between each individual variable and language proficiency, and between the individual variables, language proficiency and oral narrative task performance. Moreover, performance on the cognitively less and more complex tasks was examined in detail as well.

The findings of this study suggest that despite a positive relationship hypothesised on the basis of the literature, language aptitude and creativity seem to be negatively correlated. In the case of advanced learners, the relationships between the individual difference variables and language proficiency measures tend to be rather weak. With regard to creativity, only the fluency-free components of average originality and relative flexibility seem to be related to English proficiency. Performance on the cognitively less and more complex oral narrative tasks differs in the following respects, the cognitively less complex task results in greater fluency and lexical diversity, while the cognitively more complex task urges participants to be more accurate and talk more. More proficient learners seem to allocate their resources differently in the case of the two tasks, that is, they seem to prioritise different areas. Participant characterised by a higher level of language aptitude tend to solve the tasks in a manner similar to the more proficient students; performance on the cognitively less complex task seems to be more heavily determined by aptitude. On the contrary, creativity seems to be more strongly related to performance on the cognitively more complex task. The majority of these findings can be interpreted within the framework of the Cognition Hypothesis put forward by Robinson (2003).

As regards the relationship of creativity and oral narrative task performance, in line with the results of an earlier exploratory study (Albert & Kormos, 2004) it seems that the three components of creativity have a differential effect on the measures of task performance. However, since the relationships discovered are not entirely compatible with findings of the earlier study, further research is needed to clarify their connections. Results of this study suggest that creative individuals' superior general retrieval ability (Carroll, 1993) might be held responsible for the greater fluency, more talk and greater lexical variety exhibited by those participants who were either characterised by greater average originality or relative flexibility.

# Acknowledgements

First and foremost my gratitude goes to my supervisor, Dr Dolgosné Dr Kormos Judit, for her continuous and unwavering support and encouragement throughout the long years of completing my PhD studies, and for her invaluable insights and thorough comments during the writing process. All remaining shortcomings are, of course, my own responsibility. I would also like to thank professor Barkóczi Ilona for implanting an interest in creativity in me when I first investigated the topic under her supervision as an undergraduate student of psychology. I am also thankful to Dr Ottó István for allowing me to use his Hungarian test of language aptitude (HUNLAT) and providing me with the scores. Special thanks go to my tutors at the PhD programme, especially to Dr Kontra Edit, Dr Holló Dorottya and professor Charles Alderson, who helped me with their comments on earlier versions on parts of the dissertation. I am also indebted to my colleague Dóczi Brigitta for her help in recruiting participants. Last but not least I am immensely grateful to those students who were willing to participate in this project, and without whom conducting this study would have been impossible: Apáti Edit, Bagjos Áron, Bartha Zonga, Békefi Éva, Budai Zsolt, Czapkó Krisztina, Csemez Tamás, Csiky Anna, Csuba Katalin, Danyi Orsolya, Dselits Dávid, Erdős Dorottya, Feró Dalma, Földes Kata, Galácz Karolina, Havasi Andrea, Herman Zita, Hibsch Sándor, Hunyadi Laura, Kántor Eszter, Kézdy Orsolya, Kivovics Judit, Klausz Ágnes, Kormos Zsuzsanna, Kovács Pál, Krausz Edina, László Péter, Lázár Noémi, Marczingós Krisztina, Morovik Gabriella, Réti Péter, Szabó Adrienn, Szabó Katalin, Szabó Tünde, Szőke Viola, Tarjáni Zsófia, Tóth Katalin, Török Eszter, Udvardy Eszter, Veres Csaba, Wirth Balázs.

# Table of Contents

Abstract ..... 3
Acknowledgements ..... 4
Table of Contents ..... 5
List of Tables ..... 9
List of Figures ..... 11
Chapter 1: Introduction ..... 12
1.1 Rationale and aim of the dissertation ..... 12
1.2 Research questions ..... 13
1.3 Overview of the dissertation ..... 16
Chapter 2: Theoretical Background of the Research ..... 19
2.1 Introduction ..... 19
2.2 The role of individual variables in learning a second language ..... 19
2.2.1 An influential ID variable - Foreign language aptitude ..... 24
2.2.2 A potentially important ID variable - Creativity ..... 29
2.2.2.1 Early theories of creativity ..... 29
2.2.2.2 Psychometric theories of creativity ..... 31
2.2.2.3 Current theories of creativity ..... 33
2.2.2.4 Cognitive components of creativity ..... 34
2.2.2.5 Creativity and perception ..... 36
2.2.2.6 Creativity and memory ..... 38
2.2.2.7 Metacognition in creativity ..... 39
2.2.2.8 Measuring creativity ..... 40
2.2.2.9 Conclusion ..... 42
2.3 The role of tasks ..... 43
2.3.1 Important changes in language instruction: Communicative and task-based language learning ..... 44
2.3.2 Task definitions in TBLT ..... 46
2.3.3 Tasks in language testing ..... 47
2.3.4 Task types ..... 48
2.3.5 Theories of task characteristics ..... 50
2.3.5.1 Characteristics affecting task difficulty ..... 50
2.3.5.2 Robinson's Cognition Hypothesis ..... 54
2.3.6. Evaluating task performance ..... 56
2.3.7 Empirical research on narrative tasks ..... 58
2.3.8 Conclusion ..... 62

2.4 Conclusion: The relevance of creativity for learner performance on tasks.....63
Chapter 3: Pilot Studies.....65
3.1 Introduction.....65
3.2 MA Thesis.....65
3.3 Validation Study.....67
3.3.1 Introduction.....67
3.3.2 Oral narrative tasks described in the literature.....68
3.3.3 Issues of validity in the testing literature.....69
3.3.4 Narrative genre.....70
3.3.5 Research question.....72
3.3.6 Method.....73
3.3.6.1 Participants.....73
3.3.6.1 Instruments.....73
3.3.6.1.1 Oral narrative tasks.....73
3.3.6.1.2 Post-task interview.....74
3.3.6.2 Procedure.....75
3.3.6.3 Analysis.....75
3.3.7 Results and discussion.....76
3.3.7.1 Single picture task.....76
3.3.7.2 Picture sequence task.....79
3.3.7.3 Cartoon strip task.....81
3.3.7.4 Comparative analysis.....83
3.3.8 Conclusion and implications.....87
Chapter 4: Method.....88
4.1 Introduction.....88
4.2 Definition of constructs.....88
4.2.1 Creativity.....88
4.2.2 Oral narrative task.....89
4.3 Design.....89
4.4 Participants.....89
4.5 Instruments.....90
4.5.1 Creativity test.....91
4.5.2 Oral narrative tasks.....91
4.5.3 Proficiency tests (TOEFL-PBT, C-Test).....92
4.5.4 Language aptitude test (HUNLAT).....93
4.6 Procedures.....94
4.7 Analysis.....94

4.7.1 Measures of creativity ..... 94
4.7.2 Measures of task performance ..... 97
4.7.3 Measures of proficiency ..... 101
4.7.4 Measures of language aptitude ..... 102
4.8 Statistical analysis ..... 102
Chapter 5: Findings Concerning Individual Differences in Abilities and Proficiency ..... 103
5. 1 Introduction ..... 103
5.2 Results ..... 103
5.2.1 Creativity test ..... 103
5.2.2 Aptitude test ..... 109
5.2.3 Proficiency tests ..... 111
5.3 Discussion ..... 113
5.3.1 Creativity ..... 113
5.3.2 Aptitude ..... 115
5.3.3 Proficiency ..... 116
5.4 Conclusion ..... 117
Chapter 6: Task Performance on Cognitively Less and More Complex Tasks ..... 118
6.1 Introduction ..... 118
6.2 Results ..... 119
6.2.1 Oral narrative task variants ..... 119
6.2.2 Characteristics of the oral narrative tasks differing in cognitive complexity ..... 121
6.2.3 Comparing the tasks differing in cognitive complexity ..... 124
6.3 Discussion ..... 126
6.3.1 Oral narrative task variants ..... 126
6.3.2 Characteristics of the oral narrative tasks differing in cognitive complexity ..... 127
6.3.3 Comparing the tasks differing in cognitive complexity ..... 128
6.4 Conclusion ..... 131
Chapter 7: Correlations of Individual Differences and Proficiency ..... 133
7.1 Introduction ..... 133
7.2 Results ..... 133
7.2.1 Correlations of aptitude and proficiency ..... 133
7.2.2 Correlations of creativity and proficiency ..... 134
7.2.3 Correlations of creativity and aptitude ..... 137
7.3 Discussion ..... 141
7.3.1 Correlations of aptitude and proficiency ..... 141

7.3.2 Correlations of creativity and proficiency ..... 142
7.3.3 Correlations of creativity and aptitude ..... 145
7.4 Conclusion ..... 147
Chapter 8: Correlations of Individual Differences in Proficiency and Abilities with Task Performance Measures ..... 150
8.1 Introduction ..... 150
8.2 Results ..... 150
8.2.1 Correlations of proficiency and task performance ..... 150
8.2.2 Correlations of aptitude and task performance ..... 154
8.2.3 Correlations of creativity and task performance ..... 158
8.2.3.1 Correlations with subtests of creativity ..... 158
8.2.3.2 Correlations with composite scores of creativity ..... 163
8.3 Discussion ..... 165
8.3.1 Correlations of proficiency and task performance ..... 165
8.3.2 Correlations of aptitude and task performance ..... 169
8.3.3 Correlations of creativity and task performance ..... 172
8.4 Conclusion ..... 177
Chapter 9: Conclusions and Pedagogical Implications ..... 179
9.1 Introduction ..... 179
9.2 Summary of findings ..... 179
9.3 Limitations ..... 183
9.4 Pedagogical implications and directions for future research ..... 185
References ..... 188
Appendix A -Single picture task A ..... 205
Appendix B - Single picture task B ..... 206
Appendix C - Picture sequence task A ..... 207
Appendix D - Picture sequence task B ..... 208
Appendix E - Cartoon strip task A ..... 209
Appendix F - Cartoon strip task B ..... 210
Appendix G - Post-task interview ..... 211
Appendix H - Evaluation criteria for the oral narrative task ..... 212
Appendix I - Creativity test ..... 214
Appendix J - New picture sequence task A ..... 220
Appendix K - New picture sequence task B ..... 221
Appendix L - C-test used in the study ..... 222
Appendix M - Sample tasks from HUNLAT ..... 223

# List of Tables

Table 1 Respondents' performance on the single picture tasks ..... 78

Table 2 Respondents' performance on the picture sequence task ..... 80

Table 3 Respondents' performance on the cartoon strip task ..... 83

Table 4 Descriptive statistics of the total number of AS-units in the three tasks ..... 84

Table 5 Descriptive statistics of the total number of narrative clauses in the three tasks ..... 85

Table 6 Descriptive statistics of the ratio of the total number of narrative clauses to the total number of AS-units in the three tasks ..... 86

Table 7 Measures of creativity ..... 97

Table 8 Measures of task performance ..... 100

Table 9 Descriptive statistics of the four subtests of the standardised creativity test for 1st year English majors at ELTE (N=41) ..... 104

Table 10 Descriptive statistics of the four subtests of the standardised creativity test for a representative sample of high school graduates (N=1,098), the Hungarian national standard ..... 105

Table 11 Intercorrelations of composite measures of creativity ..... 106

Table 12 Multitrait - multimethod matrix of the creativity test ..... 108

Table 13 Descriptive statistics of the aptitude test HUNLAT for 1st year English majors at ELTE (N=41) ..... 110

Table 14 Descriptive statistics of the aptitude test HUNLAT for 1st year university students (N=130) and for 1st year university students with at least one language exam (N=65) ..... 110

Table 15 Intercorrelations of the aptitude test HUNLAT ..... 111

Table 16 Descriptive statistics of language proficiency tests for 1st year English majors at ELTE (N=41) ..... 112

Table 17 Intercorrelations of proficiency test scores ..... 113

Table 18 Independent samples t test for Task 1 versions a and b ..... 120

Table 19 Independent samples t test for Task 2 versions a and b ..... 121

Table 20 Intercorrelations of measures of Task 1 ..... 122

Table 21 Intercorrelations of measures of Task 2 ..... 123

Table 22 Paired samples t tests, comparing performance on Tasks 1 and 2 ..... 125

Table 23 Correlations of the same task performance measures on Tasks 1 and 2 ..... 126

Table 24 Correlations of aptitude test scores and scores on language proficiency tests for 1st year English majors at ELTE (N=41) ..... 134

Table 25 Correlations between different tasks of the creativity test and language proficiency scores ..... 135

Table 26 Correlations between composite creativity scores and language proficiency test scores for 1st year English majors at ELTE (N=41) ..... 137

Table 27 Correlations between different tasks of the creativity test and language aptitude scores ..... 138

Table 28 Correlations between composite creativity test scores and language aptitude test scores for 1st year English majors at ELTE (N=41) ..... 140

Table 29 Correlations between language proficiency and task performance measures for Tasks 1 and 2 (N=41) ..... 153

Table 30 Correlations between language aptitude and task performance measures on Tasks 1 and 2 ..... 157

Table 31 Correlations between the creativity test task Unusual Uses and task performance measures on Tasks 1 and 2 ..... 159

Table 32 Correlations between the creativity test task Remote Associations and task performance measures on Tasks 1 and 2 ..... 160

Table 33 Correlations between the creativity test task Circles and task performance measures on Tasks 1 and 2 ..... 161

Table 34 Correlations between the creativity test task Picture Completion and task performance measures on Tasks 1 and 2 ..... 162

Table 35 Correlations between composite scores of creativity and task performance measures on Tasks 1 and 2 ..... 163

# List of Figures

Figure 1 The Triadic Componential Framework for task classification - categories, criteria, analytic procedures, and design characteristics ..... 53

# Chapter 1: Introduction

Motto:
"Imagination is more important than knowledge. For while knowledge defines all we currently know and understand, imagination points to all we might yet discover and create." (Albert Einstein)

## 1.1 Rationale and aim of the dissertation

Many of the individual differences that exist between learners have been studied in an attempt to account for the differential success in second language acquisition. The relevance of several cognitive, motivational, personality and social factors has been revealed, but there is one complex phenomenon, the importance of which has not been thoroughly explored to this day, and this is learner creativity. If the creative process is regarded as a rare phenomenon observable only in the exceptionally talented, 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; Mednick, 1962) 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, 1985a), might be a factor that affects 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, provides 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.

It is not obvious, however, whether the effects of an individual variable like creativity can only be observed in general outcomes of second or foreign language learning, such as achievement as reflected by English grades (Ottó, 1998) and presumably proficiency test results, or if these effects can also be detected in much smaller and more specific units of learner performance, such as tasks. Since tasks are recently regarded as having central importance in language teaching and also in testing, it should be interesting to investigate the effects of learner creativity on several output variables of one particular task type, the oral narrative task. Since oral narrative tasks generally involve storytelling based on some cue, this task type seems to lend an opportunity for learners to use their imagination; therefore, it seems to be particularly well suited for demonstrating the effects of creativity.

## 1.2 Research questions

Changes in language instruction which entail greater reliance on activities that necessitate the use of imagination on the learner's part suggest that the relevance of a new ID variable creativity, which has been unexplored so far, should be examined. Although research conducted on ID variables in the past tended to concentrate on the relationship of these variables with global measures of attainment, that is language proficiency, a current trend is to

examine the effect of individual differences on task performance (see Dewaele and Furnham, 1999, on the relationship of extraversion and oral task performance and Dörnyei and Skehan, 2003, on language aptitude). This recent interest in tasks observable both in the area of language teaching and testing can be ascribed to the assumptions that transacting tasks engages naturalistic acquisitional mechanisms and drives development forward, and that by matching task features to the characteristics of those situations where the learner is likely to use the language enhances the validity of tests. In the light of this, there seems to be a need to study possible connections between learner creativity and performance on tasks.

In an attempt to study the relationship of creativity and task performance, I chose oral narrative tasks as this task type seemed to offer a good opportunity for learners to demonstrate their creativity. Therefore, keeping feasibility in view, I intended to examine their connections in the framework of a quantitative study using a correlational research design, which unfortunately does not allow conclusions to be drawn regarding causality. Although my main research interest concerned this, that is the relationship of learner creativity and oral narrative task performance, the careful examination of the topic necessitated the introduction of further variables. Concerning individual variables, besides creativity the introduction of a well-established ID variable, language aptitude seemed justifiable. I felt that with regard to language measures, language proficiency is a variable that needs the taken into consideration besides measures of task performance. It also seemed advisable to study learners' performance on two tasks differing in cognitive complexity, as on the one hand changes in cognitive complexity tend to result in changes in performance (Robinson, 2007a, Skehan, 1998), and on the other ID differences are hypothesised to have a greater effect in the case of cognitively more complex tasks (Robinson, 2003). Consequently, the study involved measuring the participants' creativity, language aptitude and English proficiency using tests and eliciting oral narrative performance from them with the help of two tasks, a cognitively less and a cognitively more complex one. I intended to find answers to the following research questions (RQ) with the help of the study:

1. What are the characteristics of a first-year English major sample with regard to creativity, language aptitude and language proficiency?
2. How do students solve oral narrative tasks characterised by different levels of cognitive complexity? What are the main differences in their output on the two tasks?
3. Is there a relationship between the ID variables measured: language aptitude and creativity, and language proficiency?
4. Is there a relationship between the ID variables themselves, that is between language aptitude and creativity?
5. How are proficiency and task performance measures related on the cognitively less and more complex tasks?
6. How are language aptitude and task performance measures related on the cognitively less and more complex tasks?
7. How are creativity and task performance measures related on the cognitively less and more complex tasks?

The importance of RQ 1 lies in the fact that in order for findings to be generalizable to any extent, is crucial that we are familiar with the characteristics of the sample from which the results originate, unless our sample is representative, which is not the case here. Answer to this research question can be found in chapter 5 of the dissertation. RQ 2 attempts to find support for the assumption that learners solve tasks characterised by different levels of cognitive complexity differently, even if the tasks themselves are of the same basic type, that is they are both oral narrative tasks. This research question is tested in chapter 6.

The rest of the research questions concern the relationships of different variables. On the one hand, RQs 3 and 4 are concerned with global measures, that is the relationships of the ID variables with measures of language proficiency, and with the relationship of creativity and language aptitude. Possible connections between these variables are discussed in chapter 7. RQs 5, 6, and 7 on the other hand refer to specific measures: task performance measures derived from a cognitively less and a cognitively more complex task

are related to language proficiency, language aptitude and creativity. These results are discussed in chapter 8.

## 1.3 Overview of the dissertation

Having introduced the research topic and the research questions in this chapter, chapter 2 of the dissertation goes on to present the theoretical background of the study in the form of a review of literature. After a brief introduction of the field of individual difference (ID) variables relevant for language learning, the constructs of language aptitude and creativity are discussed at length. Similarly, after a short introduction highlighting the recent importance of tasks in language instruction, the concepts of task and oral narrative task are examined and empirical research conducted in connection with them is cited. The separate discussion of the concepts of creativity and oral narrative tasks is then followed by an attempt to bring possible points of interplay to light.

Chapter 3 of the dissertation presents the results of pilot studies carried out prior to conducting this research study. In section 3.1 of the chapter, the main results of an exploratory study written up as my M.A. thesis are described, followed by lessons to be learnt from the exploratory study. Section 3.2 of the chapter contains a task validation study, in the framework of which the oral narrative tasks used for the dissertation were piloted and validated. Since the process of task validation is considered an important step of carrying out research, a detailed account of it is provided.

Chapter 4 describes the methods employed when conducting the research itself. The definition of constructs is followed by the description of the design and that of participants. The instruments used: the test of creativity, the aptitude test, tests of proficiency and the oral narrative tasks with different levels of cognitive complexity are discussed in detail. Measures calculated on the basis of the tests come next, followed by a description of the statistical procedures applied.

Findings of the study are discussed in the four subsequent chapters: chapter 5 is concerned with individual differences in abilities and proficiency in the sample. The sample of the study is characterised by means of descriptive statistics along the following lines: creativity, language aptitude and language proficiency. Besides characteristics of the sample, properties of the research instruments: the creativity test, the language aptitude test and the proficiency tests are also discussed.

Chapter 6 describes task-related findings. Using statistical methods, I attempt to show that the four oral narrative tasks used in the study belong to two types, a cognitively less and a cognitively more complex one. Having established this, students' performance on each task type is described. Differences in output are then attributed to differences in the task types, that is, to differing levels of cognitive complexity.

Chapter 7 presents results of the correlational analyses of the ID variables language aptitude and creativity with proficiency. It is examined whether a relationship can be detected between the ID variables and global measures of language proficiency. Possible relationships between different aspects of creativity and language proficiency are also examined here.

Chapter 8 aims to demonstrate relationships between proficiency, individual differences and task performance measures. Task performance measures are first correlated with language proficiency to see the relationships and possible differences between the cognitively less and more complex tasks. Then the relationships of aptitude and output measures are examined, also taking into consideration differences in cognitive complexity. Finally, correlations are calculated for different aspects of creativity and task performance measures. Differences in cognitive complexity between the two tasks are also taken into consideration here.

In Chapter 9 conclusions are drawn and pedagogical implications are pointed out. First, a short summary of the most important findings are presented. This is followed by the shortcomings of the research that on the one hand point out the limitations of the study, but on the other also necessitate

further research. Finally, the pedagogical implications of the study are discussed and possibilities for future research are suggested.

# Chapter 2: Theoretical Background of the Research

## 2.1 Introduction

This chapter is concerned with establishing the theoretical background of the research study I present in this dissertation. The goal of the following review of literature is to provide a detailed account of those concepts and constructs which constitute the main focus of my investigations namely, creativity and oral narrative tasks. I also indent to demonstrate why despite the lack of research studies in the field, their investigation should be imminent. In order to put these constructs into perspective a brief overview of individual differences and tasks in language teaching is also provided. This chapter is concluded by identifying points of interplay, substantiating why and how creativity is believed to be relevant for performance on oral narrative tasks based on information available in the literature.

## 2.2 The role of individual variables in learning a second language

Since the aim of this dissertation is to examine the role of a potentially important individual variable in solving a particular language learning tasks, it seems necessary to provide a short overview of the individual differences considered significant in applied linguistics. Before attempting to summarize these, it is important to define what is meant by individual differences in the dissertation. In a recent book devoted to the topic of individual differences, Dörnyei (2005) offers the following definition which I would like to adopt here. "ID constructs refer to dimensions of enduring personal characteristics that are assumed to apply to everybody and on which people differ by degree" (p. 4).

Despite failing to make it part of the definition, Dörnyei (2005) also hints at the issue of relevance. Although a vast number of ID constructs have been identified by general an personality psychology, not all of them have

direct relevance to language acquisition. Therefore, in this dissertation similarly to Dörnyei's book, only those attributes are referred to as ID variables that have been found to be relevant with regard to language learning. Highlighting the issue of relevance is important precisely because this is one of the main questions posed in connection with creativity: that is, whether creativity is relevant to language learning, more precisely for solving oral narrative tasks.

The role of individual differences in second language learning is a very popular and well-researched topic within applied linguistics. Since there are numerous articles, book chapters and even whole books devoted to the topic (e.g. Dörnyei 2005; Dörnyei & Skehan, 2003; Ellis, 2004; Kontráné, 2004; Larsen-Freeman & Long, 1991; Oxford & Ehrman, 1993; Skehan, 1989) a comprehensive summary would be beyond the scope of this dissertation. Depending on the authors, their backgrounds, research experiences, and personal orientations, they tend to lay emphasis on different variables although there are certainly a number of core ones the importance of which everyone acknowledges. In this section, I would like to provide a very brief overview of these core variables which despite their importance were not examined in this research study.

One of the most obvious differences between learners concerns the age at which the given individual started learning the language. Although noone denies its importance, some authors (Dörnyei, 2005; Ellis, 2004) refrain from discussing it on grounds that the body of research available in connection with the topic warrants that a separate book should be written on it. Debates about this issue mainly centre around differences and similarities between child and adult language learning and the critical period hypothesis. Although most authors agree that children and adults learn languages in different ways (Krashen, 1982), the explanation of these differences can be grouped around four main themes (Ellis, 1994; Larsen-Freeman & Long, 1991).

The first of these four possible explanations suggests that different cognitive processes play a role in the case of children and adults. Whereas children use the Language Acquisition Device for learning a second language in a similar way as when they acquire their mother tongue (Johnson &

Newport, 1989; Krashen, 1982), adults rely on cognitive processes used for general problem solving. According to the neurological explanation, differences in child and adult language learning can be traced back to the loss of the plasticity of the brain due to lateralization and cerebral maturation (Scovel, 1988). Certain changes in the neurological structure of the brain taking place around puberty affect learners' capacity with regard to the acquisition of pronunciation and grammar. Another theory places the emphasis on differences in the input children and adults are exposed to (Snow, 1983). Children encounter more input of the "here-and-now" type, which makes the extraction of rules easier, whereas adults are targeted with much more complicated input. The fourth type of explanation is socio-psychological in nature and states that the reason why adults do not usually achieve native-like pronunciation is because their identity as a speaker of a particular L1 is firmly established (Brown, 1987); therefore, they may prefer to speak accented L2 so that they can express their identities.

The rest of key ID variables are often divided into two large groups, the group of cognitive and that of affective factors (Gardner & MacIntyre, 1992, 1993). Language aptitude, learning styles and learning strategies are the most widely acknowledged cognitive ones, whereas language learning motivation along with anxiety are the most frequently discussed affective factors. Although this dichotomy of cognitive and affective factors is quite appealing, there tend to be some ID variables that do not fit into either group. These are typically labelled miscellaneous and along with the factor of age discussed above, we can find sex, socio-cultural experiences, personality traits such as extraversion, learner beliefs or willingness to communicate in this category.

Motivation, the most important affective factor, unarguably has a decisive role in language acquisition since no matter how talented or smart someone is, without it learning simply does not take place. It was Gardner and Lambert (1972), who first proposed a comprehensive model of language learning motivation, the so-called socio-educational model of second language acquisition. Although this model is quite complex and includes factors such as interests, attitudes and different aspects of motivation, it is often oversimplified and reduced to the dichotomy of integrative and instrumental motivation

(Dörnyei 2005), that is, whether the learner would like to interact with members of and become part of the second language (L2) community, or they would simply like to gain specific benefits by learning the L2. The passing years along with a number of new motivation theories rooted in cognitive and personality psychology also brought a change in perspective: instead of a static view of the construct, researchers started to consider motivation as a process (Dörnyei & Ottó, 1998).

Dörnyei's (2005) most recent model of language learning motivation, the L2 Motivational Self System is also clearly process-oriented. This theory has three major components: the Ideal L2 Self describing the ideal one would like to become with regard to the L2, which can be a very powerful intrinsic motivator; the Ought-to L2 Self referring to attributes one believes they ought-to possess, therefore, these motivators are more extrinsic in nature, and the L2 Learning Experience which concerns characteristics of the immediate learning experience and environment. Although the model is intuitively appealing, it is in need of empirical verification.

Anxiety, similarly to motivation, is a construct of general psychology that found its way into applied linguistics as it regarded relevant to language learning. However, whereas general psychology differentiates debilitating and facilitating anxiety (Alpert & Haber, 1960), therefore anxiety is not altogether negative but can have positive outcomes as well, foreign learning anxiety as an ID variable tends to be seen as having purely negative bearings on performance (Dörnyei 2005). Although foreign language anxiety, which refers to the worries and negative emotions when one is learning or using a foreign language (MacIntyre, 1999), seems to be a construct which can be clearly differentiated from trait anxiety (Spielberger, 1966), test anxiety or communication apprehension (Horwitz, 2001), there are still a number of issues that need to be clarified in connection with it. These include the stability of foreign language anxiety across the different foreign languages studied and also an attempt to discover its possible facilitating aspects (Dörnyei, 2005).

Among the key cognitive ID variables it is language aptitude that has the longest research tradition. (However, since language aptitude is a cognitive ID variable which was examined in the empirical study conducted for this

dissertation, it is not going to be discussed in greater detail here, but in section 2.2.1 below.) Interest in learning styles and learning strategies is more recent and is shared by educational psychology. Learning styles and strategies are distinct but related concepts, and each of them covers a wide range of constructs which will not be discussed in detail here. According to a standard definition, learning styles refer to "an individual's natural, habitual and preferred way(s) of absorbing, processing, and retaining new information and skills" (Reid, 1995, p. viii). These are typically bipolar personal preferences that can be placed on a continuum between one extreme to another, and theoretically involve no value judgement, that is, one is not regarded more advantageous than the other. They are relatively stable and operate similarly across different situations; this is what primarily differentiates them from learning strategies (Dörnyei, 2005).

The most widely acknowledged learning styles include dimensions such as field-dependence, field-independence (Witkin, Goodenough, & Karp, 1967); diverger, converger, assimilator, and accommodator thinkers (Kolb, 1984); visual, auditive, kinaesthetic, and tactile styles (Reid, 1987); and holist-analytic, verbal-imagery dimensions (Riding & Rayner, 1998). One thing that is definitely problematic in connection with learning styles is that with the numerous style dimensions identified, it is still not clear which ones are of the greatest importance and relevance for learning languages. Despite some attempts (e.g. the Ehrman&Leaver construct, Ehrman & Leaver, 2003), an empirically substantiated hierarchy of styles is missing which could provide guidance as to which dimensions should be investigated. Moreover, because of definitional and measurement problems, there are often overlaps between the different constructs (e.g. cf. visual and auditive proposed by Reid, 1987, and verbal-imagery hypothesised by Riding and Rayner, 1998), and what is subsumed under learning and cognitive styles often overlaps with other cognitive or even affective constructs. According to Dörnyei (2005), a rigorous validation of the intuitively appealing constructs is needed.

The issue of learning strategies is similarly complex. In contrast with learning styles that are stable and habitual, according to Cohen (1998) strategies are "learning processes which are consciously selected by the

learner" (p. 4). The major problem in learning strategy research concerns the distinction between strategic learning and learning per se. A possible solution is to emphasise that learning strategies are particularly appropriate for the individual learner as opposed to non-strategic learning (Riding & Rayner, 1998), and that when engaged in strategic learning, learners exert purposeful effort to select procedures that enhance their effectiveness (Dörnyei, 2005).

Similarly to learning styles, empirical research has resulted in a vast number of strategies identified. These are typically arranged into taxonomies, where the main categories include: cognitive, memory, metacognitive, compensation, affective and social strategies (Oxford, 1990), and cognitive, metacognitive, and social/affective strategies (O'Malley, & Chamot, 1990). These taxonomies, unfortunately, cannot really address the issue of usefulness of the particular strategy, as the appropriateness of any given strategy is predominantly determined by the context in which the strategies are used. Dörnyei (2005) sees the solution to the problems listed in the introduction of the concept of self-regulation, which is the individual's capacity for orchestrating strategy use.

### *2.2.1 An influential ID variable - Foreign language aptitude*

Out of the ID variables that are traditionally considered important, only one was examined in the empirical research conducted for the purpose of the dissertation and that is 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. 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 to be 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 depends to a great extent on the theory of language aptitude proposed by different authors. The lack of a generally accepted, theoretically motivated, and empirically testable definition, which is a fundamental problem of the area of language aptitude research, is well reflected by the fact that since the birth of commercial aptitude batteries "language aptitude is what language aptitude tests measure" (Dörnyei, 2005, p. 35.).

Foreign language aptitude is one of the most influential and most extensively researched ID 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.

Despite the fact that the first studies on language aptitude were conducted as early as in the 1920s, 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). Carroll and Sapon took a purely empirical approach to test design. After they 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 foreign language material. The third

factor of language aptitude is *grammatical sensitivity*, which is "the ability to recognise the grammatical functions of words (or other linguistic entities) in sentence structures" (Carroll, 1981, p. 105); whereas the last one, *inductive language learning ability*, entails "the ability to infer or induce the rules governing a set of language materials, given samples of materials that permit such inferences" (Carroll, 1981, p. 105). The latter 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. The definitions given by Carroll 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.

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, HUNLAT, (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 HUNLAT 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: 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 the 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, but 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 relevant and irrelevant information, while *accidental encoding* refers to encoding background or secondary information which can aid comprehension and production at a later stage. *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 of 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.

The theory underlying the CANAL-F test is Sternberg's (1985b, 1997, 2002) triarchic theory of human intelligence. The three aspects of intelligence

described in the theory are the analytical, creative and practical components. These are used for different purposes and are needed for success in everyday life; thus the theory is sometimes called the theory of successful intelligence. *Analytical or componential intelligence* reflects how individuals relate to their internal world, and it is concerned with processing and analysing information. It is subdivided into metacomponents such as planning, monitoring and evaluation; performance components such as execution of plans and strategies developed by the metacomponents; and knowledge acquisition components like selective encoding, selective comparison and selective combination. The *creative or experiential component* reflects how an individual connects the internal world to external reality, and it is concerned with how individuals approach new and unfamiliar tasks. This dimension is broken down to further two categories: novelty, which shows how the person deals with novel demands; and automatization, the ability to automatize information processing. *Practical or contextual intelligence* shows how the individual relates to the external world, and how they adapt to, shape, or if these are impossible, leave their environment.

Besides the fact that this theory is directly relevant for the construct of creativity to be discussed in section 2.2.2, it is also significant for another reason. It signals a trend that is also observable in other recent conceptualisations of language aptitude, namely, that knowledge gained about the cognitive processes that play a role in language learning should somehow be incorporated into theories of language aptitude. Skehan (1998) for example argues that components of language aptitude should be related to different phases of second language acquisition, and believes that besides the traditional components of phonetic coding ability, grammatical sensitivity, inductive language learning ability and rote learning ability new ones are needed. He hypothesises that for example attentional control, and working memory probably play a role in early stages of SLA when the primary processes of acquisition are input processing and noticing, while for example automatization, integrative memory, chunking and retrieval memory are relevant for pattern restructuring and manipulation, pattern control and pattern integration, that are needed at later stages.

A similar emphasis on cognitive processes can be observed in Robinson's (2001a, 2005a) theory of aptitude complexes. In his interpretation aptitude complexes are hierarchical in nature, and they are a result of primary abilities (e.g.: pattern recognition, processing speed, grammatical sensitivity) combining to form second order abilities (e.g.: noticing the gap, memory for contingent speech, deep semantic processing) which can be grouped into aptitude complexes. In the resulting aptitude complexes the relevant second order abilities can be characterised by either high or low levels, resulting in various patterns and leading to different consequences depending on the circumstances, for example the cognitive demands of tasks. Although in formulating his theory Robinson's main concern was to examine correspondences between the cognitive demands of tasks and the aptitude complexes people possess so that by making informed pedagogical decisions instruction could be enhanced, Robinson's theory also highlights the fact that the traditional view of language aptitude is probably too limited.

### *2.2.2 A potentially important ID variable - Creativity*

Having looked at language aptitude which is undoubtedly a core ID variable in the field of applied linguistics, next I am going to examine creativity, which has a long research tradition in psychology. Therefore, in the first part of this section of the dissertation, a chronological approach is taken to present different approaches to creativity, which is then followed by a description of how creativity influences some basic cognitive processes. The review of literature on creativity is concluded by a short overview of issues relevant for measuring creativity.

#### 2.2.2.1 Early theories of creativity

The origins of the study of creativity can be traced back to antiquity. Creativity as it manifests itself in imagination was already described by Plato although the explanation of the phenomenon was restricted to some vague concept of "inspiration" (Kürti, 1985). Creators' introspective reports also suggested supernatural forces as the source of their creativity (Ghiselin, 1952).

This approach, which considers creativity as a highly individualistic and unpredictable process that is qualitatively different from and is not related to the other cognitive processes, is known as the romantic approach of creativity (Ward, 1994), and it is probably still shared by some of our contemporaries. The age of enlightment brought the dominance of "scientific" explanations, which has led to the formulation of numerous theories of creativity. Galton's book "Hereditary Genius" published towards the end of the last century was the first attempt to account for the individual differences in people's abilities (Barkóczi & Zétényi, 1981). Nevertheless, during the first half of the twentieth century, probably due to its great potentials for practical application in institutionalised education and the army, it was the research of intelligence that flourished, and it was not until 1950 that the study of individual differences took up the issue of creativity.

The fact that experimental psychology did not put forward a comprehensive theory of creativity until 1950 does not mean that there were no theories at all. Almost all the different schools and approaches of personality psychology interpreted this phenomenon in their own way, the first among these being psychoanalysis and Freud (1908/1959). The central idea in Freud`s theory is the concept of "sublimation". This term describes a process during which the sexual energy, libido, abandons its originally sexual objective and becomes directed towards socially superior, non-sexual goals. Through the unconscious replacement of sexual objectives for non-sexual ones, the energy is canalised to serve socially desirable goals. Later, Kris (1952) and Kubie (1958) working within the psycho-dynamic approach proposed similar processes and termed them as "regression in service of the ego", and "regression in service of the preconscious" respectively. In this approach creativity is seen as a rare way of tension reduction, and a means to escape neurosis.

In other schools of personality psychology creativity is not an alternative of neurosis, but rather the greatest fulfilment of human potentials. Different authors within the humanistic approach use different terms to describe it: Rogers (1954) calls it a "fully-functioning person", Maslow (1968) refers to it as "self-actualisation" and "peak experience", while

Csikszentmihalyi (1988) uses the term "flow". What they all refer to is that creativity can only be achieved through the realisation of one's own potentials, instead of living up to the expectations and constraints imposed upon us by others.

### 2.2.2.2 Psychometric theories of creativity

Guilford (1950), who believed that creativity is a stable set of traits, normally distributed in the population, was among the first to put forward a list of cognitive processes involved in creativity. He believed that these processes include sensitivity to problems, synthesising ability, analysing ability, reorganisation or redefinition of organised wholes, evaluation, a high degree of complexity of the conceptual structure, creative fluency of production, ability to come up with novel ideas, and flexibility of mind. The latter three of these are seen as crucial aspects of creativity even today, and numerous creativity tests were designed to measure these abilities (Barkóczi & Zétényi, 1981; Guilford, 1967; Torrance, 1966).

Later Guilford (1959) outlined an ambitious model called the "structure of intellect", the aim of which was to account for every aspect of human cognitive abilities. This was a hypothetical model derived from the statistical process of factor-analysis, and it was to be verified later through empirical studies. The three major dimensions in his model were: operations, informational contents and products, which all contain different numbers of elements (five, four and six respectively). The sum of all the combinations of these dimensions gives a total of 120 basic skills, each of them being a unique combination of different operations, contents and products.

Divergent production is only one of the five different operations, or in other words intellectual processes, but as Guilford (1959) believed this to be the cognitive background of creativity, this was the phenomenon that he explored most extensively. He proposed that *divergent thinking*, that is the ability to produce many different ideas in response to a problem, is 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 production, the "generation of variety and amount of information, based on given information; most involved in creative potential" (Brown, 1989, p. 15) if combined with the four different contents and six products, contains 24 theoretically independent factors. If we were to follow his theory strictly, we should measure all of these independently, which would certainly be unfeasible. Thus, Guilford himself and his followers working within the psychometric tradition restrict themselves to measuring four independent facets of divergent thinking, which supposedly cover the twenty-four elementary skills. These are 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).

Mednick (1962) did not consider creativity as a result of divergent thinking, but as originating from the large numbers of associations between representations. He defined the creative thinking process as "the forming of associative elements into new combinations which either meet specified requirements or are in some way useful. The more mutually remote the elements of the new combination, the more creative the process of solution" (Mednick, 1962, p. 221). In addition to the number of associations, Mednick introduced another individual variable "associative hierarchy", which also affects a person’s creativity. Associative hierarchy is the unique way of the organisation of a person’s associations; it contains the associative links, their number and their relative strength as well. Uncreative individuals can usually make a small number of associations with very high probability, that is, they have a relatively small number of very strong associations, while highly creative individuals can make large numbers of associations with nearly equal probability. Their associations might not be as strong, but their equal probability increases the chance of producing associations between remote elements, which is creativity itself. It has to be noted that this theory has very similar implications for the testing of creativity as Guilford's (1959) model. It predicts that creative individuals will produce a large number of unusual responses on a task measuring creativity.

### 2.2.2.3 Current theories of creativity

Current models of creativity tend to be more complex as the area is dominated by multi-componential approaches. Besides cognitive components, these models heavily rely on personality and motivational variables as well. These multi-componential theories hypothesize that there are several prerequisites of creativity, that is, multiple components must converge for creativity to emerge. Sternberg and Lubart's (1991, 1996) investment theory is one example of multi-componential approaches. The model is called investment theory because it predicts that creative people are the ones who are able "to buy low and sell high in the realm of ideas" (Sternberg and Lubart, 1996, p. 683). This means that they pursue ideas that are unknown or unpopular but have growth potential, and, having developed these ideas further, they are eventually able to make profit by persuading others of the value of their ideas. According to this theory, creativity requires six distinct but interrelated resources: intellectual abilities, knowledge, styles of thinking, personality, motivation, and environment. Three intellectual abilities are judged essential with respect to creativity: *synthetic ability*, which makes it possible for the individual to see problems in new ways and escape the bounds of conventional thinking; *analytic ability*, which is useful for recognising those ideas that are worth pursuing; and the *practical-contextual ability*, which helps in persuading others of the usefulness of the individuals' ideas.

Another new feature of current theories of creativity is that they raise the issue of content-specificity. As opposed to the earlier models where the assumed position was content-generality, some recent theories of creativity argue for content-specificity, that is, they state that creative activity within one content area is independent of creativity in other content areas. This line of argumentation seems to be supported by findings that people capable of producing truly creative products, usually perform this in one certain area only (e.g. genius painters are usually not genius mathematicians as well), and that in many cases a large base of knowledge is needed on which creativity can operate. Amabile's (1983, 1996) componential model can serve as an example of the domain-specific approach, since besides the groups of creativity-relevant

skills and task motivation, she introduces a third group of variables called domain-relevant skills. These include: knowledge about the domain, technical skills required, and special, domain-relevant talent. Although the domain-specificity of creativity seems to be plausible if we consider it from the perspective of the creative product, and it is also supported by empirical evidence (Csikszentmihalyi, 1988; Gardner, 1993; Runco, 1989; Sternberg & Lubart, 1995), we can still hypothesise that it is irrelevant regarding the underlying cognitive process. It is possible that although creative individuals are characterised by a special way of cognitive functioning which is domain independent, the reason why this does not lead to creative products in all areas is that it is precisely the level of production where domain-relevant skills have a major role and not the level of underlying cognitive processes. In other words without domain-specific specific knowledge, skills, and talent, the underlying cognitive processes that play a role in creativity are unable to generate truly creative products.

Since current models of creativity tend to be multi-componential, they overlap with research on motivation, personality, and knowledge, making both the definition of the construct and conducting research quite difficult. For this reason, I intend to discuss only the cognitive components of creativity in more detail, as besides being the most extensively tested (Cropley, 1972; Guilford, 1967; Harrington, Block, & Block, 1983; Jordan, 1975; Kogan, & Pankove, 1974; Torrance, 1962), I believe they provide the basis on which creativity might be manifested if other components are also present. In other words, the intellectual abilities are essential components of creativity, without them creativity would not exist.

### 2.2.2.4 Cognitive components of creativity

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 redefinition, choosing a useful problem presentation, selecting an appropriate problem-solving strategy, and evaluating the generated possibilities effectively. It is interesting to note that some of these processes are hypothesised to be related to foreign language aptitude within the Cognitive Ability for Novelty in Acquisition of Language - Foreign (CANAL-F) theory, a framework of language aptitude recently developed by Grigorenko et al. (2000) which was briefly discussed in section 2.2.1.

Factors of creativity-relevant intellectual abilities tend to load on a common higher-order factor according to Carroll's (1993) factor-analytic research, which provides empirical evidence of the autonomous existence of this ability. Having reviewed and reanalysed 121 datasets, Carroll found nine first-order factors relevant for *idea production*, which he believes is a basic human characteristic and an equivalent of creativity. These first-order factors include ideational fluency, naming facility, associational fluency, expressional fluency, word fluency, sensitivity to problems, originality/creativity, figural fluency, and figural flexibility. 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 it 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.

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). 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 Carroll's three-stratum theory is founded, still the domain of general retrieval ability "is chiefly (but not entirely)

concerned with Guilford's divergent production operation" (Carroll, 1993, p. 638). Therefore, divergent production seems to be a valid construct even in the light of current research; thus, it seems justifiable to posit it as one of the cognitive underpinnings of creativity. It should be noted here that creativity also appears in a different theory of human intellect, discussed briefly in section 2.2.1, in Sternberg's (1985b, 1997, 2002) triarchic theory of human intelligence. In this theory, one of the three dimensions of intelligence is labelled the creative or experiential component, which is concerned with how individuals approach new and unfamiliar tasks, and which is also believed to be relevant for language aptitude.

#### 2.2.2.5 Creativity and perception

Besides those intellectual processes that are specifically regarded as relevant to creativity, such as divergent thinking, it seems that creativity can be manifested in even the most basic cognitive processes. One such cognitive process that might be influenced by creativity is perception (Flowers & Garbing, 1989). Although perception as an information reduction process is generally considered to be incompatible with creativity, it has some characteristics that can be shown to have some relevance for it. There seem to be two distinct categories of perceptual processes which might be related to creativity, one of them is the involuntary processes of perceptual organisation and the other one is the processes of executive control exercised by the individual.

The first category, the *involuntary processes of perceptual organisation* are normally aimed at information reduction, and through this they promote stability and act against the formation of novel representations of information. These involuntary processes can lead to creativity and novel ways of representation if they operate in a "loose", somewhat less deterministic but at the same time less effective way; this is what is believed to happen in schizophrenia. The second category of perceptual processes that might lead to creativity are the *processes of executive control*, which include spatial selective attention, manipulation of mental images, and controlled cross-modal representation. The production of novel representations in their case is brought

about by the effortful construction and modification of mental representations. This is the normal course of operation of the processes of executive control, whereas in the case of involuntary processes of perceptual organisation creativity resulted from the inadequate functioning of these processes. It can be easily understood that the two processes described above - one implying inadequacy, the other superb operation - lead to very different manifestations of creativity. A distinction that might account for the fact that although the mentally sick and healthy can both produce creative products, their creativity can be very different (Flowers & Garbing, 1989).

Flowers and Garbing (1989) describe a third way in which perceptual processes might contribute to creativity, and that is "sudden insight, that involves processes not under executive control nor driven by sensory data, but that produces seemingly spontaneous mental representations, often involving visual imagery" (p. 150). It might be easier to understand such insight if we know that experimental results suggest that highly creative individuals have higher susceptibility to near threshold stimuli, not detected by them consciously (Barkóczi, 1991). This means that they can make use of "unconscious" impulses and near threshold stimuli that average people cannot, which might be the key to their sudden insight.

The phenomenon which might be in the background of the perceptual peculiarities described above, thus might provide an explanation for them, is *latent inhibition* (LI). LI was already described by Pavlov, and it is brought about by the repetition of a stimulus without reinforcement, which otherwise would result in habituation. It means that the orientation reaction is no longer triggered by the stimulus, as the stimulus was learnt to be considered irrelevant. LI, therefore, is the way in which we get rid of irrelevant stimuli, and this is what schizophrenics and highly creative people are incapable of. In studies involving university students, Peterson, Smith and Carson (2002) and Carson, Peterson, and Higgins (2003) found that participants characterised by low LI were significantly more open and scored higher on Gough's (1979) Creative Personality Scale. In their interpretation, low levels of LI result in the individual being flooded by ideas if they do not have the cognitive resources to edit and constrain these; therefore, low LI typically leads to psychosis. If, however, the cognitive resources of controlling the information flow are

available for the individual as in the case of participants with high IQ, low LI tends to lead to creativity.

### 2.2.2.6 Creativity and memory

Memorization is a cognitive process that plays a role during the central processing stage of information processing. Remembering, the retaining of past ideas, similarly to perception, is seen at first sight as a process quite different from creativity. Studies revealed, however, that "remembering is not simply the reinstatement of previously experienced events but rather involves an imaginative reconstruction of the past" (Stein, 1989, p. 163). This means that some creativity is involved in the process of remembering, but the opposite might prove to be true as well, creative behaviour probably also involves elements of memory.

One area where memory is expected to have beneficial effects on creativity is that of transfer, where previous experience, knowledge and skills facilitate creativity. Numerous studies were constructed to investigate the conditions under which we can make use of our past knowledge in a creative way. The results of these studies showed that in spontaneous transfer experiments, where people are not given hints to use the specific information given to them previously, it was the similarity between the information represented in memory and the problem solving task that increased the number of creative solutions (Stein, 1989). Although this is true for every individual, in experiments where the subjects' creativity was also introduced as an independent variable, it became obvious that in tasks requiring spontaneous transfer, creative individuals can make better use of the information previously provided to them (Barkóczi, 1994a). This difference probably implies superior memory functions, that is, better encoding, storage and retrieval. The fact, that highly creative individuals demonstrate hypermnesia, in other words their performance does not decrease, but increases during the course of successive memory tests (Barkóczi, 1994b) also supports the proposed superior memory functions in case of creative individuals.

### 2.2.2.7 Metacognition in creativity

Modern cognitive psychology recognises that psychological processes involved in cognition are arranged hierarchically. At the top of the hierarchy are "the executive processes that oversee, regulate, and orchestrate the activities of cognition. These executive processes are known as metacognition" (Armbruster, 1989, p. 177). This section will examine the metacognitive processes that are believed to play a role in creativity with the help of a four stage model of the creative process proposed by Wallas and interpreted by Armbruster (1989). Armbruster (1989) describes "goal setting" as the initial stage of the creative process, although this phase is not present in Wallas' model. The metacognitive process involved at this stage is awareness; the individual becomes aware of a goal or purpose, which will further on be a driving force behind the whole creative process.

The first stage of the model proposed by Wallas is "preparation", which involves the acquisition of the necessary knowledge and skills of the field in which the creative output is to be produced. In Armbruster's (1989) interpretation, the organisation of information into flexible schemas, that is, the development of flexible knowledge representations is also very important during this phase, as these schemas can be restructured and can lead to unique recombinations later on. The formulation of rich, interconnected, and flexible cognitive structures can be enhanced by the multiple encoding of information, using different modes or styles of thought, for example. The function of metacognitive processes is twofold here. On the one hand, with their help the individuals can try to consciously enhance the formulation of flexible knowledge structures through using multiple coding deliberately. On the other hand, awareness of the current state of their knowledge structure is probably beneficial for the individuals as well.

During the second stage called "incubation", the problem is no longer consciously pursued, and the working of unconscious or partially conscious processes can be observed (Armbruster, 1989). What probably takes place is the restructuring of information into new schemas by the inherent organisational processes of the brain. The metacognitive skill possibly involved here is the "mastery of control" over the reworking of the flexible cognitive

representations, in other words, the efficient, though unconscious, control of the restructuring of schemata.

The third stage, "illumination", also called the "Eureka!" or "Aha!" experience, is the moment when the unconscious suddenly becomes fully conscious. Armbruster (1989) describes this process as the recognition of a coherent cognitive representation and supposes that creative individuals might have superior metacognitive awareness to recognise good insight. As the illumination of a fully formed work is quite unusual, the creative process usually involves a further stage, "verification".

During the course of this fourth phase, verification, correction and revision of the product takes place. Armbruster (1989) sees a very important role of metacognition here: "creative individuals seem to be especially adept at the conscious metacognitive skills that are required during the verification stage. They may be unusually sensitive to both internal and external standards and particularly able to revise the creative product accordingly" (p. 180). They may also be exceptionally good at improving these abilities with experience and practice. To sum it up, after considering each of the metacognitive processes that may play a role in creativity, we can draw the conclusion that creative individuals are probably characterised by a higher level of awareness and more efficient control of the functioning of their cognitive processes.

### 2.2.2.8 Measuring creativity

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 was done by Sternberg and Lubart (1991), who tried to establish individual creativity in 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 most characteristic of the creative process (Guilford, 1967; Torrance, 1962).

Guilford (1959), when he first produced his test of creativity, attempted to compile a tool that measured aspects of intellect not covered by traditional intelligence tests. The relationship between intelligence and creativity has been an issue ever since, and although the two are considered to be independent by many researchers, there is one phenomenon that needs to be explained. Although above a certain IQ score there is clearly no relationship between intelligence and exceptional talent that is creativity, people with low IQ will probably score low on creativity tests as well. Thus, it is very unlikely that someone who has performed poorly on an intelligence test will perform well on a creativity test. Therefore, the connection between intelligence and creativity seems to exist only in the low intelligence band of the population. Probably the best-known model that attempts to account for this phenomenon is the threshold theory (Hayes, 1989). According to this theory, a person's IQ must be above a certain threshold value if that person is to be successful in creative activities, and IQ differences above that level make no difference in creativity.

Divergent production, the ability that Guilford (1959) hypothesised to be the cognitive background of creativity, is made up of twenty-four elementary abilities in his structure of intellect model (described in section 2.2.2.2). Guilford himself was not prepared to measure each of these elementary abilities independently; thus, in his test of creativity he measured four higher level abilities: fluency, flexibility, originality and elaboration, each with the help of a separate task. This tradition, however, was not followed by other constructors of creativity tests. Torrance (cited in Oláh 1987), for example, whose tests of creativity have been used most extensively, measures all of the above abilities on the same task. In his tests all the tasks are scored for fluency, flexibility, originality and elaboration, as well (Zétényi, 1989).

The test-retest reliability of the Torrance and Guilford tests is between 0,3 and 0,93 which is probably due to the fact that performance on creativity tests is greatly influenced by motivational factors (Zétényi, 1989). One consistent finding is that the subjects' performance improves over time. 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).

The standardised creativity test used in Hungary was first developed by Barkóczi and Klein in 1968. It consists of four parts, two figural and two verbal sub-tests. The two figural test, Circles and Picture Completion, are almost identical to Torrance's Test of Creative Thinking (TCT) (cited in Oláh, 1987), while one of the verbal tests, Unusual Uses is an adaptation of a part of Guilford's test (cited in Oláh, 1987). The idea of the other verbal test Remote Associations originates from Mednick, but it was further developed by Barkóczi (cited in Oláh, 1987). The scoring of the different tasks seems to follow Torrance's tradition, each task is scored for fluency, flexibility and originality, elaboration, however, is not measured. The test has been standardised for Hungarian adults, but it still has the same reliability and validity problems as its foreign equivalents (Zétényi, 1989).

### 2.2.2.9 Conclusion

Having reviewed a selection of the most important theories of creativity chronologically, it might be easier to understand why creativity is a neglected individual variable in SLA research, and why some authors feel that it is neglected within mainstream psychology as well (Sternberg, & Lubart, 1999). First of all, although the concept of creativity is taken up and discussed by a wide range of schools and approaches within psychology, the construct of creativity is defined very differently by them, making the different constructs virtually impossible to compare. The difficulties with the measurement of the construct described above might also be partly held accountable for the fact that SLA research of individual learner variables has failed to investigate the

effects of creativity. It is probably also obvious from the presented theories that they tend to lay different emphasis on the affective and cognitive components of creativity: the psychoanalytic (Freud, 1908/1959; Kris, 1952; Kubie, 1958) and humanistic approaches (Rogers, 1954; Maslow, 1968; Csikszentmihalyi, 1988) tend to emphasise the affective components, Guilford's (1959) and Mednick's (1962) theories solely rely on cognitive ones, while today a more balanced approach is taken by the multi-componential theories (Amabile, 1983, 1996; Sternberg, & Lubart, 1991, 1996).

It seems justifiable to argue, however, that certain cognitive processes characterise creativity distinctively, one of them being divergent thinking. Despite being a relatively old construct (Guilford, 1959), divergent thinking is still regarded as part of the basic-level creativity-relevant abilities (Lubart, 1994), and it is largely compatible with Carroll's (1993) general retrieval ability. Moreover, the standardized test of creativity (Barkóczi, & Zétényi, 1981) available for the Hungarian population is a divergent thinking test, and it is still in use for measuring creativity (Gáspár, 2001, Kárpáti, 1996, Tóth, 2006). Consequently, taking feasibility issues into consideration as well, assessing the divergent thinking aspect of creativity appears to be the best choice in studies aimed at investigating the possible relationships of this new ID variable.

## 2.3 The role of tasks

Having reviewed the literature on ID variables, we will now turn our attention to the other construct examined in this research study, oral narrative tasks. The detailed discussion of tasks: definitional problems, different task types and task characteristics, however, is preceded by a brief overview of the changes that have taken place in language instruction, as these might shed light on the reasons of the recent prominence of tasks.

### *2.3.1 Important changes in language instruction: Communicative and task-based language learning*

A relatively new framework of interpretation of how language learning occurs, thus how language teaching should be constructed, is task based language teaching (TBLT). TBLT, being a descendant of communicative language teaching (CLT) whose main objective is "to develop the learner's ability to take part in spontaneous and meaningful communication in different contexts, with different people, on different topics, for different purposes" (Celce-Murcia, Dörnyei & Thurrell, 1997, p. 149), has clearly retained meaning as its main, though not exclusive, focus. Recent findings from psycholinguistics and cognitive psychology seem to suggest that learners are naturally predisposed to attend to meaning in the course of communication, and they can successfully extract meaning with the help of comprehension strategies (Clark & Clark, 1977) and convey meaning using communication strategies (Kellerman, 1991) without necessarily attending to the form of discourse. As awareness of the discrepancy between the target language and the learner's interlanguage, which is the driving force behind interlanguage development, is not possible without attending to form, the over-effective use of these strategies might carry the danger of fossilisation. For this reason researchers and practitioners working within the framework of TBLT support a focus on form, that is, they believe that besides attending to meaning, students should also be made aware of the specific features of the linguistic code. They emphasise, however, that this must not mean a return to a focus on forms, the explicit teaching of grammar (Long & Crookes, 1993).

Different authors suggest different ways for directing learners attention to form. Long (1989) hypothesises that engaging in tasks is enough in itself to trigger acquisitional processes, since the communication breakdowns emerging during task completion necessitate the negotiation of meaning between the participants. Negotiation, which usually occurs through different means such as comprehension checks and clarification requests, provides the learners with the opportunity to hear language which may later be incorporated into their interlanguage systems on the one hand, and on the other, offers them the possibility to express concepts which are beyond their linguistic capacity

(Plough & Gass, 1993). Communication breakdowns also inform learners that something went wrong during the course of interaction, thus changes need to be made. In this way they are forced to notice gaps in their knowledge, which is a prerequisite of the eventual restructuring of grammar (Gass, 1988). Skehan and Foster (1999), however, find this line of argumentation problematic, as they believe that previous research in the field provides no conclusive evidence that negotiation of meaning has an actual impact on interlanguage, in other words, that it actually brings about the changes described above. Besides, some findings suggest that negotiation of meaning does not occur as frequently as it has been argued (Foster, 1998) and that in some cases it may actually irritate language learners (Aston, 1986) and thus hinder development.

Skehan (1998) takes a different approach to the issue of directing learners' attention to form. He emphasises the information processing demands task performance places on learners and argues that because of their limited processing capacities learners cannot devote attention to every aspect of the task at the same time (Van Patten, 1990). He distinguishes three areas of task performance that can be in the focus of the learners' attention: *fluency*, where the learners' priority is meaning, and *accuracy* and *complexity*, where the learners' priority is form. Two different aspects of form are represented by the latter two performance areas: accuracy reflects the control of form, that is the correct use of acquired structures, while complexity reflects restructuring within the underlying interlanguage resulting in performance which is not necessarily correct. Skehan (1998) believes that concentrating on negotiation of meaning alone carries the danger of prioritising one area, fluency, which is the learners' capacity to mobilise linguistic resources for real-time communication, at the expense of the other two. He suggests that allocating attention to each of these three areas is necessary for balanced language development, an objective that can be achieved through the careful selection and implementation of tasks. This explanation appears plausible in the light of the current findings of cognitive psychology, providing evidence for the limited attentional and processing capacity of humans (Czigler, 1992).

### *2.3.2 Task definitions in TBLT*

The assumption that transacting tasks involves the activation of naturalistic acquisition mechanisms (Skehan, 1998), therefore, tasks should be the basic unit of instruction led to the birth of TBLT and inevitably brought along an interest in tasks among researchers. It seemed essential to learn more about the nature of tasks before they could be confidently used in task-based syllabuses. This led to a mushrooming of research on those aspects of tasks that influence the kind of output people produce on them, and a heightened interest in the definition of task. Different authors working within this framework have adopted numerous definitions, some very broad as well as more restricted ones. Long's (1985) definition is probably the most general, he defines task as:

> A piece of work undertaken for oneself or for others, freely or for some reward. Thus, examples of task include painting a fence, dressing a child, filling out a form, buying a pair of shoes, making an airline reservation, borrowing a library book, taking a driving test, typing a letter, weighing a patient, sorting letters, taking a hotel reservation, writing a check, finding a street destination and helping someone across the road. In other words, by 'task' is meant the hundred and one things people *do* in everyday life, at work, at play, and in between. Tasks are the things people will tell you to do if you ask them and they are not applied linguists. (Long, 1985, p.89)

Although Long's (1985) intention seems to be to emphasise the real-life element in tasks, a quality that he believes should also be adopted within TBLT, his definition is too broad for research purposes. Others have opted for more restricted definitions that describe tasks from a pedagogical point of view and place an emphasis on the special characteristics of language learning. Nunan (1989) points out that tasks are meaning rather than form focused, while Candlin (1987) states that tasks should involve the pursuance of some goal. A comprehensive definition of task is offered by Skehan (1996), who describes tasks as:

> an activity in which: meaning is primary; there is some relationship to the real world; task completion has some priority; and the assessment of task performance is in terms of task outcome. (p. 38)

This definition implies the importance of the real-life quality of tasks emphasised by Long, the meaning-focus stressed by Nunan, and the goal-orientedness described by Candlin, while making reference to the evaluation of task performance as well. Although different authors working within the framework of the TBLT approach undoubtedly contributed to the emergence of a rich and multi-faceted concept of task, the large number of definitions and their slightly differing interpretations pose serious problems as well, since it is not always clear whether the same term used by different authors refers to the same concept.

### *2.3.3 Tasks in language testing*

Examination of the testing literature reveals that the term task is used in a much broader sense there compared to most authors of the task-based literature. In their book "Language testing in practice" Bachman and Palmer (1996) adopt a definition given by Carroll (1993) which is much broader than the definitions typically used in the task-based literature "a task [is] any activity in which the person engages, given an appropriate setting, in order to achieve a specifiable class of objectives" (Bachman & Palmer, 1996, p. 43). The defining characteristics of a task in their opinion involve individuals using language for the purpose of achieving a particular goal in a certain situation, but the emphasis on the real-life quality and meaning-focus of tasks is clearly missing. Given this definition a task could be almost anything ranging from a cloze-test to a discussion task.

In order for a task to be used for testing purposes in a valid way, its properties must be very carefully described and matched to those circumstances that we want to draw conclusions for from our test. The authors argue that by describing target language use (TLU) tasks, that is those situations where the candidate to be tested is likely to use the language, we can attempt to design tests or test tasks with features that correspond to a large extent to those of the target language. Thus Bachman and Palmer (1996) propose a sophisticated framework for analysing task characteristics addressing the characteristics of the setting, the test rubric, the input, the expected

response and the relationship between input and response in detail. They do not discuss, however, the way each of these features affect task performance.

The question is, however, whether it is possible at all to match the features of TLU tasks and test tasks so closely so that differences between them caused by their discrepant characteristics would really be negligible. In addition, if our purpose for using tasks is other than testing proficiency, for example if we want to use tasks as an elicitation device, the effects of the above mentioned task features are of prime interest for us, as they may decisively determine the quality of the output that the subjects produce. This is the conclusion that Skehan (1998) arrives at as well saying:

> the nature of performance on a task is not something which is available at the tester's convenience, with one task being pretty much the same as another. Tasks themselves influence the nature of the performance which results, and so can have an impact upon someone's judged proficiency. (p. 175)

This being the case, it might be worth taking a closer look at these characteristics and their proposed effects as described within the framework of the task-based approach. Before doing this, problems encountered in the identification of different task types are discussed.

### *2.3.4 Task types*

The ambiguity apparent in the definition of task can also be detected when considering the way tasks appear in practice in the classroom: in the form of different task types. Based on a review of relevant literature, it may be concluded that there seems to be no consensus concerning the exact number and the precise nature of existing task types. While traditional task type labels originating from the classroom, such as "information gap", "jigsaw", "opinion exchange", "decision making" or "problem solving", also frequently occur in articles, the reader may easily get confused by phrases such as "one-way" and "two-way" tasks (Long, 1985), "convergent" and "divergent" tasks (Duff, 1986), or "static", "dynamic" and "abstract" tasks (Brown, Anderson, Shilcock & Yule, 1984). These latter task classifications are all based on certain

distinctive task features the authors judge to be of crucial importance with regard to the pedagogical usefulness or the difficulty of tasks, these are: the direction of information flow between those interacting, the number of solutions that can be arrived at, and the information type the task contains, respectively. Here the terminological diversity results from the fact that different researchers emphasise different aspects of tasks. The issue is complicated further by the varying levels of analysis employed: the traditional task types, "jigsaw" and "information gap", can also be classified as "convergent" tasks; and the different levels of specificity involved: the "Spot the difference" (Plough & Gass, 1993), "Draw the picture" (Gass & Varonis, 1985) and "Assemble the scene" (Pica, Young & Doughty, 1987) tasks are all information gap tasks but have different content.

Long and Crookes (1993) address this issue as the "problem of finiteness" (p. 42) in their article and raise questions about the number of truly different task types and the levels of analysis involved in defining them. The lack of consensus within this area is identified as a shortcoming of the TBLT approach. Long and Crookes suggest that further research is necessary since it is obvious that without a generally accepted unified framework of significant task characteristics this problem can hardly be resolved. Although some attempts at compiling a comprehensive framework have already been made (e.g.: Pica, Kanagy, & Falodun, 1993; Robinson, 2001b, 2005b, 2007a; Skehan, 1998), the existing frameworks are far from being generally accepted probably partly because they originate from different theoretical orientations. The adoption of a unified framework would still be desirable, as it would probably eliminate the use of those task type labels that are based on only one distinctive characteristic, and which because of the frequent overlaps described above can be quite confusing. Also, by providing assistance in a more sophisticated analysis of traditional task types, the tasks could at least be made comparable; thus, a unified framework might render the quest for the exact number of task types unimportant. In line with this, the current trend within TBLT seems to be researching general task characteristics instead of investigating the effect of task type.

### *2.3.5 Theories of task characteristics*

The quest for a unified framework of task characteristics should start with identifying those features of tasks that seem to have universal relevance. Having reviewed the literature, there are a number of one-dimensional approaches, for example the ones focusing on characteristics such as the way of information flow (Long, 1985), or the number of solutions agreed upon (Duff, 1986) that seem to be too narrow for our purposes. In fact, these authors do not aim to describe universal features of tasks, they are only concerned with finding tasks that are more effective for pedagogical purposes. They try to pinpoint those properties of tasks that enable them to elicit superior performance from the learners. A universally applicable, therefore, more promising trend seems to be the one which attempts to establish dimensions underlying the notion of task difficulty. These multi-componential taxonomies, primarily concerned with the cognitive demands of tasks, seem to be relevant for all kinds of tasks; thus, they are discussed in detail in section 2.3.4.1 below.

#### 2.3.5.1 Characteristics affecting task difficulty

In an early attempt to establish task difficulty, Brown et al. (1984) relied on a number of empirical studies on the basis of which they proposed two dimensions. The first is concerned with (a) the degree of difficulty as manifested in the information type, here they proceed from static, through dynamic to abstract tasks. The other dimension (b) also indicates degrees of difficulty, but this time with regard to the scale of task and the interrelationship between elements. It implies that the greater number of elements a task involves the more difficult it is, and it also suggests that the nature of the relationships between the elements also contribute to task difficulty. In fact both of the dimensions put forward by Brown et al. can be interpreted within an information processing framework, as more abstract information as well as the greater number of elements to be processed are believed to impose a greater processing load. Since their model only addresses cognitive aspects of task difficulty it is somewhat limited in scope.

A more comprehensive, although largely speculative framework is offered by Candlin (1987). This framework, however, does not address possible interrelations between the criteria listed thus can hardly be considered more than a mere checklist. Its components are: (a) *cognitive load*, referring to the general complexity of the task content and the number of elements or participants involved (cf. Brown et al. 1984 scale of task and interrelationship of elements), (b) *communicative stress*, as pressure coming from the interlocutor, (c) *particularity and generalizability*, which concerns the clarity of the goal of the task and the norms of interpretation, (d) *code complexity* and *interpretative density*, the former referring to the linguistic code, the latter to the operations which need to be carried out on the code, and (e) *process continuity*, which derives from the familiarity of the task.

Skehan (1998) uses similar criteria to outline a scheme of task difficulty, however, his scheme is a three-dimensional and multi-layered one. The first dimension is (a) *code complexity*, which includes linguistic complexity and variety, vocabulary load and variety, and redundancy and density (cf. Candlin's (1987) code complexity and interpretative density). The second one, (b) *cognitive complexity* has two facets: (1) cognitive familiarity (cf. Candlin's (1987) process continuity), that is, familiarity with the topic and its predictability, familiarity of the discourse genre, and familiarity of the task, and (2) cognitive processing, that is, information organisation, amount of 'computation', clarity and sufficiency of the information given and information type (cf. Brown et al. (1984) scale of task and interrelationship of elements, and Candlin's (1987) cognitive load). The third dimension (c) *communicative stress* includes such factors as time limit and time pressure, speed of presentation, number of participants, length of texts used, type of response and opportunities to control the interaction (cf. Candlin's (1987) communicative stress). These categories are believed to capture the three major factors that influence tasks: language, thinking, and performance conditions. Although this scheme is more comprehensive than the previous ones, it is also highly theoretical and the interrelationships between the different categories are not defined. Therefore, in spite of the fact that it could provide broad guidelines for the evaluation of tasks, it is probably an inadequate tool in those cases when

the tasks differ along various dimensions thus several variables need to be considered simultaneously.

The Triadic Componential Framework drawn up by Robinson (2001b, 2005b, 2007a) can be considered as an attempt at synthesising prior theories and research findings and at clarifying terms. He aimed to establish "theoretically motivated, empirically substantiable, and pedagogically feasible sequencing criteria" (p. 27) in order to offer guidance regarding sequencing decisions during syllabus design. In his componential framework, three independent facets of tasks: *task complexity*, *task difficulty*, and *task conditions* were distinguished. Task complexity in his interpretation is the result of various information processing demands that the structure of the task imposes on the learners; task difficulty covers learner factors: differences between learners in their cognitive and affective resources that makes certain tasks personally difficult for them, while task conditions include participation and participant factors, and the context of task performance.

Robinson (2001b, 2005b, 2007a) argues that sequencing decisions should be solely based on task complexity, as this is a fixed and invariant feature of the task; consequently, a simple task will be less demanding than a more complex one for any given learner. Task difficulty, on the other hand, explains individual differences between learners, showing why one particular task should be more or less difficult for different learners. As differences between learners in affective variables, such as motivation, and social factors, such as group cohesion, that were shown to contribute to differences in task performance (Dörnyei & Kormos, 2000) are variable and temporal, they should form the basis of on-line methodological decisions according to Robinson. The effects of the more stable cognitive abilities such as intelligence, aptitude or even creativity could be taken into consideration as well if conclusive results were available about the way they affect performance on tasks. Although Robinson believes that the three factors of task complexity, task difficult and task conditions interact with each other and their interactions should be empirically studied, he expects that individual differences contributing to task difficulty play a greater role in complex task performance than they do on

simple tasks (Robinson, 2003). Figure 1 presents the Triadic Componential Framework for task classification.

Figure 1
*The Triadic Componential Framework for task classification - categories, criteria, analytic procedures, and design characteristics*

| ***Task Complexity*** (Cognitive factors) | ***Task Condition*** (Interactive factors) | ***Task Difficulty*** (Learner factors) |
|---|---|---|
| (Classification criteria: cognitive demands) | (Classification criteria: interactional demands) | (Classification criteria: ability requirements) |
| (Classification procedure: information-theoretic analyses) | (Classification procedure: behaviour-descriptive analyses) | (Classification procedure: ability assessment analyses) |
| *(a) Resource-directing variables* making cognitive/conceptual demands | *(a) Participation variables* making interactional demands | *(a) Ability variables* and task-relevant resource differentials |
| +/- here and now | +/- open solution | h/l working memory |
| +/- few elements | +/- one-way flow | h/l reasoning |
| -/+ spatial reasoning | +/- convergent solution | h/l task-switching |
| -/+ causal reasoning | +/- few participants | h/l aptitude |
| -/+ intentional reasoning | +/- few contributions needed | h/l field independence |
| -/+ perspective taking | +/- negotiation not needed | h/l mind/intention reading |

| ***Task Complexity*** (Cognitive factors) | ***Task Condition*** (Interactive factors) | ***Task Difficulty*** (Learner factors) |
|---|---|---|
| (Classification criteria: cognitive demands) | (Classification criteria: interactional demands) | (Classification criteria: ability requirements) |
| (Classification procedure: information-theoretic analyses) | (Classification procedure: behaviour-descriptive analyses) | (Classification procedure: ability assessment analyses) |
| *(b) Resource-dispersing variables* making performative/procedural demands | *(b) Participant variables* making interactant demands | *(b) Affective variables* and task-relevant state-trait differentials |
| +/- planning time | +/- same proficiency | h/l openness to experience |
| +/- single task | +/- same gender | h/l control of emotion |
| +/- task structure | +/- familiar | h/l task motivation |
| +/- few steps | +/- shared content knowledge | h/l processing anxiety |
| +/- independency of steps | +/- equal status and role | h/l willingness to communicate |
| +/- prior knowledge | +/- shared cultural knowledge | h/l self-efficacy |

*Note.* From "Criteria for grading and sequencing pedagogic tasks" by P. Robinson, 2007, In M. P. Garcia Mayo (Ed.) *Investigating tasks in formal language learning* (p. 14). Clevedon: Multilingual Matters.

### 2.3.5.2 Robinson's Cognition Hypothesis

Since it is the Cognition Hypothesis put forward by Robinson (2001c, 2003, 2005b) which elaborates the theory behind the Triadic Componential Framework (Robinson, 2001b, 2005b, 2007a) described above, it also needs to be addressed briefly. Robinson's Cognition Hypothesis, which is a modified version of Cromer's (1974) Cognition Hypothesis of L1 acquisition claims the following:

> increasing the cognitive demands of tasks contributing to their relative complexity along certain dimensions will; (a) push learners to greater accuracy and complexity of L2 production in order to meet the greater functional and conceptual communicative demands they place on the learner; (b) promote interaction, and heightened attention to and memory for input, so increasing learning from the input, and incorporation of forms made salient in the input; as well as (c) longer term retention of input; and that (d) performing simple to complex sequences will also lead to automaticity and efficient scheduling of the components of complex L2 task performance. (Robinson & Gilabert, 2007, p.162)

The words "certain dimensions" (Robinson & Gilabert, 2007, p.162) in the quotation above refer to a very important distinction introduced by Robinson. According to the Cognition Theory (see also the Triadic Componential Framework in Figure 1), a task can be more complex along two different dimensions: a resource directing dimension, that is the task is changed in a way that necessitates the expression of more difficult concepts; and a resource dispersing dimension, that is increased demand is put on the learners' resources without their attention being directed to any particular aspect of the linguistic system. If we make a task more complex along the resource directing dimension, for example by varying whether the task requires reference to events happening now or in the past, or whether it requires simple information transmission or reasoning, it will probably lead to greater accuracy and complexity. If, however, the change in cognitive complexity is brought about by manipulating the task along the resource dispersing dimension, for example by giving or taking away planning time or background information, then a decrease in accuracy and complexity can be expected. Fluency is expected the be affected negatively in both cases. If, however, a task is made more complex along both dimensions, as it often happens in real life, then synergetic effects are likely to be witnessed, which can only be revealed through empirical research. The introduction of the resource directing/dispersing distinction is important because previously authors tended to acknowledge resource dispersing effects of increases in cognitive complexity only. Consequently,

Skehan (1998) for example argued that increasing a task's cognitive complexity results in decrease in accuracy, complexity and fluency.

### *2.3.6. Evaluating task performance*

When analysing task performance, most authors use measures that somehow reflect their priorities; for instance those primarily concerned with the negotiation work involved in the tasks use negotiation of meaning indices for evaluating task performance (Plough & Gass, 1993). Foster and Skehan (1996), however, who question the primacy of negotiation work suggest more general measures of language performance. Drawing on a fundamental distinction between form and meaning, they call for the use of *accuracy* (a conservative orientation towards the focus on form), *complexity* (focus on form as using more elaborate language), and *fluency* (primacy of meaning while coping with real-time communication) measures and define these variables operationally. Citing evidence from a factor analytical study (Skehan & Foster, 1997), they argue effectively for the independence of these three measures of language performance. By choosing the above measures they succeed in reconciling the form-meaning dichotomy underlying the theory of task-based instruction with adequate evaluation measures.

Accuracy, complexity and fluency can be operationalized either as specific or as general measures. Foster and Skehan (1996) argue for using general measures as these are likely to be more sensitive and reflect weaker effects. Accuracy, as a general measure of language performance, is mostly operationalized as the proportion of error-free clauses (Foster & Skehan, 1996; Iwashita, McNamara, & Elder, 2001; Robinson, 2001b, 2007b; Skehan, 2001; Skehan & Foster, 1997, 1999), whereas syntactic complexity is often measured through a subordination index. In order to measure syntactic complexity, data are coded into communication units (c-units; Brock, 1986), then an index is calculated showing the number of clauses per these units (Bygate, 1999; Foster & Skehan, 1996; Iwashita, et al., 2001; Robinson, 2001b, 2007b; Skehan, 2001; Skehan & Foster, 1997, 1999). Robinson (2001b, 2007b) also attempts to capture *lexical complexity* reflecting lexical variety, and he uses the type-token

ratio (TTR) for this purpose, which is the total number of different words (types) divided by the total number of words (tokens) produced.

The issue of measurement of fluency appears to be less straightforward than that of accuracy and complexity, or at least there are a greater range of possible measures that can be considered. Based on findings from factor analysis, Foster and Skehan (1999) differentiated two broad factors of fluency: one is termed *breakdown fluency,* and it can be measured by the number of pauses and the amount of silence; while the other is called *repair fluency* encompassing repetition, false starts, reformulations and replacement. Later Skehan (2001) argued for measuring fluency by the number of pauses greater than 1 second in duration per five minutes of performance. Another possibility for determining fluency involves measuring speech rate, that is, the total number of syllables produced by the participant divided by the amount of total time, including pause time, required to produced the text, as it was done by Robinson (2007b). Calculating speech rate seems to be a feasible option as besides the relative ease of measurement, speech rate was found to be a reliable measure of fluency in a number of studies (Ejzenberg, 2000; Freed, 2000; Kormos, & Dénes, 2004; Lennon, 1990; Riggenbach, 1991).

Besides these general measures of task performance, specific measures can also be used to measure accuracy, complexity and fluency. Robinson and Gilabert (2007) argue that specific measures which are relevant to the particular resource directing variables making conceptual/linguistic demands should be used to supplement general measures. For example, tasks requiring complex reasoning about intentional states of others can be expected to result in the greater use of psychological and cognitive state terms (e.g.: think, expect, know) and of the complex syntactic predication the use of these terms requires. Therefore, these features should also be operationalized when examining tasks requiring complex reasoning about intentional states. Skehan (2001) also states that the specific measures used should be sensitive to experimental differences and clearly definable for operational purposes. Since these specific measures of task performance are necessarily diverse, and the nature of the measure to be used totally depends on the specific task at hand, they will not be discussed here in detail.

### *2.3.7 Empirical research on narrative tasks*

Narrative tasks were in the focus of Robinson's (1995) research that aimed to establish the effects of different levels of complexity on task completion. He used three cartoon strips (depicting stories) and six prompts (containing the first one or two sentences of each story) to encourage his participants to generate narratives under two conditions. The Here-and-Now condition called for the present tense description of the cartoons as the participants had the cartoons in front of them, while the There-and-Then condition called for past tense description without the presence of the pictures. Drawing on findings from research into first language acquisition (Brown and Bellugi, 1964, cited in Robinson 1995), empirical results of second language development (Meisel, 1987), and functional linguistic theory (Givón, 1989) Robinson hypothesised that the cognitive complexity of the There-and-Then narratives would be reflected by syntactically more complex language and greater accuracy, while the Here-and-Now narratives, presenting less processing burden, would result in more fluent performance. He only found empirical evidence for greater lexical density as shown by a higher ratio of lexical words, and for greater accuracy as indicated by more target-like use of articles in the There-and-Then condition. The lack of significant findings for the other two hypotheses might be accounted for by the small sample size (altogether 12 participants were included) or by the presumption that the tasks might not have been dissimilar enough with regard to the availability of contextual support, that is, they did not differ enough in their cognitive complexity.

Foster and Skehan (1996) set out to investigate the influence of task type and planning on second language performance using three types of tasks: a personal information exchange, a narrative, and a decision making task. The narrative task involved making up a storyline from five loosely but obviously connected pictures and sharing the story with a partner. The authors attempted to rank these tasks on a scale of difficulty proposed by Skehan (1996) and judged the narrative task to be of medium difficulty. The empirical findings showed, however, that a one-dimensional notion of task difficulty is unable to

handle the fact that tasks may differ along a number of dimensions. Thus though with regard to fluency the narrative task seemed to be of medium difficulty, it elicited the most complex but the least accurate language from the respondents.

Foster and Skehan (1996) also investigated the effects of two types of planning: detailed and undetailed. In accordance with the hypotheses put forward by the authors, the performance of the participants gradually improved through the no planning, undetailed planning and detailed planning conditions with regard to measures of fluency and complexity on the narrative task. The results were less clear in the case of accuracy; although they were able to identify more accurate performance in the undetailed planning condition than in the no planning one, detailed planning actually resulted in less accurate performance than the no planning condition. A possible explanation for this might lie in limited attentional capacity and in the trade-off effects between competing goals. The detailed planning condition might have pushed the learners above their competence level taking up a large portion of their attentional resources leaving no spare capacity to attend to form (Foster and Skehan, 1996).

In their next piece of research Skehan and Foster (1997) studied the effects of the above described three task types as well as two implementation conditions: planning, and awareness of a post-task activity. In this case the narrative task was slightly different, the participants were asked to tell a story depicted by a cartoon strip consisting of ten frames without a dialogue. Consequently, the results also differ somewhat from the findings of the previous research. In line with the hypotheses and the previous research findings, the participants' performance became more fluent and accurate in the planning than in the no planning condition. However, no significant effects were found for complexity. With regard to the other implementation condition, namely awareness of a post-task activity, the authors expected to find a selective effect for accuracy. This was not confirmed by the findings with regard to the narrative task, as the post-task condition did not affect the accuracy of the participants' performance, but it influenced one of the fluency

measures significantly (those aware of the post-task paused more during task completion).

Skehan and Foster (1997) also drew conclusions with regard to task characteristics. They argued that when performing tasks with a clear inherent structure, participants tend to invest the gain from planning into accuracy, as the obvious given storyline does not call for complex language thus attention can be directed towards accuracy. The narrative used in their previous study (Foster and Skehan, 1996), however, involved a great degree of on-line computation, and the need to express complex ideas pushed learners towards using more complex language taking up a considerable portion of attentional capacity. Skehan and Foster call attention to the important role of limited processing abilities, and thus to the need of the prioritisation of attentional resources and offer an information processing framework for the interpretation of the results.

The issue of task structure and processing conditions is taken one step further in a later study of Skehan and Foster (1999). The participants here were required to retell the story of one of two short television episodes of the Mr Bean series under different conditions. The episodes differed in the predictability of their structures; the restaurant script had a sequential task structure, a predictable storyline, while the golf episode had a far less predictable structure. Four performance conditions were used to manipulate the processing load of the task, and they were believed to represent decreasing levels of difficulty. Under the first condition the participants were required to watch and tell the story simultaneously, while the second condition was eased by giving a brief storyline of the episode before watching it. Further help was given under the third condition, the participants were first allowed to watch the whole episode before having to watch and tell it simultaneously for the second time. The easiest condition involved first watching then telling the story, thus removing the time pressure involved in the previous conditions.

Skehan and Foster (1999) hypothesised that clear inherent story structure would result in more fluent and more accurate performance yet would have no effect on the complexity measures. The decreasing processing load was believed to increase fluency, accuracy, as well as complexity. The

empirical findings lent support to only some of these hypotheses: the structured task was found to generate more fluent performance while performance condition (delayed or simultaneous narration) turned out to be related to the complexity measure. Here the greatest difference was found between the non-simultaneous as opposed to the three simultaneous conditions indicating that the need to keep up with the pace of the video in real time imposed an overwhelming processing burden on the subjects. The results concerning the accuracy measure are less clear than findings regarding fluency and complexity: it seems that it is an interaction of the variables of task structure and processing conditions that determine accuracy.

Iwashita et al. (2001) studied oral narrative tasks in an attempt to find further empirical support for Skehan's claims (1998) stating that task performance varies if a task is manipulated along different dimensions. They made oral narrative tasks more difficult along the dimensions of perspective, immediacy, adequacy, and planning time and expected decreases in accuracy, complexity, and fluency. Interestingly, they found no statistically significant differences between students' performance under the different conditions. Moreover, their only significant finding defied their expectations. They found that when the immediacy of narratives was varied, that is, students were required to tell a narrative in the There-and-Then condition instead of the Here-and-Now one (cf. Robinson, 1995), students' accuracy increased. Their results seem to lend partial support to Robinson's (2001b, 2003, 2005b) Cognition Hypothesis claiming that when a task is made more complex along resource-directing dimensions the language produced tends to be more complex and accurate but less fluent.

In a recent study Robinson (2007b) examined how students solve interactive narrative tasks, where the cognitive complexity of tasks was varied along a resource-directing dimension, intentional reasoning, that is in the more complex versions of the task more reference needed to be made to the characters' thoughts and intentions. Robinson's results are largely in line with the claims of the Cognition Hypothesis (Robinson's 2001b, 2003, 2005b): greater task complexity led to an increase in complexity when measured by specific measures designed on the basis of the conceptual and linguistic

demands of the task; however, general measures of accuracy, complexity and fluency were not affected. The cognitively more complex tasks led to more interaction and more uptake of premodified input, which lends support to a further claim of the Cognition Hypothesis. Concerning the effect of individual differences, it was revealed that output processing anxiety had a greater effect on performance as cognitive complexity increased.

### *2.3.8 Conclusion*

In conclusion, it can be stated that since tasks seem to play an important role in language teaching and testing, their features and characteristics and the effect these have on language performance should definitely be studied. Although several authors theorize about the impact of certain task features, there are only two cognitively motivated comprehensive frameworks of task characteristics which their authors attempted to substantiate using empirical evidence. However, results in connection with them are far from being conclusive. As regards Skehan's (1998) scheme of task difficulty, there are some empirical studies that lend support to it (e.g.: Foster & Skehan, 1996, 1999; Skehan & Foster, 1997, 1999), but there is also some counterevidence (Iwashita et. al, 2001) defying his hypotheses. The most recent theory is Robinson's (2001b, 2003, 2005b) Cognition Hypothesis and his Triadic Componential Framework for task classification (Robinson, 2001a, 2005b, 2007a), which is also supported by some empirical findings (Iwashita et al., 2001; Robinson, 1995, 2001, 2007b). Since the frameworks proposed by Skehan and Robinson hypothesise different outcomes of increasing task complexity with regard to certain task performance measures, it would be possible to test them empirically.

The review of literature suggests that oral narrative tasks are a frequently used task type in research studies; therefore, they are probably often used in classrooms and tests as well. Since I believe that this task type also provides an opportunity for the use of imagination, it seems suitable for demonstrating the effects of learner creativity. Although empirical findings about this task type are somewhat contradictory, some general conclusions can

be drawn based on the results. It seems that task performance benefits from available planning time, and that changes in the cognitive complexity of tasks is likely to affect performance measures. Moreover, individual differences are more likely to influence performance on more complex tasks.

## 2.4 Conclusion: The relevance of creativity for learner performance on tasks

Having reviewed the literature on creativity and tasks separately, I will now attempt to highlight possible points of interaction. The relevance of creativity for learner performance on tasks can be examined on two levels. One of them is the level of specific cognitive mechanisms that are believed to contribute to creativity. Since the instrument I intend to use as a test of creativity in my research is aimed at identifying divergent thinkers, it should be pointed out why I feel that divergent thinking might be advantageous for foreign language learners when tackling language tasks. The other level is the wider context of language teaching methodology, more specifically the use of communicative methods and more recently task-based instruction; in these approaches the use of drills is discouraged and emphasis is placed on conveying meaning. Despite the fact that these two levels can be considered as separate theoretically, I am aware that they interact to a great extent in practice: in most cases language learning is mediated by some kind of methodology.

On the basis of the literature review, I hypothesise that since creativity is usually manifested in production, that is, in creative products, its effects would probably be more easily detectable in output as opposed to comprehension. I believe that there are a number of reasons why language tasks, especially open-ended ones like narrative tasks, where there is no correct solution but a large number of solutions are possible, could be better suited for creative foreign language learners. Since creative learners are characterised by greater fluency, that is they provide a larger number of solutions in a given amount of time (Baer, 1993), they might be able to talk more during the tasks. As it has been suggested by Swain (1985), producing more comprehensible output has a beneficial effect on language acquisition. Flexibility, the second

facet of creativity measured by divergent thinking tests, which reflects the ability to produce a wide variety of ideas (Baer, 1993), might be manifested directly in the way language is used by the learners: if their language competence is sufficient, they might in fact use a wider range of vocabulary items in order to express their wide range of ideas. Similarly, originality, the ability to produce unusual ideas (Baer, 1993), might also prompt learners to employ a wide range of vocabulary in an attempt to give an account of the interesting ideas they have in mind. Although the above-mentioned qualities of creative people might be advantageous in any language task, I feel that narrative tasks, which obviously rely on learners' imagination, might intensify the effect of creativity on language performance. Therefore, despite the fact that the imaginativeness or creativity of the stories themselves is measured, I believe that narrative tasks would be suitable for conducting research on the effects of creativity on output.

# Chapter 3: Pilot Studies

## 3.1 Introduction

This chapter discusses findings of preliminary empirical investigations carried out prior to conducting the research for the dissertation. Since I first started to investigate the effects of creativity on oral narrative task performance in my MA thesis, I would like to summarize my findings briefly, and draw conclusions from the problems I encountered there. Then, the results of a validation study aiming at the piloting and validation of oral narrative tasks used for the purpose of the dissertation are presented in detail.

## 3.2 MA Thesis

In my MA thesis I examined the effects of learner creativity on the performance of oral narrative tasks. The research conducted involved a small sample of twenty-four high-school students instructed by communicative methods. The sample was selected on the basis of creativity scores received on a standardized test of creativity (Barkóczi, & Zétényi, 1981) in a way to include students with high and low creativity. The language output of these students was examined on two oral narrative tasks. The results showed that after eliminating the effect of the subjects' different levels of proficiency using partial correlations, the output produced by highly creative students on these tasks was both quantitatively and qualitatively superior to the output of students with low creativity. Students characterised by greater creative fluency produced more words, while original students used a higher ratio of different verbs in their stories. It is important to note that based on the results, it seems that creativity is not a uniform, one-faceted quality; it is much more like a set of cognitive characteristics involved in different aspects of solving a task: that is, original students were not characterized by high creative fluency for example. Another interesting finding was that the third component of creativity measured by this test, relative flexibility, seemed to be related to language

proficiency as measured by C-test scores although the relationship could only be demonstrated at a tendency level ($p<0.1$).

The process of conducting the research and writing up the results brought some problems to light. First of all, it seemed that the narrative task used in the research prompted descriptions rather than narratives in some cases. Therefore, it seemed important to find oral narrative tasks that validly elicit narratives from the respondents; this issue is addressed in the validation study discussed in section 3.3 of this chapter. Also, the measures I employed for assessing the quality - language and content - of the stories, especially the measures used for lexical variety, were unfortunately not comparable to the measures found in the literature. Moreover, the three-point scale I used for evaluating the contents of the stories was probably too robust and did not work very well; no significant or even tendency level correlations could be established for any of the variables. Apart from this scale, however, I had no instrument for judging the non-linguistic aspects of the stories. These problems necessitated the introduction of some new task performance measures, which are discussed in detail in section 4.7.2. Furthermore, since it was suggested that the relationship detected between creativity and language proficiency might be an artefact in the sense that the C-test as a task-type might favour learners with this particular type of creativity, using a more comprehensive measure of language proficiency seemed advisable.

Another group of problems concerned the amount of background data available about the participants. Since the intended aim of my research is to identify creativity as a possibly significant individual variable related to oral narrative task performance, it is important to have background information about at least one other, already established, significant individual variable that is known to be related to language achievement, such as language aptitude. Having background data would enable me to demonstrate the relative weakness or strength of creativity compared to other individual variables. In addition, since recent models of language aptitude cover some creativity-related processes (Grigorenko, et al. 2000, cited in section 2.2.1) it seems justifiable to investigate whether creativity and language aptitude are related in any way.

## 3.3 Validation Study

This section presents a detailed description of a validation study. The aim of this pilot study was the validation of oral narrative tasks that could later be used as elicitation devices.

### *3.3.1 Introduction*

The reliability and validity of all tasks and items are important questions of test and questionnaire construction, but these issues usually do not receive serious attention when tasks are used as elicitation instruments in research. Since research articles, on oral narrative tasks for example (Bygate, 1999; Foster & Skehan, 1996; Iwashita et al., 2001; Robinson, 1995, 2001, 2007b; Skehan & Foster, 1997, 1999), disregard this issue completely, it seems that researchers only rely on assumptions, hunches and experience when they design their tasks. The problem is that this way they can only hope that their expectations, that is, that a task is of a certain type and will result is a certain kind of performance, will be met by their students. The current practice seems to be that as long as no obvious discrepancy becomes apparent between how the task is supposed to function and the way it really works, there is no need to be concerned about the validity of the tasks. This approach, however, is certainly problematic. Since these issues are of crucial importance with regard to the usability and usefulness of any task, they should be addressed not only in connection with test tasks, but also regarding research tasks.

The research described in this section of the dissertation was aimed at the validation of different types of oral narrative tasks, which could later be used to elicit stories from students. For the in-depth investigation of the issue, the task-based literature was surveyed for oral narrative tasks used as elicitation devices, and the testing literature was consulted to provide guidance with regard to the validity issues relevant for our purposes. Since the construct underlying oral narrative tasks is the narrative genre itself, a conceptual definition and characteristics of the narrative genre also had to be compiled from the literature. The conceptual definition of narratives was then converted into an operational definition, which served as a starting point for the empirical

investigation. In the course of the empirical research, three types of oral narrative tasks were designed and piloted, and the narratives elicited were analysed in order to establish the validity of each task type.

### *3.3.2 Oral narrative tasks described in the literature*

In order to find the basic task types used for the elicitation of oral narratives, research articles of authors working within the task-based framework were consulted, and on the basis of these, three basic types of oral narrative tasks could be identified. A frequently used technique involves giving cartoons to the respondents and asking them to narrate the story depicted by the pictures (Skehan & Foster, 1997). In some cases the cartoons are supplemented with prompts, such as the first one or two of sentences of the story to be told (Robinson, 1995). Another variant of the same task type is when the stimulus is a short film instead of a cartoon, and respondents are asked to narrate the story of the film (Skehan & Foster, 1999). Another type of oral narrative task is when the participants are given loosely but obviously connected pictures, and they are asked to invent the story themselves (Foster & Skehan, 1996). This task is less specific and less structured than the previous one, as the respondents have to come up with a logical arrangement for the pictures and create the story themselves. The third type of task used for the elicitation of narratives is when the participants are given only one picture, which can either be the beginning, the middle or the last stage of the story, but they are given complete freedom regarding the invention of the rest of the story (Csölle & Károly, 1998). This task type is the least structured and the least specific of the three, and this is the one that draws on respondents' creativity the most. The first task type is used much more frequently as an elicitation device in second language research than the other two, probably because on the one hand the specificity and the structured nature of the task seem to guarantee the desirable output, narratives, and on the other hand because it does not draw so obviously on students' imagination. Despite the fact that narrative tasks are frequently used for research purposes in many different variants, no reference to task validation could be identified in any of the research articles consulted (Foster

& Skehan, 1996; Iwashita, et al., 2001; Robinson, 1995, 2001, 2007b; Skehan & Foster, 1997, 1999).

### *3.3.3 Issues of validity in the testing literature*

In the glossary of their book on testing, Alderson, Clapham, and Wall (1995) define validity as "the extent to which a test measures what it is intended to measure: it relates to the uses made of test scores and the ways in which test scores are interpreted, and is therefore always relative to test purpose" (p 296). This definition needs to be modified when discussing the validity of an elicitation device, instead of a test. Defining *validity as the extent to which the device elicits what it is intended to elicit relative to the purpose of the elicitation device*, seems to be more appropriate here; and this statement will be adopted as a working definition of validity here. The modification of the definition of validity results in the fact that aspects of external validity, such as concurrent and predictive, cannot be employed. Since the tasks to be validated are elicitation devices, there is no valid and reliable data available about the tasks from other sources that could be used for concurrent validation. The same is true for predictive validity: there is no data about future performance that could serve as a basis of comparison. Conversely, measures of internal validity, such as face and content, and construct validity of the elicitation device could probably be examined, and they could contribute to the overall validity of the device.

Construct validity in the case of an elicitation task should be interpreted as the correspondence between the construct hypothesised to underlie output and the characteristics of the actual output. Therefore, in the case of an oral narrative task where the underlying construct is the narrative genre which is hypothesised to be manifested in narratives as output, the actual output of the task should be a narrative as well. In the course of the validation procedure, we should attempt to find evidence that the texts produced in response to such a task bear the features of narratives; therefore, the identification of the basic characteristics of oral narratives is crucial in this respect. The content validity of an elicitation device should be evaluated as overlaps between task content

and the content of the output produced by participants on the task. Finally, its face validity should be based upon respondents' judgements about the instrument's purpose, the difficulties encountered, and issues concerning its administration, such as for example instruction and planning time.

An essential part of the validation procedure should be establishing the reliability of our instrument, proving that whatever our task elicits, it elicits that consistently, which turned out to be a somewhat problematic undertaking in the present research. Reliability in the testing literature (Alderson, et al., 1995) refers to the consistency of test scores, and there are several ways of demonstrating this consistency in the case of 'objective' tests, for example by calculating test-retest, parallel form or split-half reliability. The reliability of 'subjective' tests is usually measured by calculating the reliability of the marking. Since the aim of the present study is neither the validation of a scoring scheme for tasks performance nor the validation of task performance measures that can be used to evaluate language elicited on tasks, we are left without a sophisticated scoring scheme where the reliability of marking could be demonstrated. Moreover, since establishing reliability usually involves statistical procedures, a relatively large sample size is needed for this purpose. For these reasons, the issue of reliability will not be addressed explicitly in connection with the elicitation tasks. It will only be hinted at when attempting to identify balanced performance of students on the elicitation tasks.

### *3.3.4 Narrative genre*

The characteristics of the oral narrative genre were studied by a number of authors, focusing on different types of narratives both written an oral. It was Labov, however, who first offered a comprehensive, still relatively easily manageable framework for the analysis of spoken narratives. In his influential book Language in the inner city, Labov (1972) defines "a *minimal narrative* as a sequence of two clauses which are *temporally ordered*: that is, a change in their order will result in a change in the temporal sequence of the original semantic interpretation" (p 360). According to Labov the skeleton of any narrative consists of narrative clauses, which are temporarily ordered

independent clauses connected by temporal junctures. Subordinate clauses cannot serve as narrative clauses as it is possible to change their order without changing the original semantic interpretation. Those independent clauses that do not have a fixed temporal order, that is they are not joined by temporal junctures, do not qualify as narrative clauses for the same reason; a change in their order does not necessarily bring about altered semantic interpretation.

Although according to Labov's (1972) definition, two temporally ordered narrative clauses already qualify as a narrative, the majority of oral narratives have a considerably larger number of narrative clauses and a similar basic structure, consisting of several stages. Labov describes six stages of the fully formed narrative; these are the abstract, the orientation, the complicating action, the evaluation, the result or resolution and the coda. The narrative usually starts with an abstract serving as a title introducing what the story will be about. The next stage of a story is the orientation; this informs the audience about the world of the story: the time and place, the participants and the circumstances. Having introduced the background, the story proceeds to the complicating action, that is the events making up the story; on the basis of our narrative definition this is the only truly obligatory part of the narrative. Evaluation as a separate phase is usually introduced into the story at this point, offering an interpretation of the events before the resolution, where the events are sorted out. The coda as the last stage provides a bridge between the events in the story and the present situation of the narration.

Labov (1972) considers evaluation as having central importance in narratives, and he argues that it is perhaps the most important element in narratives besides narrative clauses. Its significance lies in the fact that evaluation is what signals the reason why the narrative is told; thus, the term evaluation covers all the means that the storyteller uses for indicating the point of the narrative. In a large number of cases, however, evaluation does not appear as a separate stage of the story but is interwoven into the whole text with the help of dispersed evaluative devices, such as external evaluation, embedded evaluation, evaluative action and evaluation by suspension of the action.

For the purpose of this empirical study, Labov's (1972) definition was adopted as a minimal requirement for any text to be identified as a narrative. The presence or lack of the six stages of the fully formed narrative also served as an important guideline in the identification of narratives; the more stages a narrative contained the better. In order to ease the identification of the six stages of narrative, the works of authors (Hatch, 1992; Liskin-Gasparro, 1996; McCarthy & Carter, 1994) working within the field of applied linguistics were consulted, who used the Labovian framework for analysis. However, based on Labov's definition the complicating action stage was the only obligatory part. As the backbone of narratives is made up of narrative clauses, the texts elicited by the tasks were analysed for these. Narrative clauses reflect the events of the story; therefore, a higher number of such clauses probably signals longer and more complex stories as far as their event structure is concerned. For this reason, a relatively high ratio of narrative clauses was expected in fully formed narrative texts. Although the evaluative structure of narratives is believed to be crucial as well, evaluation was only discussed together with the six stages of the narrative in the empirical analysis. The reason for this is that the circumstances under which these narratives were elicited are somewhat artificial; thus, it would not have been realistic to expect the respondents to make a point by telling the story, to provide a reason for telling it other than the fact that this was the task.

### *3.3.5 Research question*

The aim of the pilot study was to identify tasks that elicit narratives as opposed to texts belonging to other genres, such as description. Based on a review of literature, three basic variants of the narrative task were identified: one using a single picture as prompt, another one employing loosely connected pictures and a third one presenting cartoons. The research question guiding the pilot study was which of the three task variants qualify as valid oral narrative tasks. Since the language output produced on the tasks would later be analysed using linguistic measures not discussed here, it would also be important to know which tasks can be expected to result in longer texts and a relatively balanced performance.

### *3.3.6 Method*

#### 3.3.6.1 Participants

The participants of the pilot study were 11 English major university students, three males and eight females, who attended the same first year academic writing class partly instructed by the researcher.

#### 3.3.6.1 Instruments

*3.3.6.1.1 Oral narrative tasks*

The three variants of narrative tasks to be tested were selected on the basis of the review of literature. The selected narrative task variants were the "single picture task", the "picture sequence task" and the "cartoon strip task". Since students were required to solve the tasks in pairs, two versions of each of the three oral narrative tasks were prepared; thus, altogether six oral narrative tasks were used. For the single picture task, the stimuli were adopted from a psychological test aimed at eliciting narratives, Murray's Thematic Apperception Test (TAT); one of the pictures chosen (Picture A) showed a young woman in the doorway (see Appendix A) while the other one (Picture B) portrayed a young boy with his violin (see Appendix B). The picture sequences, one (Sequence A) showing different stages of a quarrel (see Appendix C) the other (Sequence B) presenting scenes from the life of a man (see Appendix D) were selected from various teaching resource books, similarly to the comic strips: one (Cartoon A) about a shipwrecked man (see Appendix E) and another (Cartoon B) about a wizard (see Appendix F). The English language instruction in each case asked for the invention of a story based on the stimulus (see the instructions for each task in Appendices A-F), which could be supplemented by any additional detail. In the case of the single picture task, which was the least structured of the three, participants could decide whether the picture given to them represented the first, the last or a middle stage in their story, and they were granted freedom regarding other details as well. The picture sequence task involved first arranging the separate pictures in a sensible order, then telling the story on the basis of them. The

most structured task was narrating the story shown by a comic strip, where the sequence of the pictures was predetermined. The instructions allowed for five minutes of preparation time, after which one member of the pair narrated the story with the other member listening, then the roles were reversed.

The reason why the narrative tasks were performed in pairs was that this is the way this task type is usually performed in the classroom or sometimes even at exams. My aim was to approximate usual implementation conditions as far as possible. Planning time was set for the tasks because research results show that planning affects task performance positively. Mehnert (1998) suggests that the use of different amounts of planning time might be favourable when trying to enhance various aspects of performance. On a sample of intermediate students, he found that while planning intervals up to ten minutes progressively improve the fluency measures, time given above this limit seems to have little additional effect. In the case of accuracy, the same limit can be found at one minute, after which no significant improvements of accuracy can be detected. With regard to complexity, the optimal planning time is regarded to be ten minutes, shorter intervals having no obvious effect on this variable. Since students in my sample were language majors, that is advanced learners, providing five minutes of planning time seemed sufficient.

#### *3.3.6.1.2 Post-task interview*

An additional instrument, a post task interview (see Appendix G) was used as well, for validation purposes. It contained questions regarding the overall liking, the easy and difficult features, and the hypothesised purpose of the tasks. The participants' opinion concerning the content and wording of the instruction, the length of preparation time and their past experiences with similar tasks were also explored. Since the native language of the participants was Hungarian, the post task interview was conducted in Hungarian in order to guarantee the unhindered expression of ideas.

### 3.3.6.2 Procedure

Data collection took place in May 2000, on a voluntary basis, after the ending of the participants' academic writing course. Those wishing to take part in the study made an appointment with the researcher in pairs or threes, and completed one to three tasks depending on the amount of time they had; the majority of the students completed two tasks. Altogether 22 tasks: eight single-picture, seven picture sequence and seven comic strip, were completed by the participants. The pairs performing the tasks were interviewed together after the completion of each task. The task completion and the proceeding post-task interview was audio-recorded with the participants' consent for later analysis.

### 3.3.6.3 Analysis

The recordings of the tasks were first transcribed, and the transcripts were analysed in various ways later on. The first step of the analysis involved finding the six stages of the fully-formed oral narrative as identified by Labov (1972) and described in the literature review section. The detailed evaluation criteria were compiled on the basis of articles using the Labovian system in the field of applied linguistics (Hatch, 1992; Liskin-Gasparro, 1996; McCarthy & Carter, 1994) (see Appendix H). The next stage was an in-depth analysis of the narratives uttered; an important issue here was the identification of appropriate chunks within the spoken data, which would reflect the psychological planning process (Foster, Tonkyn, & Wiggleworth, 2000) that is, the thoughts of the speaker. From the possible ways of segmentation, the method of dividing the texts into analysis of speech units (AS-units) (Foster et al., 2000) was adopted, as this system was specifically designed to handle oral data. The analysis followed the guidelines suggested by Foster et al. (see their article for a thorough description of the process). I hoped that by counting the AS-units, the number of thoughts formulated by the respondent could be captured. The next step of the analysis involved identifying the narrative clauses (Labov, 1972) in the texts as defined in the literature review section above. The total number of narrative clauses as well as their ratio to the total number of AS-units were thought to reveal qualitative differences between the narratives produced. As regards the post-task interviews, notes were taken on the basis of the

recordings and these were analysed later on; full transcripts of the post-task interviews were not prepared.

### *3.3.7 Results and discussion*

Findings of the task validation study are presented below. First the three variants of oral narrative tasks: the single picture task, the picture sequence task, and the cartoon strip task are discussed separately. Issues of face validity are addressed first in each section followed by aspects of content validity and construct validity. The discussion of individual tasks is concluded by a comparative analysis of the three tasks.

#### 3.3.7.1 Single picture task

The face validity of the single picture task was explored through the post-task interviews, in which respondents expressed an overall liking for the task highlighting the fact that it gave them an opportunity to use their imagination and express their own ideas and thoughts. The majority of students were familiar with the task type, and one respondent mentioned that in many ways it is similar to telling tales to children. Considering that the aim of the task is the elicitation of narratives, this statement is favourable with regard to the face validity of the instrument. None of the respondents reported problems understanding the instruction, and they were able to recall it precisely. Despite the fact that none of the respondents spent five minutes with preparation, they judged the amount of planning time adequate or even too short. When asked about the researcher's intentions, they mentioned that the task probably measures language proficiency, imagination, or clichés in stories.

As the single picture task was the least structured of the three tasks, and it would have been possible to produce a large number of different narratives on the basis of the picture, its content validity is not very easy to evaluate. Only 4 out of the 8 respondents incorporated the event portrayed by the picture explicitly in their stories, which suggests a loose relationship between the picture and the stories. It seems, however, that all the respondents perceived

the atmosphere depicted in the pictures, and they created their stories in accordance with that.

The two single picture tasks resulted in varied performance on the part of the participants as far as the quality of narratives produced by them is concerned. Out of the eight narratives, one definitely cannot be classified even as a minimal narrative in the Labovian sense, as it does not contain two consecutive events, the minimal requirement set in our definition (Labov, 1972). It is more like a listing of possible interpretations of the atmosphere portrayed by the picture. Out of the remaining seven stories, three are not very elaborate, they contain very short orientations and a very limited number of complicating actions. In the remaining four narratives, the orientation, the complicating action and the resolution stages of the Labovian narrative can be identified; these are full-blown and adequately elaborated. They also contain evaluation, though it is never a separate stage but interwoven into the whole of the text. The lack of abstract and coda, which was characteristic of all the narratives elicited regardless of the specific task type, can probably be accounted for by the somewhat unnatural conditions under which these narratives were created. The fact that these stories were produced in response to a task probably affected the respondents in a way that they felt no urge either to give a title in order to grab the listeners attention, which resulted in a lack of abstract, or to evaluate the story's relevance to the present situation, which led to a failure to provide a coda.

A more thorough analysis, the results of which are presented in Table 1, confirmed the impressions formed on the basis of the global analysis and revealed further problems with the task as well. The number of AS-units, reflecting the number of thoughts or ideas the speaker produced, varied from 5 to 27, which suggests big differences in the output across participants in line with our previous finding. A novel aspect of task performance in this analysis is the number of narrative clauses, and their ratio to the total number of AS-units. As we can see the number of narrative clauses varies greatly as well, from a minimum value of 1 to a maximum value of 16, and this wide range is not a consequence of the different lengths of the narratives, that is, short narratives do not necessarily have a low number of narrative clauses and vice

versa. The ratios of the number of narrative clauses to the total number of AS-units seem to suggest that while some respondents included a large number of events, others talked a lot without using narrative clauses; therefore, the quality of narratives produced is quite varied.

Table 1
*Respondents' performance on the single picture tasks*

| code of narrative | total number of AS-units in narrative | number of narrative clauses in narrative | ratio of number of narrative clauses to total number of AS-units |
|---|---|---|---|
| 1/A | 20.00 | 15.00 | 0.75 |
| 2/A | 17.00 | 6.00 | 0.35 |
| 3/A | 12.00 | 2.00 | 0.17 |
| 4/A | 5.00 | 2.00 | 0.40 |
| 5/B | 20.00 | 10.00 | 0.50 |
| 6/B | 27.00 | 16.00 | 0.59 |
| 7/B | 11.00 | 1.00 | 0.09 |
| 8/B | 7.00 | 3.00 | 0.43 |

*Note.* Figures referring to narratives produced in response Task A - "woman in doorway" are unmarked, whereas figures referring to narratives produced in response to Task B - "boy with violin" are highlighted in grey.

The great extent of variance displayed in the task performance indicates that although there were some respondents who were able to comply with the task requirements and produce well-formed narratives, the single picture task did not necessarily elicit narratives. Since, according to the working definition adopted in this paper, the elicitation of narratives is the central issue regarding the construct validity of an oral narrative task, it cannot be claimed that the single picture task is indeed an oral narrative task, as it either fails to elicit or results in poor quality narratives in a number of cases.

### 3.3.7.2 Picture sequence task

The post-task interview revealed that respondents generally liked the task, they found it interesting and challenging, but they also added critical remarks about the quality of the drawings and the limited possibility of interpretation. The majority of the participants were familiar with the task, they were able to recall the instruction and judged it comprehensible. With regard to preparation time, their opinions varied: some respondents found the five minutes planning time adequate, while others thought it either too long or too short, which is rather surprising in the light of the fact that none of them spent five minutes on preparation. The respondents attributed a number of different purposes to the task, they thought it might be used for assessing language proficiency, creativity, imagination, logical thinking and the ability to produce a narrative. The respondents' remarks suggest that although the picture sequence task as a type of oral narrative task is perceived as having some face validity, the task content, the actual pictures might be worth experimenting with or even changing.

The content validity of this task is somewhat easier to judge than that of the single picture task, although several interpretations were possible here as well. All three respondents receiving the life scenes narrated a short life story, while three out of the four interpreted the quarrel scenes as phases of a disagreement occurring for various reasons. This suggest that they relied on the information conveyed by the pictures, which reinforces the content validity of this task.

The construct validity of this task was again analysed in two steps. When attempting to identify the six stages of the Labovian narrative in the two picture sequence tasks, one gets the impression that the seemingly similar tasks are in fact quite different. While the narratives, two shorter and two longer ones, elicited by the quarrel scenes (see Appendix C) all seem to be full-blown stories with orientation, complicating actions, resolution as well as some evaluation, the narratives produced in response to the life scenes (see Appendix D) appear as sketches, loosely connected events from a life with scarce evaluation.

The analysis of the number of AS-units and true narrative clauses, shown in Table 2, presents a more favourable picture, however. The respondents' performance on this task was obviously more balanced than on the single picture task, the minimum and maximum values for the AS-units and the true narrative clauses being 6 and 17, and 3 and 12, respectively. This interpretation is supported by the ratios of the number of narrative clauses to the total number of AS-units as well; with the exception of the lowest value 0.33, the ratios fall into quite a narrow range here. It means that in the case of this task those respondents who produced more output also used a higher number of narrative clauses, while those who expressed fewer ideas employed fewer narrative clauses as well. This suggests that despite the variance in the length of narratives produced, the quality of the narratives was quite balanced on this task.

Table 2
*Respondents' performance on the picture sequence task*

| code of narrative | total number of AS-units in narrative | number of narrative clauses in narrative | ratio of number of narrative clauses to total number of AS-units |
|---|---|---|---|
| 1/C | 17.00 | 11.00 | 0.65 |
| 2/C | 17.00 | 12.00 | 0.71 |
| 3/C | 11.00 | 6.00 | 0.55 |
| 4/C | 9.00 | 6.00 | 0.67 |
| 5/D | 10.00 | 7.00 | 0.70 |
| 6/D | 6.00 | 3.00 | 0.50 |
| 7/D | 15.00 | 5.00 | 0.33 |

*Note.* Figures referring to narratives produced in response Task C - "quarrel" are unmarked, whereas figures referring to narratives produced in response to Task D - "life scenes" are highlighted in grey.

Although the analysis of the stages revealed problems with one of the tasks drawing our attention to the effect of task content and through this to the need of piloting, the picture sequence task seems to comply more with the requirements set for a valid oral narrative task than the single picture task did. Thus, it is probably justifiable to call this type of task an oral narrative task, as it results in the production of similar, narrative texts in the majority of cases.

### 3.3.7.3 Cartoon strip task

The results of the post task interviews show that respondents evaluated this task positively as well; they found it entertaining and imaginative. The cartoon strip task was perceived as qualitatively different from the single picture and picture sequence tasks in a sense that the ready-made story portrayed by the pictures made the task more specific. This had a twofold consequence: the need to rely on creativity was reduced making the task easier on the one hand, but on the other there was also a need to use specific words, which made the task more difficult. The respondents were familiar with the task type, found the instruction easy to follow, and despite the fact that noone used up the five minutes planning time entirely, they considered it adequate. The possible purposes of the task were identified as assessment of proficiency, vocabulary, creativity and the ability to make up coherent stories.

Because of its greater specificity, this was the task where content validity, in our case the relationship between the events represented by the pictures and those appearing in the narratives, was the easiest to examine. When analysing the narratives, it becomes apparent that there is a close correspondence between the pictures and the stages of the story. Although it seems to reflect high content validity, this phenomenon proved to be harmful in certain cases as it induced the narrator to alternate between narration and picture description. It is also interesting to note that some of the participants failed to understand the punch line of the cartoon, but this did not prevent them from giving a full account of the events.

The analysis of the stages of the narrative revealed that this task resulted in the most similar output. Although, similarly to the single picture and

picture sequence tasks, the abstract and the coda were absent, the orientation, the complicating action and the resolution stages are present and quite elaborated. The relatively stable number of complicating actions can probably be attributed to the specificity and highly structured nature of the task. Evaluation is not included as a separate stage, but evaluative remarks can be identified in each narrative.

A more detailed analysis of the number of AS-units presented in Table 3 confirms that the narratives produced in response to this task are not extremely different in terms of length, since the minimum number of narrative clauses is 9, while the maximum is 20. By examining the number of narrative clauses in each text, we may conclude that five out of the seven respondents used six to eight narrative clauses, which might be in connection with the fact that each cartoon strip consisted of six pictures describing six events. The ratios of the narrative clauses to the total number of AS-units, similarly to the picture sequence task, do not display great variance, which can be interpreted as a tendency for longer narratives to contain more events and for shorter ones to contain fewer. There seems to be no qualitative difference between the stories from this respect.

Table 3
*Respondents' performance on the cartoon strip task*

| code of narrative | total number of AS-units in narrative | number of narrative clauses in narrative | ratio of number of narrative clauses to total number of AS-units |
|---|---|---|---|
| 1/E | 20.00 | 8.00 | 0.40 |
| 2/E | 12.00 | 6.00 | 0.50 |
| 3/E | 9.00 | 4.00 | 0.44 |
| 4/F | 15.00 | 11.00 | 0.73 |
| 5/F | 19.00 | 7.00 | 0.37 |
| 6/F | 16.00 | 8.00 | 0.50 |
| 7/F | 12.00 | 8.00 | 0.67 |

*Note.* Figures referring to narratives produced in response Task E - "shipwrecked man" are unmarked, whereas figures referring to narratives produced in response to Task F - "wizard" are highlighted in grey.

Evidence from the post-task interviews and the analysis of the cartoon strip task suggests that it indeed elicits narratives; therefore, it is a valid oral narrative task.

### 3.3.7.4 Comparative analysis

Having analysed the tasks separately in detail in the previous section, I wish to shed light on some differences between the task variants with the help of a comparative analysis of the tasks along the variables discussed above. Although it would have been possible to use statistical tests (one-way ANOVA) for the comparison of the different task types, generalizations would have been impossible to draw on the basis of them because of the small sample sizes; therefore, the interpretations offered are based on eyeballing the results.

The descriptive statistics for the total number of AS-units presented in Table 4 indicate that the single picture task and the cartoon strip tasks are quite similar as far as the mean number of AS-units are concerned. However, the variability of performance revealed by the dispersion measures of range and standard deviation suggest a similarity between respondents' performance on the picture sequence and cartoon strip tasks. These tasks appear to result in more balanced performance.

Table 4
*Descriptive statistics of the total number of AS-units in the three tasks*

| variable | number of cases | minimum value | maximum value | range | mean | standard deviation |
|---|---|---|---|---|---|---|
| number of AS-units in single picture task | 8 | 5.00 | 27.00 | 22.00 | 14.87 | 7.43 |
| number of AS-units in picture sequence task | 7 | 6.00 | 17.00 | 11.00 | 12.14 | 4.25 |
| number of AS-units in cartoon strip task | 7 | 9.00 | 20.00 | 11.00 | 14.71 | 3.98 |

In the case of narrative clauses a similar tendency can be detected; the ranges and standard deviations of the picture sequence and cartoon strip tasks differ much less compared to the single picture task as shown in Table 5. This finding again reflects the similarity of participants' output on these two tasks, and also suggests low variance within each of these task types.

Table 5
*Descriptive statistics of the total number of narrative clauses in the three tasks*

| variable | number of cases | minimum value | maximum value | range | mean | standard deviation |
|---|---|---|---|---|---|---|
| number of narrative clauses in single picture task | 8 | 1.00 | 16.00 | 15.00 | 6.87 | 6.05 |
| number of narrative clauses in picture sequence task | 7 | 3.00 | 12.00 | 9.00 | 7.14 | 3.23 |
| number of narrative clauses in cartoon strip task | 7 | 4.00 | 11.00 | 7.00 | 7.42 | 2.14 |

The analysis of the ratios of narrative clauses to the total number of AS-units show that on the picture sequence and cartoon strip tasks the rate of the narrative clauses to the AS-units was higher, that is these tasks contained a higher proportion of narrative clauses as shown in Table 6. Apart from this, both the range and the standard deviation of these tasks is lower; the participants' performance was less varied on these tasks.

Table 6
*Descriptive statistics of the ratio of the total number of narrative clauses to the total number of AS-units in the three tasks*

| variable | number of cases | minimum value | maximum value | range | mean | standard deviation |
|---|---|---|---|---|---|---|
| ratio of number of narrative clauses to total number of AS-units in single picture task | 8 | 0.09 | 0.75 | 0.66 | 0.41 | 0.21 |
| ratio of number of narrative clauses to total number of AS-units in picture sequence task | 7 | 0.33 | 0.71 | 0.37 | 0.58 | 0.13 |
| ratio of number of narrative clauses to total number of AS-units in cartoon strip task | 7 | 0.37 | 0.73 | 0.36 | 0.51 | 0.13 |

Evidence from the comparative analysis suggests that the picture sequence and cartoon strip tasks are more similar to each other in a number of respects than to the single picture task. First of all, the former two tasks always resulted in less variance, the performance of participants on these tasks was always more similar both regarding the number of AS-units and the number of narrative clauses. Secondly, the output produced in response to the picture sequence and cartoon strip tasks seem to conform better to the requirements of a narrative as they include a higher ratio of narrative clauses, which is crucial as they provide the skeletal structure of narratives.

### *3.3.8 Conclusion and implications*

On the basis of the validation procedure of the three supposedly oral narrative tasks, it can be concluded that only two out of the three tasks are in fact oral narratives as only they can be said to elicit narratives validly. The single picture task is not valid as there is too much variability in the output produced by the respondents in response to it. Although according to the definition of a minimal narrative adopted from Labov (1972) the majority of these texts would qualify as narratives, they contain a lower ratio of narrative clauses; thus, their quality is poorer as compared to the texts produced under the other two conditions. The picture sequence and the cartoon strip tasks on the other hand are valid oral narrative tasks; they result in balanced performance and a relatively higher ratio of narrative clauses. Despite these favourable results a slight modification of the picture sequence task seemed necessary because of problems with task content. The description of the modified version of the picture sequence tasks can be found in 4.5.2.

# Chapter 4: Method

## 4.1 Introduction

This chapter of the dissertation discusses the methods employed when conducting the empirical study. First the definitions of constructs under examination are provided, which is followed by the presentation of the research design, the participants, and the instruments applied. The subsequent sections contain description of the procedures, the steps of data analysis, and a detailed description of the measures and statistical procedures employed.

## 4.2 Definition of constructs

The most important constructs of my research are creativity and the oral narrative task, both of which are concepts that are interpreted in several ways. Therefore, I feel it is important to provide a working definition of these constructs, in order to clarify how they were used in my research.

### *4.2.1 Creativity*

Since the construct of creativity seems to be too complex and therefore quite impossible to define for research purposes, I opted for an operational definition of creativity. Therefore, in my dissertation, creativity was defined as the person's ability to come up with a large number of novel and statistically rare solutions on a given task and was operationalised as the total score achieved on a standardised creativity test (Barkóczi, & Zétényi, 1981). Also, since the test measures three distinct facets of creativity: *average originality*, the ratio of total originality and total creative fluency scores measuring the ability to produce unusual ideas; *relative flexibility*, the ratio of total flexibility and total creative fluency measuring the ability to produce a wide variety of ideas; and *creative fluency*, the total number of responses measuring the ability to produce a large number of ideas (Baer, 1993), these sub-scores were also

used to identify individuals creative in one of these sub-fields (see a more detailed description of the measures used in section 4.7.1).

### *4.2.2 Oral narrative task*

In my dissertation I adopted Skehan's (1996) comprehensive definition of tasks which states that a task is "an activity in which: meaning is primary; there is some relationship to the real world; task completion has some priority; and the assessment of task performance is in terms of task outcome" (p. 38). In line with this definition, successful task completion involved telling a story on the basis of the pictures, where a story consisted of at least two temporally ordered narrative clauses, and it contained at least four of the six stages of a fully formed narrative (Labov, 1972): the orientation, the complicating action, the evaluation, and the resolution.

## 4.3 Design

The empirical research carried out for the dissertation was quantitative, and its design was correlational. In a correlational research design, existing relationships between variables are examined, but the presence of a relationship does not imply that it is a causal relationship. In order to establish causality, conducting an experiment would have been needed (Hatch & Lazaraton, 1991; Seliger & Shohamy, 1989). This, however, would have been unfeasible with these variables.

## 4.4 Participants

Participants of the research were first-year English major students who attended language practice or academic skills classes. This choice of participants was feasible for three reasons. First, since it was hypothesised that creativity manifests itself in the language the participants use for solving the narrative tasks, more precisely in lexical diversity, I assumed that a relatively high level of proficiency is required to demonstrate this effect. Also at the time

when the study was conducted, first-year English majors at ELTE were required to pass a proficiency exam at the end of the second semester, which contained a task that is quite similar to one of the oral narrative tasks used in the study. Therefore, the findings of my research study would have born direct relevance for the exam. The third reason for choosing this population was that since I taught first-year students, they were easily accessible for me. The fact that I needed a large amount of data from each individual made accessibility an important consideration, as well.

As I did not expect gender and age to influence the phenomenon under investigation, I did not set any constraints with regard to these variables. I assumed that the ideal sample size would be around forty, since the time-consuming and labour-intensive analysis of the narrative tasks and the creativity test does not make a bigger sample feasible. Moreover, investigations reported in various journals usually had thirty to forty participants (Foster and Skehan, 1996; Robinson, 1995; Skehan and Foster, 1997, 1999) which might be due to the fact that some statistical procedures, for example correlational analysis, can already be performed on a sample of this size and the results are, to some extent, generalizable.

Keeping the above considerations in mind, the final research sample consisted of 41 English majors, 11 males and 30 females, who were in the first year of their studies at the School of English and American Studies at ELTE. Their age ranged from 18 to 23; the majority, that is 17 participants, being at the age of 19. 12 students attended a language practice course and 21 participants an academic skills course instructed by the researcher. The remaining eight students volunteered to take part in the study from the group of a colleague instructing a language practice seminar. Students provided data on a voluntary basis and they were given no financial reward for their contribution, but they were informed about their test results.

## 4.5 Instruments

This section presents the instruments used for the empirical study. The standardised test of creativity, the oral narrative tasks, the two tests of language

proficiency a (C-test and a TOEFL test), and the Hungarian test of language aptitude (HUNLAT) are discussed below.

### *4.5.1 Creativity test*

The standardised creativity test (Barkóczi & Zétényi, 1981) (see Appendix I) used in the study 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 and last evaluated tasks of the test are verbal tasks, they 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 "Remote Associations" task (in a similar fashion to Mednick's (1962) Remote Associates Test), students 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 second and third tasks of the test are figural, and they require responses in drawing (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 ("Circles" task); and to finish abstract shapes in a creative manner ("Picture Completion" task). The four tasks last for five, eight, ten, and six minutes respectively.

### *4.5.2 Oral narrative tasks*

The two variants of oral narrative tasks used for the purpose of data collection were the ones that proved to elicit narratives in the validation study discussed in section 3.3, that is the "cartoon strip" and the "picture sequence" tasks. Although the cartoon strip task was exactly the same as the one used in the pilot study (see Appendices E-F), the picture sequence task needed to be changed. The reason for this is that the analysis and the post-task interviews revealed that task content is a decisive factor regarding the validity of oral

narrative tasks. Since one sequence of pictures was problematic in this respect because it resulted in sketchy stories with hardly any evaluation, it became obvious that these pictures would have to be changed.

As a follow-up to the validation study, some new pictures were piloted for the picture sequence task (see Appendices J-K), which led to a slight modification of the task itself. I attempted to make the picture sequence task less similar to the cartoon strip task, and wanted to give more freedom to students to manifest their creativity. As a consequence, the new pictures were not connected in the sense that they had no common protagonist as in the previous version of the task. They only showed six simple drawings of some story ingredients: one object, one means of transport, one natural phenomenon, and three scenes, either natural or built. The participants were not given pictures of protagonists in order to allow them to use their imagination freely. The analysis of the stories and the following post-task interviews showed that these pictures elicited narratives validly. Also, since the pictures were basically unconnected still they showed essential elements of narratives, the issue of problematic content could be avoided, as well.

The resulting two narrative tasks differ in their cognitive complexity; therefore, their structure imposes different information processing demands on the participants. The picture sequence task being less specific and structured probably requires more on-line processing; thus, it is more complex (Skehan, 1998). On the basis of Robinson's (2001b, 2005b, 2007a) Triadic Componential Framework of task characteristics, this increase in cognitive complexity was brought about by resource dispersing variables, that is, the lack of inherent task structure and an increase in the number of tasks to be carried out (inventing and telling the story).

### *4.5.3 Proficiency tests (TOEFL-PBT, C-Test)*

Two tests were used for measuring language proficiency. One of them is a C-test validated for Hungarian learners of English (Dörnyei & Katona, 1992) (see Appendix L), which is the same C-test that was used for measuring language proficiency for my MA thesis. Although for my purposes it would

have been very useful to measure oral skills with an oral test, the Test of Spoken English (TSE) for example, this was not feasible since the testing of oral skills cannot be performed validly and reliably by one researcher. Since this is a problem with all oral tests, I decided to use a proficiency test that has no compulsory oral component, is widely used, and is relatively easy to evaluate. Therefore, I administered a paper based Test Of English as a Foreign Language (TOEFL-PBT) practice test (Phillips, 1989) to the participants since I wanted to use a more comprehensive test of language proficiency than the C-test. The TOEFL-PBT consists of three parts: listening comprehension, structure and written expression, and vocabulary and reading comprehension. The test format of the 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.

The rationale behind using both a C-test and a TOEFL-PBT test is that their relationship with creativity would help me draw conclusions concerning whether there is in fact a relationship between creativity and language proficiency, or, as it was suggested, it is only C-tests that favour creative students. The third possibility is that creativity and language proficiency are unrelated, and the tendency-level relationship found in my MA thesis was only by chance.

### *4.5.4 Language aptitude test (HUNLAT)*

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 [Hungarian Language Aptitude Test] (HUNLAT) 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. Sample tasks from HUNLAT can be found in Appendix M.

## 4.6 Procedures

Data collection was carried out throughout the academic year 2001/2002 at the School of English and American Studies, at ELTE. Participants were asked to fill in the creativity test (Barkóczi & Zétényi, 1981), the TOEFL test (Phillips, 1989), the C-test (Dörnyei & Katona, 1992) and the language aptitude test (Ottó, 2002) in the presence of the researcher. The reason for this is that all of the above tests are timed, and reflect the knowledge or skills of the individual; therefore, the presence of an invigilator is highly recommended. In order to simulate classroom-like and exam-like conditions, the oral narrative tasks were performed in pairs, where each member of the pair was randomly assigned one of the "cartoon strip" and one of the "picture sequence" tasks. Students were allowed to spend five minutes with preparation, after which they told their story to their partner. Then they changed roles, and the listener became the speaker. Each participant told two stories altogether, which were tape-recorded by the researcher.

## 4.7 Analysis

The following section contains the steps of data analysis and the description of the resulting data. The analysis of the creativity test is discussed first simultaneously with the type of data gained; this section is followed by the detailed analysis and measures of the narrative tasks. Finally, measures of language proficiency and language aptitude are summarised.

### *4.7.1 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) (this creativity test 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. Barkóczi and Zétényi suggest that the resulting raw scores should be converted to a standardised T-profile, but as the conversion of scores using the figures of the test booklet was judged to be rather imprecise, I decided to use standard scores, Z-scores, for further calculations. Z scores indicate distance from the mean in terms of standard deviations; therefore, they are directly comparable to one another considering their relative location in their respective distributions (Salkind, 2004).

The *fluency* score, which in this dissertation is called *creative fluency* in order to differentiate it from the temporal variable also called fluency, equals the number of responses given on a given task, while the *flexibility* score reflects the number of categories the subjects select their answers from on each task (the categories were set up in the course of the standardisation procedure by Barkóczi and Zétényi, 1981). The *originality* score was 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. *Verbal creativity* can be determined by adding up the creative fluency, flexibility and originality scores on the two verbal tasks, while *figural creativity* is calculated in the same way for the two drawing tasks. Apart from these total scores, 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.

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. The

high correlations between the fluency, originality and flexibility sub-scores are the reason why some authors (Hargreaves & Bolton, 1972) argued for dropping the originality and flexibility scores altogether since in this form they provide little additional information. I also believe that this scoring system is biased, since for example, if a person produces two highly original ideas worth the maximum score 0.99 point each, his or her total originality score will be 1.98 points for the given task. If, however, another respondent produces five statistically more common responses worth 0.50 point each, his or her originality score will be higher (2.5 points), than that of his or her less fluent peer, and in this way a misleading picture is gained about the two individuals' true originality. Since a similar scoring method is applied for the flexibility scores (each new category is rewarded with 1 point without considering the total number or responses), these scores can be said to be affected by creative fluency to a great extent, as well.

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 flexibility and creative fluency) and *average originality* (the ratio of originality and creative fluency) should also be calculated, in line with the procedure specified in the test (Barkóczi & Zétényi, 1981). This way the *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. Measures of creativity employed in the study are summarised in Table 7.

Table 7
*Measures of creativity*

| Measures | Description |
|---|---|
| Originality | the sum of originality scores received on each task |
| Flexibility | the sum of flexibility scores received on each task |
| Creative Fluency | the sum of responses on each task |
| Average Originality | the ratio of originality and fluency scores (can be calculated on each task and for the whole test as well) |
| Relative Flexibility | the ratio of flexibility and total fluency scores (can be calculated on each task and for the whole test as well) |
| Total Originality | the sum of originality scores received on the four tasks added up for the whole test |
| Total Flexibility | the sum of flexibility scores received on the four tasks added up for the whole test |
| Total Creative fluency | the sum of responses given by the respondent on the four sub-tasks |
| Verbal Creativity | the sum of originality, flexibility and fluency scores on the two verbal tasks |
| Figural Creativity | the sum of originality, flexibility and fluency scores on the two drawing tasks |
| Total Creativity score | the sum of total originality, total flexibility and total fluency scores |

### *4.7.2 Measures of task performance*

The respondents' audio-recorded performance on the narrative tasks was first transcribed, and these transcripts were used for analysis later on. I measured six aspects of the respondents' performance: *accuracy*, *syntactic complexity*, *lexical complexity/variety*, *fluency*, *quantity of talk*, and *narrative structure*. As it was argued in section 2.3.6 of the dissertation, accuracy, complexity (syntactic and lexical) and fluency are widely used measures of task performance (see for example Foster & Skehan, 1996; Iwashita et al.,

2001; Robinson, 2001, 2007b; Skehan, 2001; Skehan & Foster, 1997, 1999). Since there seems to be a consensus regarding the measurement of accuracy and syntactic complexity in the research articles consulted, I adopted the way these were operationalised there with slight modifications. Therefore, I calculated *accuracy* as the proportion of error-free clauses relative to the total number of clauses. *Syntactic complexity,* however, was measured by the ratio of the total number of clauses to the total number of analysis of speech units (AS-units), which is different from what is suggested by Foster and Skehan (1996), as they argued for using c-units as the unit of measurement. The reason why I decided to apply AS-units, where an AS-unit is defined as "single speaker's utterance consisting of an independent clause or sub-causal unit, together with any subordinate clause(s) associated with either" (Foster et al., 2000, p. 365) is because AS units were designed specifically for handling spoken data.

In contrast to accuracy and syntactic complexity, operationalising fluency and lexical complexity/variety was much less straightforward. Although fluency was operationalised in several different ways in the research articles consulted, for example as the number of pauses, the amount of silence, the number of repetition, false starts, reformulations and replacement and speech rate (Foster & Skehan, 1999; Robinson, 2007b), I opted for measuring *fluency* as speech rate, as besides being feasible, it was found to be a reliable measure of fluency in a number of studies (Ejzenberg, 2000; Freed, 2000; Kormos, & Dénes, 2004; Lennon, 1990; Riggenbach, 1991). Based on Riggenbach's (1991) recommendations, fluency was calculated as the total number of syllables produced by the participant divided by the total amount of time, including pause time, required to produce the text. Regarding the issue of *lexical complexity/variety*, the problem is caused by the fact that not many researchers measured this aspect of task performance. Those who did (e.g. Robinson, 2001b) used the type-token ration (TTR) for this purpose, which is a measure reflecting the total number of different words (types) divided by the total number of words (tokens) produced.

The problem with the TTR is, however, that it depends to a great extent on the sample size, that is, on the number of words spoken by the participants.

It is easy to see that longer texts usually involve some repetition which is a natural process and should not be interpreted as a signal of deteriorating lexical complexity/diversity. Richards (1987) found that the "type-token ratio falls rapidly as the number of tokens increases" (p. 205). In a recent study Jarvis (2002) found that two formulas based on the type-token ratio: Dugast's (as cited in Jarvis, 2002) Uber U formula and Malvern's and Richards' (1997) D-formula can be used to measure lexical richness in L2 texts in a reliable way. Both measures can, however, only be used with a curve-fitting approach. Because there exists an available software only for the D-formula (VOCD is available at the CHILDES web-site: http//childes.psy.cmu.edu), I applied this software to establish a measure of lexical diversity. The calculation of the D-value is based on a mathematical probabilistic model, and the software uses random sampling of tokens in plotting the curve of TTR against increasing token size for the text to be investigated. Malvern and Richards argue that the D-value is a valid measure of diversity because it does not depend on the length of the sample, and it uses all the words produced by the participants.

A different aspect of *lexical complexity/variety* is captured if we consider whether the words comprising the stories are among the most frequently used English words or they are relatively rare. PLex (Meara, 2001, available at http://www.swan.ac.uk/) is a computer program designed to capture this aspect of lexical complexity/variety of texts, and as such it bears some resemblance to the Lexical Frequency Profile (LFP) (Laufer & Nation, 1995) although unlike LFP, Plex is claimed to be able to handle relatively short texts as well (Meara & Bell, 2001). Plex operates in a way that it divides the text into segments of ten words, then counts the number of 'difficult words', that is the words not included in the list of the 1.000 most frequent English content words, in the segments. It then calculates a figure, the plex lambda, indicating the likelihood of the occurrence of difficult words; the higher the figure the more likely the use of rare words. Although to the best of my knowledge the use of this type of measure of lexical complexity/variety is unprecedented in task-based research, it has been successfully used in studies investigating the lexical richness of spoken and written texts (Espinosa, 2005; Meara & Bell, 2001; Read, 2005).

The fifth measure I used is one that seemed to be connected to creativity on the basis of my MA thesis, and it is the *quantity of talk*. The quantity of talk students produce was measured by the total number of words (see also Dewaele & Pavlenko, 2003; Dörnyei & Kormos, 2000). The last measure applied is a non-linguistic one. It provides information about the contents of the stories, in the sense that it attempts to establish the event structure of stories. The analysis of the *narrative structure* of the students' texts was based on Labov's (1972) classification of the elements of a narrative; it is in fact the same framework I used for the validation study of narrative tasks (see the detailed description in section 3.3.4). According to Labov, the skeleton of any narrative consists of narrative clauses, which are temporarily ordered independent clauses connected by temporal junctures. Therefore, the texts elicited with the help of the tasks were analysed for the number of narrative clauses and also for the number of narrative clauses per AS-unit. Narrative clauses reflect the events of the story; thus, a high percentage of such clauses can be assumed to signal complex stories as far as their event structure is concerned. Measures of task performance used in the study are presented in Table 8.

Table 8
*Measures of task performance*

| Measures | Description |
|---|---|
| *Accuracy* | |
| Correct clauses per clauses | The number of grammatically correct clauses divided by the total number of clauses |
| *Syntactic complexity* | |
| Number of clauses per AS-unit | The total number of clauses divided by the total number of AS-units |

| Measures | Description |
|---|---|
| *Lexical complexity/variety* | |
| D-index | A value calculated with the help of a mathematical probabilistic model that plots the curve of type-token ratio against increasing token size |
| Plex lambda | A value showing the likelihood of the occurrence of less frequent (not included in a list of the 1.000 most frequent words) words in the text |
| *Fluency* | |
| Speech rate | The total number of syllables produced divided by the total amount of time required to produce them, including pause time, expressed in seconds |
| *Quantity of talk* | |
| Number of words | Total number of words produced in English |
| *Narrative structure* | |
| Number of narrative clauses | The total number of temporally ordered independent clauses connected by temporal junctures (Labov, 1972) |
| Ratio of narrative clauses per AS-unit | The total number of temporally ordered independent clauses connected by temporal junctures (Labov, 1972) divided by the total number of AS-units |

### *4.7.3 Measures of proficiency*

The three texts comprising the C-test contain 63 gaps altogether. The scoring of the C-test meant adding up the number of correct solutions. The evaluation of the TOEFL-PBT test was carried out according to the procedures specified in the practice test booklet (Phillips, 1989). After adding up the

number of correct solutions, 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 scale, thus enabling the computation of a composite TOEFL-PBT score.

### *4.7.4 Measures of language aptitude*

Since the aptitude test used in this study is protected by copyright, the participants' answers, recorded on answer sheets, were analysed by Ottó (2002), the author of the aptitude test. Having computed the scores, he sent the results to me. This way all 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.

## 4.8 Statistical analysis

Data deriving from the analytical procedures were analysed using the software SPSS 11.0 for Windows. With the help of this software, descriptive statistics (mean, standard deviation) were calculated to shed light on characteristics of the sample, whereas correlation coefficients were used to describe the features of instruments. Independent samples t-tests were used for examining the two versions of the two oral narrative task variants, and paired samples t-test were employed to compare participants' performance on the two oral narrative tasks differing in cognitive complexity. In order to detect relationships between variables, correlations were calculated between ID variables, creativity and language aptitude, language proficiency and task performance measures. Unfortunately, the relatively small sample size did not permit the use of more sophisticated statistical procedures.

# Chapter 5: Findings Concerning Individual Differences in Abilities and Proficiency

## 5. 1 Introduction

This section of the dissertation presents those results of the study which concern participants' language proficiency and their individual differences in abilities, such as creativity and language aptitude. In an attempt to provide a detailed picture of the research sample and shed light on the characteristics of the measuring instruments, first descriptive statistics for each of the variables are provided, which is followed by different correlational analyses. The first variable examined is creativity; means and standard deviations of the sample are compared to the national standard provided for high school graduates. Besides intercorrelations of the different measures of creativity for the whole test, a multitrait - multimethod analysis of the creativity test across the four subtasks is also provided. Descriptive statistics and intercorrelations of the different measures are presented both for the language aptitude test (HUNLAT) and for TOEFL test scores. In the case of HUNLAT, scores awarded to members of the sample are compared to results of first year university students and first year university students with at least one language exam (Ottó & Nikolov, 2003). The presentation of the results is then followed by a detailed discussion of the findings, in which characteristics of the research sample and the tests used to assess individual differences are analysed in detail.

## 5.2 Results

### *5.2.1 Creativity test*

If we examine the mean and standard deviation figures of the individual variable of creativity (see Tables 9 and 10), they show that the means of the 1st year language major sample tend to be considerably higher than the corresponding means of the national standard for high school graduates (Barkóczi & Zétényi, 1981). The only exception is relative flexibility since the mean value on the remote association task is $M$=0.69 for the research sample,

whereas the national standard is *M*=0.73. Relative flexibility on the unusual uses and circles tasks is similar for the research and the national sample, *M*=0.78 and *M*=0.61 respectively. Although standard deviation scores also tend to be somewhat higher for the research sample, this change is not as salient as the change in the average figures. The reason why scores achieved on the different subtests and not composite scores are used in Tables 9 and 10 is that only these figures were available for the national sample (Barkóczi & Zétényi, 1981).

Table 9
*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 | | Remote Associations | | Circles | | Picture Completion | |
|---|---|---|---|---|---|---|---|---|
| Variable | mean | SD | mean | SD | mean | SD | mean | SD |
| 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 10
*Descriptive statistics of the four subtests of the standardised creativity test for a representative sample of high school graduates (N=1,098), the Hungarian national standard*

| Title of subtest | Unusual Uses | | Remote Associations | | Circles | | Picture Completion | |
|---|---|---|---|---|---|---|---|---|
| Variable | mean | SD. | mean | SD. | mean | SD | mean | SD. |
| 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 |

Note. From "*A kreativitás vizsgálata"* [The examination of creativity]. by I. Barkóczi and T. Zétényi, 1981, Budapest: Országos Pedagógiai Intézet (p. 32).

Table 11 presents intercorrelations of the composite creativity scores. Total creativity being a sum of the originality, creative fluency and flexibility scores on the four different tasks has very high, significant correlations with all the other scores except for average originality ($r_s$=0.26). Verbal creativity, which is the sum of the originality, creative fluency and flexibility scores on the two verbal tasks, and figural creativity, which is the sum of the originality, creative fluency and flexibility scores on the two drawing tasks, do not correlate significantly with average originality ($r_s$=0.21 and $r_s$=0.24 respectively); moreover, figural creativity does not correlate with relative flexibility ($r_s$=0.18). Verbal creativity and figural creativity are also unrelated ($r_s$=0.30). Out of the five remaining variables originality, creative fluency and flexibility are very strongly related, correlation coefficients range from $r_s$=0.86 to $r_s$=0.89; the other two variables, average originality and relative flexibility, have lower correlations. The only significant correlation of average originality is with the originality score ($r_s$=0.44), whereas relative flexibility correlates

significantly not only with flexibility ($r_s$=0.63), but also with originality ($r_s$=0.42) and creative fluency ($r_s$=0.31). Because of their high correlations with the creative fluency score and one another, the originality and flexibility scores that are not fluency free will not be used in further analyses.

Table 11
*Intercorrelations of composite measures of creativity*

| | Creative fluency | Flexibility | Average originality | Relative flexibility | Verbal creativity | Figural creativity | Total creativity |
|---|---|---|---|---|---|---|---|
| Originality | .89** | .86** | .44** | .42** | .74** | .77** | .95** |
| Creative fluency | | .89** | .12 | .31* | .74** | .75** | .95** |
| Flexibility | | | .20 | .63** | .77** | .73** | .95** |
| Average originality | | | | .22 | .21 | .24 | .26 |
| Relative flexibility | | | | | .63** | .18 | .48** |
| Verbal creativity | | | | | | .30 | .79** |
| Figural creativity | | | | | | | .77** |

* $p < .05$. ** $p < .01$.

The multitrait-multimethod matrix developed by Campbell and Fiske (1959) (see Table 12) is an approach used for assessing the construct validity of a set of measures in a study. Although in its original form the various traits are supposed to be tested using different methods (e.g. paper-pencil test, observation, etc.), this kind of analysis was already used in research on divergent thinking where the methods were substituted by the various tasks of the creativity test (Runco, 1986). Figures shaded in light grey show the validity diagonals, that is, correlations between measures of the same trait, in our case the same facet of creativity, using different methods, that is, different tasks. Since convergent validity is believed to contribute to construct validity, we

expect that different measures of a certain facet of creativity should have the highest correlations as the same concept is being assessed using different tasks. Figures shaded in dark grey show the so-called heterotrait-monomethod triangles; these are correlations among measures sharing the same method of measurement, that is, correlations between the various facets of creativity measured on the same task. If the method of measurement is influential, that is, there is a method effect, these correlations tend to be relatively high as well. The rest of the figures show heterotrait-heteromethod correlations, that is, these correlations share neither method nor trait. These are correlations of different facets of creativity measured by different tasks. Since these figures indicate discriminant validity, they are expected to be the lowest correlations in the matrix.

Table 12
*Multitrait - multimethod matrix of the creativity test*

| | | Unusual Uses | | | Remote Associations | | | Circles | | | Picture Completion | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A.o. | C.f. | R.f | A.o. | C.f. | R.f | A.o. | C.f. | R.f | A.o. | C.f. | R.f |
| U.U. | A.o. | - | | | | | | | | | | | |
| | C.f. | .27 | - | | | | | | | | | | |
| | R.f. | .46** | .72** | - | | | | | | | | | |
| R.A. | A.o. | -.09 | .10 | .13 | - | | | | | | | | |
| | C.f. | .37* | .47** | .27 | -.16 | - | | | | | | | |
| | R.f. | .26 | .19 | .19 | -.11 | .68** | - | | | | | | |
| C. | A.o. | .14 | .13 | .14 | .13 | -.03 | -.03 | - | | | | | |
| | C.f. | -.20 | .24 | .07 | .08 | .18 | -.06 | .04 | - | | | | |
| | R.f. | .07 | .10 | .18 | -.05 | -.01 | .14 | .56** | -.40** | - | | | |
| P.C. | A.o. | .35* | .23 | .39** | .17 | .01 | .01 | -.08 | -.06 | -.04 | - | | |
| | C.f. | -.07 | .13 | .06 | .30 | .12 | .22 | -.07 | .27 | -.13 | .02 | - | |
| | R.f. | .16 | -.01 | .11 | .18 | -.06 | .05 | .19 | -.01 | .22 | .10 | .21 | - |

*Note.* U.U. = Unusual Uses; R.A. = Remote Associations; C = Circles; P.C. = Picture completion; A.o. = Average originality; C.f. = Creative fluency; R.f. = Relative flexibility.

* $p < .05$. ** $p < .01$.

Figures in Table 12 show that the validity diagonals shaded in light grey do not contain the highest correlations in the matrix. Out of 18 correlations only two are statistically significant, these are between creative fluency in unusual uses and remote associations ($r_s$=0.47), and average originality in unusual uses and picture completion ($r_s$=0.35). The heterotrait-

monomethod triangles shaded in dark grey contain the highest correlation of the matrix, and the highest number of significant correlations is also found there. These five significant correlations are between average originality and relative flexibility, and creative fluency and relative flexibility in unusual uses; creative fluency and relative flexibility in remote associations; and between average originality and relative flexibility, and relative flexibility and creative fluency in circles. Although the heterotrait-heteromethod correlations tend to be low, there are two which are statistically significant: the correlation between average originality in unusual uses and creative fluency in remote associations $r_s$=0.37, and relative flexibility in unusual uses and average originality in picture completion $r_s$=0.39.

### *5.2.2 Aptitude test*

Table 13 presents descriptive statistics for the aptitude test (HUNLAT) (Ottó, 2002) for the population of my study, 1st year English majors at ELTE. In order to aid the interpretation of figures, Table 14 provides 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 (*SD*=6.68 for English majors and *SD*=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 (*SD*=10.24) is greater than that of language majors.

As the maximum score for each of the subtasks measuring components of language aptitude was 20, it can be concluded that rote learning ability measured by the vocabulary learning section (*M*=17.02) and inductive language learning ability as measured by the language analysis section (*M*=16.92) were those skills where members of the research sample scored the highest. These are the subtasks on which university students and university students with at least one language exam also performed the best, but for these

populations instead of the vocabulary learning section, the language analysis part (*M*=16.28 for university students, and *M*=17.54 for university students with at least one language exam) seems to result in the highest scores.

Table 13
*Descriptive statistics of the aptitude test HUNLAT for 1st year English majors at ELTE (N=41)*

| Variable | Mean | Standard deviation | Minimum | Maximum |
|---|---|---|---|---|
| Hidden Sounds | 15.87 | 2.27 | 9 | 18 |
| Language Analysis | 16.92 | 1.42 | 12 | 18 |
| Words in Sentences | 14.78 | 3.48 | 5 | 20 |
| Vocabulary Learning | 17.02 | 2.97 | 8 | 20 |
| Language Aptitude | 64.60 | 6.68 | 50 | 75 |

Table 14
*Descriptive statistics of the aptitude test HUNLAT for 1st year university students (N=130) and for 1st year university students with at least one language exam (N=65)*

| Variable | 1st year university students (*N*=130) | | 1st year university students with at least one language exam (*N*=65) | |
|---|---|---|---|---|
| | Mean | Standard deviation | Mean | Standard deviation |
| Hidden Sounds | 13.53 | 3.89 | 14.98 | 3.62 |
| Language Analysis | 16.28 | 3.70 | 17.54 | 3.34 |
| Words in Sentences | 10.95 | 4.40 | 12.48 | 4.12 |
| Vocabulary Learning | 15.04 | 4.31 | 16.03 | 4.2 |
| Language Aptitude | 55.79 | 11.61 | 61.03 | 10.24 |

*Note.* From "Magyar felsőoktatási intézmények elsőéves hallgatóinak nyelvérzéke [The language aptitude of first year college students in Hungary]," by I. Ottó and M. Nikolov, 2003, *Iskolakultúra, 13,* (6-7), pp.39-40.

When examining the intercorrelations of the subtasks of HUNLAT in Table 15, it can be seen that only two, low but statistically significant, correlations can be found between the subtasks; the Hidden Sounds scores correlate positively with Language Analysis scores ($r_s$=0.33) and Words in Sentences ($r_s$=0.34) scores. The rest of the subtasks are unrelated, they only correlate significantly with the total score.

Table 15
*Intercorrelations of the aptitude test HUNLAT*

| | Language Analysis | Words in Sentences | Vocabulary Learning | Total Language Aptitude score |
|---|---|---|---|---|
| Hidden Sounds | .33* | .34* | .11 | .64** |
| Language Analysis | | .14 | .15 | .47** |
| Words in Sentences | | | .21 | .76** |
| Vocabulary Learning | | | | .63** |

* $p < .05$. ** $p < .01$.

### *5.2.3 Proficiency tests*

Table 16 presents the participants' language proficiency measures: the scores they received on the TOEFL-PBT, and their scores on a standardised C-test. The maximum scores that can be achieved on the Listening Comprehension and the Structure and Written Expression sections are 68 points, and it is 67 points for the Vocabulary and Reading Comprehension section. The maximum total score on a TOEFL-PBT test is 677 points. Members of the research sample scored highest on the Structure and Written Expression section and lowest on the Listening Comprehension section, their

mean total score was 571.31 points. The highest possible score for the C-test was 63 point, and the mean score of the participants was 48.46 on this test.

Table 16
*Descriptive statistics of language proficiency tests for 1st year English majors at ELTE (N=41)*

| Variable | Mean | Standard deviation | Minimum | Maximum |
|---|---|---|---|---|
| TOEFL Listening Comprehension | 56.00 | 5.12 | 42 | 64 |
| TOEFL Structure and Written Expression | 59.19 | 5.65 | 46 | 68 |
| TOEFL Vocabulary and Reading Comprehension | 56.48 | 4.73 | 44 | 63 |
| Total TOEFL score | 571.31 | 46.61 | 450 | 643 |
| C-test score | 48.46 | 6.05 | 32 | 60 |

Intercorrelations of the subtasks of the TOEFL-PBT and the C-test can be found in Table 17. The intercorrelations between the three subtasks of the TOEFL-PBT are all statistically significant and rather high. Listening Comprehension positively correlates with Structure and Written Expression ($r_s$=0.56), and also with Vocabulary and Reading Comprehension ($r_s$=0.71). These latter two subtests also correlate positively ($r_s$=0.67), and obviously, all of the subtasks correlate with the total score. The other proficiency test score, the C-test score also correlates positively and significantly with the subtests, and with the total TOEFL-PBT score as well ($r_s$=0.77).

Table 17
*Intercorrelations of proficiency test scores*

| Variable | TOEFL Structure and Written Expression | TOEFL Vocabulary and Reading Comprehension | Total TOEFL score | C-test score |
|---|---|---|---|---|
| TOEFL Listening Comprehension | .56** | .71** | .84** | .65** |
| TOEFL Structure and Written Expression | | .67** | .87** | .71** |
| TOEFL Vocabulary and Reading Comprehension | | | .88** | .62** |
| Total TOEFL score | | | | .77** |

* $p < .05$. ** $p < .01$.

## 5.3 Discussion

### *5.3.1 Creativity*

If we examine the measures of the ID variable creativity (see Tables 9 and 10), 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 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. 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.

Intercorrelations of the creativity test scores (see Table 11) suggest that figural and verbal creativity seem to exist independently as their correlation is

statistically not significant. Having examined the correlation matrix, we can also conclude that creative fluency scores determine originality and flexibility scores to a great extent, which is obvious from the very high level of correlation that can be found between them. Calculating fluency-free scores, that is average originality and relative flexibility was partly successful as an attempt at establishing two further independent measures of creativity. Average originality seems to meet the criterion of being independent as it has no significant correlations with either creative fluency or relative flexibility. Relative flexibility, however, is significantly correlated with creative fluency although this correlation is rather low.

The multitrait-multimethod matrix (see Table 12) seems to reflect some problems with the measure of creativity used in the study, which is unfortunately in line with what Runco (1986) found in connection with other divergent thinking tests consisting of several tasks. Assuming that the different facets of creativity measured by the creativity test are different traits and hypothesising that the different tasks represent different methods for measuring these traits, the resulting matrix should have been very different. The highest correlations should have been located on the validity diagonals shaded in light grey, with moderately high correlations in the heterotrait-monomethod triangles shaded in dark grey. The remaining heterotrait-heteromethod correlations should have been the lowest in the matrix. Contrary to these expectations, however, the highest and the highest number of correlations were found in two of the heterotrait-monomethod triangles, suggesting that there might be a method effect for the Unusual Uses and Circles tasks. The fact that there were only two statistically significant correlations on the validity diagonals, and there were also two statistically significant correlations among the heterotrait-heteromethod correlations clearly shows that these four tasks did not really measure the same traits, that is, the same facets of creativity. If they had been measuring the same traits with different methods, the resulting matrix would have shown the above described pattern as a result of convergent and discriminant validity.

There are two possible interpretations of this result: it can either be argued that the lack of discriminant validity reflects a problem of with the

constructs or with their operationalization. Since the theoretical constructs: fluency, originality and flexibility have a long history as facets or creativity and their existence is also supported by empirical research (Runco, 1985; Runco & Albert, 1985; Runco, Okuda, & Thurston, 1987), it is more likely that the problem concerns their operationalization. There are a number of things that might be flawed: the method of calculation itself, some or all of the creativity tasks, or it is also possible that the tasks can all be used to measure creativity, but they measure slightly different aspects of it. Although it would be quite difficult to tell which aspects of creativity are tapped by the individual tasks, the figural a drawing tasks clearly belong to different content areas or domains. In this sense, results of the multitrait-multimethod analysis might provide support for the content-specificity of creativity (Amabile, 1983, 1996; Csikszentmihalyi, 1988; Gardner, 1993; Runco, 1989; Sternberg & Lubart, 1995). At present I am not in a position to rule out any of these options. Consequently, the only justifiable course of action seems to be examining each task of the creativity test separately instead of using composite measures, which is the approach taken in chapters 7 and 8 of the dissertation.

### *5.3.2 Aptitude*

The descriptive statistics presented in Tables 13 and 14 about the language aptitude test used in the study reveal that the research sample, that is, the group of English majors, have the highest score when compared to university students, or university students with at least one language examination. Although the difference between their average and that of the latter group is not high, standard deviation figures indicate that the group of language majors is much more homogeneous as far as language aptitude is concerned. It seems, therefore, that the group of 1st year English majors at ELTE can be characterised by a high level of language aptitude, and this group is quite homogeneous with regard to this individual variable.

When examining the particular subskills of language aptitude, it seems that inductive language learning ability as measured by the Language Analysis subtest is the skill which has the highest average among 1st year university students and 1st year university students with at least one language

examination, and it is the second highest among 1st year English majors. This finding might be a consequence of the fact that university students are generally believed to have good analytical skills, as these skills are known to be advantageous in formal education. Furthermore, language majors and also probably students with at least one language exam are likely to have studied grammar explicitly, so because of their high metalinguistic awareness they probably know how to infer grammatical rules from the language input. The other high-scoring task is the Vocabulary Learning subtest measuring rote learning ability; this is the strongest subskill of the English major research sample. Since rote learning probably also has some importance in higher education, especially when learning a language, for example when learning new words, this finding should not be surprising.

When Carroll and Sapon (1959) devised the MLAT, their intention was to create independent indicators of language aptitude. Since the HUNLAT (Ottó, 2002) rests on the same theoretical basis as the MLAT, and even some tasks are similar, it is expected that the skills underlying language aptitude be independent. Therefore, the scores gained on the tasks measuring them should have no significant correlations with one another. This criterion can be said to be partly met by HUNLAT, as out of the six possible correlations among the tasks only 2 are statistically significant, and they are relatively low.

### *5.3.3 Proficiency*

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, standard deviation figures 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. Having examined the means of different sub-tests, it can be concluded that the Structure and Written Expression part proved to be the easiest for the research sample, whereas the Listening Comprehension part was the most difficult. Since the participants were language majors, a

heightened awareness for grammar is understandable, although this result might also indicate that grammar is probably still emphasised at the expense of communication, as is can be seen from the listening scores, in our schools.

The high positive correlations between the C-test and the TOEFL-PBT scores lend support to the hypothesis that the C-test used in the study is a reliable measure of proficiency as well. The very high intercorrelations between the different parts of the TOEFL-PBT suggest that they might not be measuring independent skills; therefore it might be advisable to consider the composite scores only.

## 5.4 Conclusion

The aim of this section of the dissertation was to provide a detailed picture of the sample examined and the research instruments used. With regard to the research instruments, the following can be concluded: since based on the results of the multitrait-multimethod analysis the different tasks of the creativity test do not seem to measure the same construct, the use of composite scores is not justifiable. Instead of them subscores of the separate tasks will be used in subsequent chapters of the dissertation. A different problem arose in connection with the TOEFL-PBT scores where subscores of the test seem to correlate too highly, which makes it doubtful whether they indeed measure independent skills. HUNLAT seems to be the least problematic among the tests used as the subscores of the different skills measured are more or less independent.

As far as the characteristics of the research sample are concerned, it seems that the 1st year English major population from ELTE exhibits almost uniformly high language aptitude with English proficiency levels ranging from intermediate to advanced. It can be hypothesised that this seeming discrepancy might be ascribed to differences in the time devoted to learning English. Although the sample on average seems to be more creative than the national standard, distribution figures show that low and high creativity individuals can be found among them as well, which is important in a study that attempts to examine the relationship of creativity and oral task performance.

# Chapter 6: Task Performance on Cognitively Less and More Complex Tasks

## 6.1 Introduction

This part of the dissertation presents findings concerning the participants' task-performance on the oral narrative tasks differing in cognitive complexity. Altogether four different oral narrative tasks were used in the study, supposedly two cognitively less and two cognitively more complex ones. In order to analyse whether the tasks used were of two different types or not, descriptive statistics for the two different versions of each task were presented and means were compared with the help of independent *t* tests. Since the research design necessitated the use of four tasks, it was important to establish whether these can be considered as variants of two narrative task types, cognitively less and more complex, or not.

Then in order to characterise the cognitively less and more complex tasks independently, intercorrelations were calculated for output measures such as accuracy, fluency, complexity, quantity of talk, lexical variety and the number and ratio of narrative clauses on each of the tasks separately. In order to answer the research question about how the cognitively less and more complex tasks differ, output measures on the two tasks were compared with the help of paired samples *t* tests. The Cognition Hypothesis put forward by Robinson (2001c, 2003, 2005b) suggests that differences in task complexity result in differential task performance; therefore, the cognitive complexity of tasks should be considered when making pedagogical decisions. Findings of the study might lend support to this hypothesis and enhance a more conscious way of using tasks.

## 6.2 Results

### *6.2.1 Oral narrative task variants*

Task 1, that is the cartoon strip task, was the cognitively less complex oral narrative task used in the study; versions "a" and "b" (see Appendices E-F) involved telling two completely different stories. 21 students were given version "a", the shipwrecked man story, whereas 20 students received version "b", the girl and wizard story. In order to decide whether we can consider the two versions as variants of the same task, the means of various output measures were compared using independent samples *t* test. Before performing the *t* test, there is a need to check distributions of the variables using an *F* test. Since the *F* test did not produce significant results in either of the cases thus indicating the comparability of distributions, the *t* tests could safely be performed. Table 18 reveals that there were no statistically significant differences between the two versions of the task for any of the output measures examined.

Table 18
*Independent samples t test for Task 1 versions a and b*

| | N | | mean | | SD | | mean diff. | F | Sig | t | df | Sig (2-tailed) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | a | b | a | b | a | b | | | | | | |
| Accuracy | 21 | 20 | .83 | .84 | .12 | .10 | -.01 | .42 | .51 | -.45 | 39 | .65 |
| Fluency | 21 | 20 | 127.12 | 138.42 | 31.28 | 31.62 | -11.30 | .195 | .66 | -1.15 | 39 | .25 |
| Syntactic complexity | 21 | 20 | 1.49 | 1.61 | .30 | .23 | -.12 | .173 | .68 | -1.44 | 39 | .157 |
| Quantity of talk | 21 | 20 | 142.19 | 166.70 | 35.81 | 64.03 | -24.50 | 3.46 | .07 | -1.52 | 39 | .13 |
| Lexical variety (d index) | 21 | 20 | 52.16 | 54.18 | 14.08 | 15.10 | -2.01 | .012 | .91 | -.44 | 39 | .66 |
| Lexical variety (plex lambda) | 21 | 20 | 1.03 | 1.08 | .29 | .30 | -.05 | .07 | .78 | -.62 | 39 | .53 |
| Number of narrative clauses | 21 | 20 | 8.14 | 10 | 2.81 | 4.21 | -1.85 | .44 | .50 | -1.66 | 39 | .10 |
| Ratio of narrative clauses | 21 | 20 | .62 | .69 | .16 | .12 | -.07 | 1.79 | .18 | -1.65 | 39 | .10 |

The "a" and "b" versions of Task 2, the cognitively more complex version of the oral narrative task used in the study, involved creating stories on the basis of two different sets of pictures (see Appendices I-J). 21 students performed task "a" while 20 students did task "b". As it can be seen from Table 19, the equality of variances can be assumed on the basis of the *F* test; therefore, the independent *t* test can be performed. Similarly to Task 1, no statistically significant differences between means of the output variables could be detected.

Table 19
*Independent samples t test for Task 2 versions a and b*

| | N | | mean | | SD | | mean diff. | F | Sig | t | df | Sig (2-tailed) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | a | b | a | b | a | b | | | | | | |
| Accuracy | 21 | 20 | .86 | .88 | .10 | .09 | -.01 | .24 | .62 | -.52 | 39 | .60 |
| Fluency | 21 | 20 | 128.48 | 127.32 | 35.02 | 28.35 | 1.16 | .89 | .34 | .11 | 39 | .90 |
| Syntactic complexity | 21 | 20 | 1.47 | 1.56 | .24 | .26 | -.08 | .01 | .95 | -1.12 | 39 | .27 |
| Quantity of talk | 21 | 20 | 212.14 | 266.70 | 136.41 | 142.52 | -54.55 | .65 | .42 | -1.25 | 39 | .21 |
| Lexical variety (d index) | 21 | 20 | 41.52 | 42.60 | 8.23 | 10.82 | -1.07 | 3.08 | .08 | -.36 | 39 | .72 |
| Lexical variety (plex lambda) | 21 | 20 | .85 | .80 | .26 | .34 | .05 | 1.51 | .22 | .62 | 39 | .53 |
| Number of narrative clauses | 21 | 20 | 14.28 | 16.40 | 7.97 | 7.33 | -2.11 | 0.01 | .92 | -.88 | 39 | .38 |
| Ratio of narrative clauses | 21 | 20 | .71 | .67 | .17 | .10 | .04 | 2.43 | .12 | .87 | 39 | .38 |

### *6.2.2 Characteristics of the oral narrative tasks differing in cognitive complexity*

Having seen that there are no statistically significant differences between versions "a" and "b" of Tasks 1 and 2, data from the two variants of the same task were merged. Table 20 contains data on the intercorrelations of the different output measures on Task 1. The correlation matrix shows that some output measures tend to change together, and there are also some independent ones. Fluency and one measure of lexical variety showing the ratio of rare words (plex lambda) seem to be independent measures as they have no significant correlations with any other measure. The ratio of narrative clauses

seems to be independent as well, as its only significant and relatively low correlation is with the number or narrative clauses ($r_s$=0.35), whereas syntactic complexity only moderately correlates with one measure of lexical variety showing the type token ratio (d index) ($r_s$=0.34). The remaining four measures tend to change together with three other measures each: accuracy correlates with the quantity of talk ($r_s$=.38), with the d index ($r_s$=0.55), and with the number of narrative clauses ($r_s$=0.31). Apart from accuracy, the quantity of talk also correlates with the d index ($r_s$=0.37) and with the number of narrative clauses ($r_s$=0.79).

Table 20
*Intercorrelations of measures of Task 1*

| | Fluency | Complexity | Quantity of talk | Lexical variety (d index) | Lexical variety (plex lambda) | Number of narrative clauses | Ratio of narrative clauses |
|---|---|---|---|---|---|---|---|
| Accuracy | .30 | -.04 | .38* | .55** | .29 | .31* | -.20 |
| Fluency | | .08 | .16 | .25 | -.07 | .02 | -.16 |
| Syntactic complexity | | | .22 | .34* | -.10 | .06 | .26 |
| Quantity of talk | | | | .37* | .21 | .79** | .02 |
| Lexical variety (d index) | | | | | .15 | .22 | -.28 |
| Lexical variety (plex lambda) | | | | | | .21 | .03 |
| Number of narrative clauses | | | | | | | .35* |

* $p < .05$. ** $p < .01$.

Table 21 presents the intercorrelations of output measures for Task 2. Interestingly, the number of significant correlations is much lower in this table; the majority of variables seem to be independent. It is only the quantity of talk that correlates with two other variables: similarly to the previous task, it has a high correlation with the number of narrative clauses ($r_s$=0.87), but unlike in the previous task it correlates with syntactic complexity ($r_s$=0.32). The last significant correlation in the table is between fluency and one measure of lexical variety, the d index ($r_s$=0.38).

Table 21
*Intercorrelations of measures of Task 2*

| | Fluency | Syntactic complexity | Quantity of talk | Lexical variety (d index) | Lexical variety (plex lambda) | Number of narrative clauses | Ratio of narrative clauses |
|---|---|---|---|---|---|---|---|
| Accuracy | .23 | -.05 | .07 | .19 | -.01 | .06 | -.18 |
| Fluency | | .20 | .18 | .38* | .01 | .02 | -.06 |
| Syntactic complexity | | | .32* | .23 | -.13 | .01 | -.01 |
| Quantity of talk | | | | .25 | -.24 | .87** | -.13 |
| Lexical variety (d index) | | | | | .25 | .17 | .03 |
| Lexical variety (plex lambda) | | | | | | -.22 | -.01 |
| Number of narrative clauses | | | | | | | .06 |

* $p < .05$. ** $p < .01$.

### *6.2.3 Comparing the tasks differing in cognitive complexity*

Having examined the different variants and the intercorrelations of the output measures on each task, we should now turn our attention to comparing students' performance on Tasks 1 and 2. For this purpose paired samples *t* tests were used, the results of which can be found in Table 22. It can be seen that there are a number of statistically significant differences in the output measures. Students tended to be more accurate in Task 2 (Task1 $M$=0.84, Task 2 $M$=0.87), and they also talked more while performing Task 2 (Task1 $M$=154.14, Task 2 $M$=238.75). When comparing means of indices of lexical variety as shown by the type token ratio (Task 1 d index $M$=53.14, Task 2 d index $M$=42.05) and the ratio of less frequent words (Task 1 plex lambda $M$=1.05, Task 2 plex lambda $M$=0.83), it seems that students used a wider range of words and a higher ratio of difficult words when performing Task 1. Students produced a higher number of narrative clauses in Task 2 (Task 1 $M$=9.04, Task 2 $M$=15.31), but the difference in the ratio of narrative clauses between the two tasks is statistically not significant. Although students tended to be more fluent on Task 1 (Task 1 $M$=132.63, Task 2 $M$=127.91), this difference in fluency is statistically not significant in this research sample, it is only a tendency level relationship ($p<0.1$). It seems that the two tasks differing in cognitive complexity did not differ in syntactic complexity, that is, in the ratio of subordination.

Table 22
*Paired samples t tests, comparing performance on Tasks 1 and 2*

| | N | mean | | SD | | mean of paired diff. | t | df | Sig (2-tailed) |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 1 | 2 | | | | |
| Accuracy | 41 | .84 | .87 | .11 | .10 | -.03 | 2.40 | 40 | .021* |
| Fluency | 41 | 132.63 | 127.91 | 31.57 | 31.55 | 4.71 | 1.99 | 40 | .053 |
| Syntactic complexity | 41 | 1.55 | 1.51 | .27 | .25 | .03 | .76 | 40 | .446 |
| Quantity of talk | 41 | 154.14 | 238.75 | 52.37 | 140.41 | -84.60 | -4.22 | 40 | .001* |
| Lexical variety (d index) | 41 | 53.14 | 42.05 | 14.44 | 9.48 | 11.09 | 4.75 | 40 | .001* |
| Lexical variety (plex lambda) | 41 | 1.05 | .83 | .29 | .30 | .22 | 4.09 | 40 | .001* |
| Number of narrative clauses | 41 | 9.04 | 15.31 | 3.64 | 7.64 | -6.26 | -5.47 | 40 | .001* |
| Ratio of narrative clauses | 41 | .65 | .69 | .15 | .14 | -.03 | -1.32 | 40 | .194 |

* $p < .05$. ** $p < .01$.

After comparing the means of students' performance on Tasks 1 and 2, it might be interesting to see whether the same output measures on the two tasks tend to change together. Table 23 presents the correlations of the pairs of output measures. There are three output measures that do not display significant positive correlations between the two tasks: syntactic complexity and the two indices of lexical variety (d index and plex lambda). For two other variables, the number of narrative clauses ($r_s$=0.32) and the ratio of narrative clauses ($r_s$=0.31) the relationship is statistically significant but rather low.

However, the quantity of talk ($r_s$=0.41), and especially accuracy ($r_s$=0.64) and fluency ($r_s$=0.88) measures display high correlations across the two tasks.

Table 23
*Correlations of the same task performance measures on Tasks 1 and 2*

| | N | Correlation | Sig. |
|---|---|---|---|
| Accuracy on Task 1 and Task 2 | 41 | .64** | .001 |
| Fluency on Task 1 and Task 2 | 41 | .88** | .001 |
| Complexity on Task 1 and Task 2 | 41 | .27 | .081 |
| Quantity of talk on Task 1 and Task 2 | 41 | .41** | .008 |
| Lexical variety (d index) on Task 1 and Task 2 | 41 | .27 | .084 |
| Lexical variety (plex lambda) on Task 1 and Task 2 | 41 | .30 | .057 |
| Number of narrative clauses on Task 1 and Task 2 | 41 | .32* | .039 |
| Ratio of narrative clauses on Task 1 and Task 2 | 41 | .31* | .043 |

* $p < .05$. ** $p < .01$.

## 6.3 Discussion

### *6.3.1 Oral narrative task variants*

The oral narrative tasks used in the study were designed to be of two different types, a cognitively more and a cognitively less complex one. Since the oral narrative tasks were performed in pairs, this necessitated that two versions of each task type should be employed. Thus, altogether four different oral narrative tasks were used in the study. In order to establish whether the tasks that were intended as variants were similar enough to be considered as variants of the same task, independent samples *t* tests were performed, the results of which can be found in Tables 18 and 19. The means of each output variable measured were compared for the two variants of the same task, and no statistically significant differences were found between them for either of the

two task types. Therefore, despite the fact that the cartoon strips presented different stories and the pictures to be used for the cognitively more complex task showed different story ingredients, they can be considered as variants of the same task, and data deriving from them can be merged for further analysis.

### *6.3.2 Characteristics of the oral narrative tasks differing in cognitive complexity*

Having established that the four oral narrative tasks used in the study belong to two types since there is no statistically significant difference between the variants for any of the measures, intercorrelations of the output measures for each task were examined. Table 20 presented intercorrelations for Task 1, while Table 21 showed the same intercorrelations for Task 2. Interestingly, there seems to be a cluster of output measures that tend to change together in the case of Task 1, these are accuracy, the quantity of talk, the d index and the number of narrative clauses (see Table 20). The rest of the output measures seem to be more or less independent of each other. No such cluster can be detected in the case of Task 2, almost all the output measures seem to be independent of one another. It is only the quantity of talk that correlates with two other measures, complexity and the number of narrative clauses (see Table 21). This might indicate that language proficiency probably determined performance on Task 1 to a greater extent, while other variables, like individual differences, probably exerted a greater effect when students performed the cognitively more complex Task 2.

More proficient learners can be expected to outperform their less proficient peers in general, that is regarding the quantity and quality of their performance. They probably talk more, do this more accurately and display greater lexical variety than less proficient speakers. Evidence for this can be found in section 8.2.1, Table 29 presenting the correlational analyses of TOEFL scores and task performance measures. Therefore, it is probably not very surprising if we find that these variables tend to change together, as they do in the case of Task 1, the cognitively less complex task. What needs to be explained is when they fail to correlate; this is what happened in the case of Task 2, the cognitively more complex task. In this case it is plausible to hypothesise that other factors, for

example individual differences, might have come into play and acted in a way that is different from the way English proficiency affected task performance. Support for this line of argumentation can be found in section 8.2.2, Table 30 showing correlations of task performance measures and aptitude, and sections 8.2.3.1 and 8.2.3.2, Tables 31 through 35 presenting correlations of creativity and oral narrative task performance measures. Aptitude seems to be related to performance on Task 1 and in a similar way as proficiency; therefore, it does not upset correlations, whereas creativity is in connection with performance on Task 2 and displays a different pattern from proficiency. The finding that performance on the cognitively more complex task was more affected by ID variables is also line with Robinson's (2003) Cognition Hypothesis claiming that individual differences play a greater role when learners perform a cognitively more complex task. A recent study conducted by Robinson (2007b) provides further empirical support for this in connection with output processing anxiety.

### *6.3.3 Comparing the tasks differing in cognitive complexity*

After examining performance on the two task types separately, it is interesting to see whether differences can be observed on the cognitively less and more complex task regarding the output measures used. Paired samples *t* tests were used to compare the means of students' performance on Task 1 and Task 2 (see Table 22). Results indicate that students talked more, they used a larger number of narrative clauses and they were more accurate on Task 2, the cognitively more complex task. The larger number of narrative clauses is probably a result of the fact that students talked more on this task, as the difference in the ratio of narrative clauses is statistically not significant. The significantly higher number of words uttered might a be a consequence of task design; while students were given a complete story in 6 pictures in Task 1, the 6 pictures in Task 2 were not connected, they simply showed story ingredients. Probably more cognitive effort, and also a higher number of words were needed to connect and make sense of these unrelated pictures.

Robinson (2001c, 2003, 2005b) believes that an increase in cognitive complexity which is brought about by making a task less structured has an attention dispersing effect; therefore, it does not result in more accurate performance. In contrast to his expectations, in my study the more complex task seemed to have an attention directing effect, and students became more accurate. A plausible explanation for this can be found in the characteristics of the tasks. In the cognitively less complex task, the protagonists of the stories were depicted in the pictures, as well as the changes in location, and the order of events. Therefore, the participants might not have felt the need to be very precise with personal pronouns and tenses since the storylines were self-evident for them. In the cognitively more complex version of the task, only some story ingredients were provided, and the participant had to create the stories themselves; therefore, nothing could be taken for granted. Accuracy plays an important part in storytelling in the sense that protagonists, changes in location and the timeline must be kept clear for the listeners to be able to follow the story (Brown, & Yule, 1983). It seems that students felt more compelled to meet this expectation in the cognitively more complex task; therefore, this task directed their attention to accuracy more than Task 1.

In the case of Task 1, the cognitively less complex task, participants displayed significantly greater lexical variety for both measures examined (d index, plex lambda). This might indicate that in the cognitively less complex task, students had more resources at hand that they could direct to this aspect of the task, probably at the expense of accuracy. It seems that having received a ready-made story urged them to concentrate less on accuracy and demonstrate their command of English by using a higher ratio of different words, as reflected by the d index, and a higher ratio of difficult or rare words, as shown by the plex lambda. Therefore, this particular task type seems to direct learners' attention to using sophisticated vocabulary. Moreover, this is the point where the results are in line with the claims of the Cognition Hypothesis (Robinson 2001c, 2003, 2005b), the attention dispersing effect of removing task structure resulted in a decrease in lexical complexity on the more complex task.

Although the difference in fluency is statistically not significant, there seems to be a trend that students were more fluent on Task 1. This is in line

with what might be expected on the basis of the Cognition Hypothesis (Robinson, 2001c, 2003, 2005b) and also with what Skehan (1998) predicts in his framework of task difficulty, that is, learners' performance tends to be more fluent on cognitively less complex tasks. A replication of this study on a larger sample, or increasing sample size by recruiting more participants might be sufficient for this finding to reach significance level.

Besides comparing participants' performance on the two tasks with the help of paired samples *t* tests, correlations of the same task performance measures on Tasks 1 and 2 were also examined (see Table 23). These correlations suggest that there are some measures that correlate very highly, these are fluency and accuracy; some correlate moderately, these are the quantity of talk, the number of narrative clauses and the ratio of narrative clauses; while the rest do not correlate significantly at all across the two tasks, these are the measures of lexical variety and syntactic complexity. A plausible interpretation of these findings is that those measures that correlate very highly are relatively constant characteristics of the given learners or of their proficiency. It seems that speech rate is one such stable characteristic, and so is accuracy to some extent; therefore, if a learner is fluent and accurate on one task, they tend to be fluent and accurate on another as well. The amount of talk produced, and as a consequence the number of narrative clauses used and the ratio of narrative clauses uttered, seems to have some stability as well; it seems that some learners tend to speak less while others tend to speak more regardless of the specific task at hand.

It is probably not very surprising that the ratio of subordination used as reflected by the measure of syntactic complexity is not a stable characteristic of any learner; it probably depends to a great extent on the task to be performed. Similarly, displaying lexical variety, as shown both by the type token ratio (measured by the d index) and the ratio of rare words (as operationalized by the plex lambda), is only a possibility in the sense that learners do not necessarily demonstrate it from task to task. If the task requires them to use varied vocabulary and they have the necessary resources at hand, they probably tend to use varied vocabulary; in other cases this might not happen.

Findings of this study are generally in line with the results of research reported in the literature concerning oral narrative task performance on cognitively less and more complex tasks although comparisons are not always easy to make because of the different dimensions of task characteristics examined. Citing only those research studies that seem to be comparable to this one, it can be seen that greater accuracy on the cognitively more complex task was also identified by Robinson (1995) and Iwashita et al. (2001) while the cognitively less complex (more structured) task was also found to lead to greater fluency (Skehan & Foster, 1999). If we consider planning time as a resource dispersing dimension (Robinson, 2001b, 2005b, 2007a), we can also refer to studies where reducing cognitive complexity by giving planning time resulted either in greater fluency and syntactic complexity (Foster & Skehan, 1996) or in greater fluency and accuracy (Skehan & Foster, 1997), which is not entirely compatible with my results. Making tasks cognitively more complex along resource directing dimensions seems to increase lexical complexity/diversity (Robinson, 1995) or syntactic complexity (Robinson, 2007b) although these changes are not always detectable in general measures, only in specific ones.

## 6.4 Conclusion

On the basis of the findings presented in this chapter, it can be concluded that the four oral narrative tasks used in the study seem to belong to two distinct types differing in cognitive complexity. Correlational analyses performed on the two task types separately suggest that global language proficiency might be more determinant when students solve a cognitively less complex task, whereas other factors, like individual differences tend to play a greater role in the cognitively more complex task. This finding and similar results reported by Robinson (2007b) seem to provide empirical evidence for one of the claims of the Cognition Hypothesis (Robinson, 2001c, 2003, 2005b). They substantiate the assumption that ID variables play a greater role in performance on cognitively more complex tasks.

In the present study, participants tended to talk more and use a higher number of narrative clauses on the cognitively more complex task, and they were also more accurate. The cognitively less complex task resulted in greater lexical variety and fluency though the difference is not statistically significant in the case of the latter. Cognitive complexity did not seem to affect syntactic complexity and the ratio of narrative clauses in this sample. These results are generally in line with findings reported in the literature (Foster & Skehan, 1996; Iwashita, et al., 2001; Robinson, 1995, 2007b; Skehan & Foster, 1997, 1999), and partially support claims of the Cognition Hypothesis (Robinson, 2001c, 2003, 2005b). They also suggest that in case of a seeming contradiction, the specific characteristics of the task need also be taken into consideration. This way it can be revealed that a task that was believed to be cognitively more complex along resource dispersing dimensions might also have some features that are resource directing. It seems that what we witnessed here are synergic effects (Robinson, 2005b).

# Chapter 7: Correlations of Individual Differences and Proficiency

## 7.1 Introduction

This section of the dissertation presents findings from correlational analyses of individual differences and proficiency. In the study reported here, two individual difference variables, a well-known and much-researched one, language aptitude, and a potentially important one, creativity, were examined. Correlational analyses were performed in order to find answers to the research questions concerning whether these are directly related to ultimate attainment in L2. First correlations of aptitude and proficiency were computed to examine how language aptitude and L2 competence measures are related. Then correlations of creativity and language proficiency were calculated. Finally, the relationship of the two individual variables: aptitude and creativity, was examined.

## 7.2 Results

### *7.2.1 Correlations of aptitude and proficiency*

Table 24 presents correlations between aptitude test scores (HUNLAT) and language proficiency measures (TOEFL-PBT, C-test). There are hardly any statistically significant correlations in the table; and the ones that exist between the Language Analysis part of HUNLAT and the Total TOEFL score ($r_s$=0.32) and the TOEFL Structure and Reading Comprehension score ($r_s$=0.33) are quite low.

Table 24
*Correlations of aptitude test scores and scores on language proficiency tests for 1st year English majors at ELTE (N=41)*

| | TOEFL Listening Comprehension | TOEFL Structure and Written Expression | TOEFL Vocabulary and Reading Comprehension | Total TOEFL score | C-test score |
|---|---|---|---|---|---|
| Hidden Sounds | 0.20 | 0.23 | 0.22 | 0.24 | 0.15 |
| Language Analysis | 0.23 | 0.26 | 0.33* | 0.32* | 0.28 |
| Words in Sentences | 0.04 | 0.13 | 0.13 | 0.10 | 0.21 |
| Vocabulary Learning | 0.04 | 0.08 | 0.11 | 0.10 | 0.01 |
| Total Language Aptitude score | 0.11 | 0.26 | 0.24 | 0.24 | 0.21 |

* $p < .05$. ** $p < .01$.

### *7.2.2 Correlations of creativity and proficiency*

As the multimethod-multitrait analysis has shown that the different tasks which comprise the creativity test do not seem to measure the same construct, correlations between various measures (average originality, creative fluency and relative flexibility) of the different tasks and the TOEFL and C-test scores were correlated for each of the four tasks. Table 25 presents the correlation coefficients. It seems that students' performance on the non-verbal, figural tasks of the creativity test (Circles and Picture Completion) is unrelated to students' language proficiency measures. In the case of the Unusual Uses task, relative flexibility correlates significantly with the Listening Comprehension part of TOEFL ($r_s$=0.31), and with the C-test score ($r_s$=0.38). Results of the Remote Association task indicate that average originality correlates with the Vocabulary and Reading Comprehension part of TOEFL

($r_s$=0.31). Therefore, in the two verbal creativity tasks, which showed some connection with language proficiency, two different aspects of creativity, relative flexibility and average originality proved to be relevant.

Table 25
Correlations between different tasks of the creativity test and language proficiency scores

| | TOEFL Listening Comprehension | TOEFL Structure and Written Expression | TOEFL Vocabulary and Reading Comprehension | Total TOEFL score | C-test score |
|---|---|---|---|---|---|
| Unus. Uses Aver. orig. | 0.24 | -0.08 | 0.13 | .011 | 0.16 |
| Unus. Uses Creat. flue. | 0.19 | 0.06 | 0.10 | 0.13 | 0.21 |
| Unus. Uses Relat. flex. | 0.31* | 0.14 | 0.22 | 0.27 p=0.08 | 0.38* |
| Rem. Asso. Aver. orig. | 0.25 | 0.15 | 0.31* | 0.23 | 0.23 |
| Rem. Asso. Creat. flue. | 0.05 | -0.21 | -0.05 | -0.09 | 0.06 |
| Rem. Asso. Relat. flex. | 0.05 | -0.15 | 0.03 | -0.04 | 0.06 |
| Circles Aver. orig | 0.22 | 0.03 | 0.12 | 0.17 | 0.03 |

| | TOEFL Listening Comprehension | TOEFL Structure and Written Expression | TOEFL Vocabulary and Reading Comprehension | Total TOEFL score | C-test score |
|---|---|---|---|---|---|
| Circles Creat. flue. | 0.10 | 0.02 | -0.06 | -0.01 | 0.02 |
| Circles Relat. flex. | 0.05 | -0.01 | 0.11 | 0.11 | 0.05 |
| Pict. Com. Aver. orig | 0.11 | 0.13 | 0.21 | 0.15 | 0.24 |
| Pict. Com. Creat. flue. | -0.10 | -0.22 | -0.14 | -0.20 | -0.12 |
| Pict. Com. Relat. flex | -0.06 | -0.20 | -0.04 | -0.11 | -0.08 |

*Note.* Unus. Uses = Unusual Uses; Rem. Asso. = Remote Associations; Pict. Com. = Picture Completion; Aver. orig. = Average originality; Creat. flue. = Creative fluency; Relat. flex. = Relative flexibility.

* $p < .05$. ** $p < .01$.

Despite having reservations about the precise meaning and usefulness of composite creativity scores (see the results of the multitrait-multimethod analysis in section 5.2.1), in order to assure comparability with previous studies (Albert & Kormos, 2004), the composite scores were also used in the correlational analyses with language proficiency. From Table 26 it seems that only one aspect of creativity, average originality is related to language proficiency when measured by two parts of the TOEFL test, Listening Comprehension ($r_s$=0.34) and Vocabulary and Reading Comprehension ($r_s$=0.37). The tendency level ($p$<0.1) relationship that was found between relative flexibility and the C-test scores in the previous study (Albert & Kormos, 2004) disappeared; therefore it was probably caused by chance. In the

present sample a tendency level relationship exists between average originality and the C-test score ($r_s$=0.27, $p$=0.84) and the total TOEFL-PBT score ($r_s$=0.30, $p$=0.54). Although the findings presented in Table 26 suggest that the only aspect of creativity that might be related to language proficiency is average originality, this relationship tends to be rather weak even in those cases when it reaches the level of 5% statistic significance.

Table 26
*Correlations between composite creativity scores and language proficiency test scores for 1st year English majors at ELTE (N=41)*

| | TOEFL Listening Comprehension | TOEFL Structure and Written Expression | TOEFL Vocabulary and Reading Comprehension | Total TOEFL score | C-test score |
|---|---|---|---|---|---|
| Average originality | 0.34* | 0.12 | 0.37* | 0.30 $p$=0.054 | 0.27 $p$=0.084 |
| Creative fluency | 0.02 | -0.14 | -0.07 | -0.10 | 0.01 |
| Relative flexibility | 0.21 | 0.01 | 0.18 | 0.17 | 0.23 |
| Total creativity | 0.10 | -0.11 | -0.01 | -0.03 | 0.10 |
| Verbal creativity | 0.04 | -0.04 | 0.06 | 0.05 | 0.17 |
| Figural creativity | 0.043 | -0.06 | -0.04 | -0.05 | 0.06 |

* $p < .05$. ** $p < .01$.

### *7.2.3 Correlations of creativity and aptitude*

Table 27 presents correlations of the different parts of the aptitude test, and the three relevant measures of creativity: average originality, creative fluency, and relative flexibility for each of the tasks of the creativity test. On the verbal

tasks, one aspect of creativity, creative fluency, seems to be negatively related to language aptitude as it has negative correlations with the Hidden Sounds subtest in the case of Unusual Uses ($r_s$=-0.34), and with the Language Analysis subtest in the case of Remote Associations ($r_s$=-0.33). The tendencies portrayed by the figural tasks are different from those suggested by the verbal ones. Although similarly to the verbal tasks creative fluency correlates negatively with both the Hidden Sounds ($r_s$=-0.50) and the Language Analysis ($r_s$=-0.31) subtasks of HUNLAT in the Picture Completion task; surprisingly, there seems to be a significant positive correlation between relative flexibility and the total language aptitude score on this particular aptitude test task. In a similar manner, the average originality score correlates positively with the total language aptitude score on the Circles task.

Table 27
*Correlations between different tasks of the creativity test and language aptitude scores*

| | Hidden Sounds | Language Analysis | Words in Sentences | Vocabulary Learning | Total Language Aptitude score |
|---|---|---|---|---|---|
| Unus. Uses Aver. orig. | 0.02 | -0.02 | -0.01 | 0.10 | 0.03 |
| Unus. Uses Creat. flue. | -0.34* | -0.27 | 0.04 | 0.05 | -0.10 |
| Unus. Uses Relat. flex. | -0.25 | -0.01 | 0.11 | 0.02 | 0.01 |
| Rem. Asso. Aver. orig. | -0.06 | 0.26 | 0.06 | 0.08 | 0.09 |

| | Hidden Sounds | Language Analysis | Words in Sentences | Vocabulary Learning | Total Language Aptitude score |
|---|---|---|---|---|---|
| Rem. Asso. Creat. flue. | -0.29 | -0.33* | 0.04 | -0.18 | -0.25 |
| Rem. Asso. Relat. flex. | -0.30 | -0.25 | 0.15 | -0.22 | -0.17 |
| Circles Aver. orig | 0.14 | 0.17 | -0.01 | 0.38* | 0.24 |
| Circles Creat. flue. | 0.01 | -0.07 | 0.16 | 0.11 | 0.11 |
| Circles Relat. flex. | -0.08 | 0.09 | -0.06 | 0.21 | 0.09 |
| Pict. Com. Aver. orig | -0.01 | -0.02 | 0.21 | 0.07 | 0.12 |
| Pict. Com. Creat. flue. | -0.50** | -0.31* | 0.01 | 0.01 | -0.20 |
| Pict. Com. Relat. flex | 0.14 | 0.28 | 0.28 | 0.15 | 0.32* |

*Note.* Unus. Uses = Unusual Uses; Rem. Asso. = Remote Associations; Pict. Com. = Picture Completion; Aver. orig. = Average originality; Creat. flue. = Creative fluency; Relat. flex. = Relative flexibility.

* $p < .05$. ** $p < .01$.

Correlations between different parts of the aptitude test and composite scores of creativity are presented in Table 28. The figures in the table reinforce the main trends already witnessed when examining the creativity tasks separately above, that is, there seems to be a negative relationship between

creative fluency and some aspects of the language aptitude test, as creative fluency correlates negatively with the Hidden Sounds task ($r_s$=-0.46) and with the Language Analysis task ($r_s$=-0.41). The same negative relationship can be seen when looking at the total creativity score and its correlations with language aptitude: total creativity correlates negatively with the Hidden Sounds task ($r_s$=-0.48) and with the Language Analysis task ($r_s$=-0.33). The last statistically significant relationship in the table is the negative correlation between verbal creativity and the Hidden Sounds task ($r_s$=-0.37).

Table 28
*Correlations between composite creativity test scores and language aptitude test scores for 1st year English majors at ELTE (N=41)*

| | Hidden Sounds | Language Analysis | Words in Sentences | Vocabulary Learning | Total Language Aptitude score |
|---|---|---|---|---|---|
| Average originality | .04 | .18 | .09 | .26 | .20 |
| Creative fluency | -.46** | -.41** | .10 | .01 | -.17 |
| Relative flexibility | -.24 | -.05 | .16 | .01 | .03 |
| Total creativity | -.48** | -.33* | .13 | .06 | -.12 |
| Verbal creativity | -.37* | -.28 | .10 | -.07 | -.15 |
| Figural creativity | -.27 | -.17 | .22 | .15 | .07 |

* $p < .05$. ** $p < .01$.

## 7.3 Discussion

### *7.3.1 Correlations of aptitude and proficiency*

This part of the dissertation examined correlational analyses calculated between individual differences and language proficiency. Besides the individual difference of creativity, which is the main focus of my investigations, a well-established and much-researched individual difference, aptitude was also examined. It seems that although language aptitude is one of the best predictors of the rate of progress when it comes to learning a foreign language (Dörnyei, 2005; Ehrman, & Oxford, 1995), it might not be a good predictor of ultimate attainment. Since it seems that only relatively weak relationships exist between one task of the language aptitude test, Language Analysis, and one part of the TOEFL, the Structure and Reading Comprehension part, and consequently the total score (see Table 24); it can be argued that language aptitude and proficiency appear to be almost unrelated at this level. Kormos and Sáfár (2006) reported similar results in a study conducted with intermediate learners; they found that the strongest relationship can be found between inductive language learning ability and proficiency, whereas grammatical sensitivity played a limited role. In an attempt to explain low correlations with phonetic coding ability, they tend to agree with Skehan (1998) who hypothesises that phonetic coding ability probably has a role in earlier stages of language learning, but it is almost irrelevant at an intermediate level. Moreover, based on other findings (e. g. Carroll, 1990), they questioned the validity of the rote learning ability subtest altogether. My findings seem to support this line of argumentation, and suggest that in the case of advanced learners only inductive language learning ability seems to retain its importance out of the traditional components of language aptitude. There is a possibility, however, that other aptitude complexes (Robinson, 2001a, 2005a) might come into play in this later phase of language learning.

### *7.3.2 Correlations of creativity and proficiency*

When examining correlations of language proficiency and different aspects of creativity on the four tasks of the creativity test (see Table 25), it becomes obvious that there is no statistically significant relationship between the students' English proficiency and their performance on the figural, that is, drawing tasks of the creativity test. On the verbal tasks, however, there seems to be some relationship between creativity and proficiency. On the Unusual Uses subtask, relative flexibility correlates positively with the TOEFL-PBT Listening Comprehension part. The relative flexibility score of the creativity test reflects the number of categories the respondents choose their answers from, irrespective of the actual number of responses they give. The ability of giving a wide variety of answers seems to be moderately related to listening comprehension skills in this sample.

It can be hypothesised that the reason why relative flexibility might be helpful when someone is solving a listening comprehension task is that a student characterised by higher flexibility keeps a wider range of options available while performing the task. This might be advantageous when the learner is trying to understand the text as they probably choose from several possible interpretations instead of narrowing down the topic prematurely. High flexibility can also be beneficial when selecting the answer to a particular question based on the text. Comprehension questions on listening texts usually require listeners to go beyond what is evident, and the most likely answer often turns out to be a distractor. Flexible students might be better at avoiding obvious distractors, and as a result of keeping several interpretations alive make good guesses.

Besides listening comprehension, relative flexibility is positively related to a global measure of language proficiency, the C-test score (see Table 25). Moreover, relative flexibility and the other global measure of English proficiency the total TOEFL-PBT score also seem to be related although their correlation is not significant at the 5% level. It is only to a tendency level relationship ($p<0.1$) which might reach significance on a larger sample. Interestingly, the relationship between a composite measure of relative flexibility and the C-test score was already demonstrated in a previous

exploratory study (Albert & Kormos, 2004). Nevertheless, further research would be needed to substantiate the existence of this moderate but direct link between one aspect of creativity and language proficiency.

Although on the basis of this finding it can be hypothesised that a direct link exists between the relative flexibility aspect of creativity and English proficiency, the pattern of correlations suggests an alternative interpretation as well. This alternative interpretation is that high relative flexibility is advantageous in every testing situation, when someone has to choose from several options. Relative flexibility might be beneficial under these circumstances because through considering several options, flexible people might avoid being distracted by obvious but incorrect answers. If the relationship really exists between proficiency and creativity, it is not easy to explain why it is significant in the case of one test, the C-test, but not in the case of the other, the TOEFL-PBT. If, however, relative flexibility is considered as a helpful test-taking skill, it can be argued that it has a greater role in the case of a C-test, as not only a choice has to be made there, but also the options have to be generated by the respondent.

The other relationship that is statistically significant at the 5% level is between the average originality measure of the Remote Associations task and the Vocabulary and Reading Comprehension part of the TOEFL-PBT test (see Table 25). The average originality score of the creativity test reflects the statistical rarity of the respondents' answers. This positive but moderate correlation shows that those students who gave unusual answers on this task, in other words, who had unusual associations, scored higher on the Vocabulary and Reading Comprehension part of TOEFL-PBT. In order to reach high scores on this part of the TOEFL-PBT, students need to have a wide range of vocabulary and good reading skills, a precondition of which is again having good knowledge of vocabulary (Alderson, 2000; Wagner, Muse, & Tannenbaum, 2007). On the basis of this finding, a link might be hypothesised to exist between having unusual ideas and knowing a large number of words, that is, a large foreign language lexicon and a good access to it.

In order to ensure comparability with our previous exploratory study (Albert & Kormos, 2004) composite scores of creativity were also used in the

analysis (see Table 26). These findings, however, fail to support the results of the previous study since C-test scores correlated positively with relative flexibility there. In contrast to this, correlations with composite scores on this sample show that positive correlations exist between average originality and some aspects of language proficiency. Average originality correlates positively with the Listening Comprehension and the Vocabulary and Reading Comprehension parts of the TOEFL-PBT, and there is a tendency level relationship ($p<0.1$) between global measures of language proficiency, that is, the total TOEFL-PBT score and the C-test score. It seems that students who produced statistically rare, that is, unusual answers, scored higher on these parts of the TOEFL-PBT, and there might be a moderate but direct link between average originality and English proficiency. These findings again suggest that there might be a direct relationship between creativity and students' foreign language vocabulary, discussed in the previous paragraph. However, in the light of these somewhat contradictory findings, it seems that further research would be needed to determine which aspect of creativity, relative flexibility or average originality, is in direct relationship with language proficiency.

Findings of this and our previous exploratory study (Albert & Kormos, 2004) suggest that if there is a relationship between any aspect of creativity and language proficiency, it probably exists between a fluency-free component of creativity, either relative flexibility or average originality, and proficiency, but creative fluency and proficiency are probably unrelated. The fact that the creative fluency score was not found to be related language proficiency either in this or in earlier studies (Albert, & Kormos, 2004) might provide an explanation why no relationship was found between creativity and language proficiency in earlier studies (NYEK kutatócsoport, 2004). It is quite likely that instead of the fluency free scores, composite scores of creativity were used in these studies which are heavily determined by the creative fluency score, that is, the number of responses provided by the participants. Therefore, while these studies lend support to the independence of creative fluency and language proficiency apparent in my research as well, they unfortunately do not provide

information about the correlations of fluency-free scores and language proficiency where a relationship might actually exist.

### *7.3.3 Correlations of creativity and aptitude*

When examining the relationship between scores of language aptitude and the three aspects of creativity on the different tasks, it is obvious that three out of the four tasks, the two verbal and one figural, seem to reflect a similar pattern (see Table 27). In the Unusual Uses, Remote Associations and Picture Completion tasks, creative fluency seems to be negatively related to some aspect of language aptitude, either to the Hidden Sounds subtask as in the case of Unusual Uses and Picture Completion, or to the Language Analysis subtask as in the case of Remote Associations and Picture Completion. Creative fluency reflects the number of responses that the respondent gave on the open-ended tasks. It seems that the more answers the participants gave, the lower they scored either on the Hidden Sounds task reflecting phonetic coding ability, or on the Language Analysis part measuring inductive language learning ability.

Although this negative relationship is not easy to account for, there is a group of people who show similar characteristics: dyslexics. Learning sound-letter correspondences and extracting rules, which are the skills tested by the Hidden Sounds and Language Analysis parts of HUNLAT, cause problems for dyslexics even in their mother tongue. While these people tend to have difficulties when learning and using reading and writing, they are often found to be exceptionally talented and creative. West (1997) ascribes this finding to differences in brain functioning when compared to the average population, which can be made responsible for both phenomena, dyslexia and creativity. He believes that while the majority of people are characterised by left-hemisphere dominance which is associated with logic, language and sequential time, dyslexics can be characterised by a visual-spatial or right-hemisphere mode of thought. Right hemisphere activation or dominance is also believed to play a role in creativity (Martindale, 1999). Although it is obvious that conclusions concerning brain functioning cannot be drawn on the basis of this

present study, the possibility of the existence of a common neurological background, that is relative right-hemisphere dominance, cannot be excluded either.

In the light of the above it is quite surprising that on the fourth task of the creativity test, Circles, not only does creative fluency appear to be independent of all the aspects of language aptitude, but another aspect of creativity, average originality, is positively correlated with the Vocabulary Learning subtask measuring rote learning ability. Similarly, there is also a significant positive relationship between relative flexibility and the total language aptitude score on the Picture Completion task. Although on the basis of the literature reviewed in section 2.1.1, positive correlations between language aptitude and creativity could have been hypothesised as there is at least one new trend in aptitude research (see section 2.1.1 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, empirical findings seemed to contradict this expectation. A possible explanation of this contradiction is that the instrument used for measuring language aptitude in the study has different theoretical background from the one that suggests such a relationship. HUNLAT uses the four components of language aptitude identified by Carroll and Sapon (1959), whereas the aptitude test which rests on the CANAL-FT theory (Grigorenko et al., 2000) aims to measure quite different components.

As mentioned in the previous paragraph, there are two positive relationships which are in line with our expectations formed on the basis of the literature review. One of them is the positive correlation between relative flexibility and the total language aptitude score in the case of the Picture Completion task, which suggests that students who chose their responses from a wider range of categories displayed higher language aptitude on average. The other such relationship is the positive correlation between the average originality score in the Circles task and the Vocabulary learning task score. This means that those students who produced statistically rare, that is, unusual responses on the Circles task, were able to remember a higher number of words

in the Vocabulary Learning task, that is, they were better at rote learning on this particular task. When we examined the correlations between creativity and language proficiency in section 7.3.2, it was revealed that average originality also correlates with the Vocabulary and Reading Comprehension section of TOEFL-PBT. A plausible explanations for this is that higher average originality on the Circles task might be related to a better rote learning ability resulting in a wider range of vocabulary that is manifested in the higher TOEFL-PBT score.

Correlations between the aptitude test and the composite creativity scores reflect the same trends as described above (see Table 28). However, the positive correlation between average originality and Vocabulary learning that reached statistical significance on the circles task is only a tendency level ($p<0.1$) relationship here. Therefore, a larger sample would be needed to substantiate the existence of this relationship. The negative correlation between creative fluency and the Hidden Sounds and Language Analysis subtests that could be observed in three of the tasks is statistically significant on the composite measures as well. Since the total creativity score is very much determined by the creative fluency score, it shows the same correlations as creative fluency. The reason why verbal creativity reflects a similar pattern but figural creativity does not lies in the fact that one figural task, Circles task failed to show the above pattern. Since correlations with the composite scores clearly reflect similar trends to the ones observed on individual tasks, but in some cases less explicitly, relying on the creativity measures of the individual tasks appears to be more appropriate here as well.

## 7.4 Conclusion

Having examined the correlations of the two individual variables, language aptitude and creativity, with each other and with English proficiency, the following conclusions can be drawn. It seems that when examining learners whose proficiency is between the intermediate and advanced levels, the relationship between language aptitude and their proficiency becomes rather

weak. Since language aptitude is intended as a predictor of the rate of progress and not of ultimate attainment, this finding is not unexpected.

As regards composite scores of creativity, findings of this study fail to support the results of our exploratory study (Albert & Kormos, 2004) since in the exploratory study it was relative flexibility that correlated positively with C-test scores, whereas here average originality seemed to be connected to proficiency. Thus, further research is necessary to clarify the relationships of these variables. On the basis of results originating from correlations with the individual creativity tasks, we can hypothesise that relative flexibility might be advantageous when being tested by certain types of proficiency tests, while there might be a direct link between average originality and the participants' vocabulary knowledge. What seems quite certain is that if there is any direct relationship between creativity and proficiency, it exists between a fluency-free component of creativity, either relative flexibility or average originality, and proficiency. This finding also provides a possible explanation why no direct relationship was found between creativity and proficiency previously (NYEK kutatócsoport, 2004) in studies where measurement probably heavily relied on the number of responses provided by the participant.

The relationship of the two individual variables, language aptitude and creativity appears to be problematic. First of all, hardly any support, except for two correlations on the drawing tasks, was found for the positive relationship hypothesised on the basis of a current theory of language aptitude, CANAL-FT (see section 2.1.1). This is probably caused by the fact that the instruments used in the study measure different components of the constructs from the ones that are discussed by the CANAL-FT theory. Differences in the theoretical background which largely determine the method of measurement are likely to be held accountable for this discrepancy. Based on our findings, negative relationships exist between the creative fluency component of creativity showing the number of responses a person gives on an open-ended task and two components of language aptitude: phonetic coding ability and inductive language learning ability. These relationships are particularly strong in the cases of the verbal tasks of the creativity test. Since a similar phenomenon is observable in dyslexics (West, 1997), a possible right-hemisphere dominance

also characteristic of creative people (Martindale, 1999) can be hypothesized to be in the background.

# Chapter 8: Correlations of Individual Differences in Proficiency and Abilities with Task Performance Measures

## 8.1 Introduction

This section of the dissertation focuses on oral narrative task performance, and the way it is related to individual differences: language aptitude and creativity. As it seems to be evident that performance on a foreign language task is dependent on the learners' level of proficiency, correlations of oral narrative task performance and English proficiency measures are examined first. Since the oral narrative tasks used in the study differ in cognitive complexity, I will also explore whether the level of English proficiency has a differential role in the case of the cognitively less and more complex tasks. Next, correlations of aptitude and oral narrative task performance measures are examined in an attempt to detect any differences between the less and more complex task. The last section of the chapter investigates the relationships between students' performance on different tasks of the creativity test and task performance measures. The presentation of the results is followed by the discussion of the findings.

## 8.2 Results

### *8.2.1 Correlations of proficiency and task performance*

Although it appears to be evident that performance on a foreign language task should be determined by language proficiency, this seemingly straightforward relationship might not exist, or might not be equally strong in the case of all the performance measures examined in the study. Table 28 presents the correlations of proficiency and task performance measures for Tasks 1 and 2 together for ease of comparability, but before comparing the cognitively less and more complex tasks in these respects, the relationship of proficiency and task performance measures are discussed separately for each task.

It is evident from Table 29 that there are some task performance measures on Task 1 that are related to the participants' level of English proficiency. One such measure is accuracy, that is, the ratio of error-free clauses, which is positively correlated with both the overall language proficiency measures of the total TOEFL- PBT score ($r_s$=0.54) and the C-test score ($r_s$=0.57), and all the different parts of TOEFL-PBT. Although the correlation coefficients are somewhat lower, the same relationship can be seen in the case of the quantity of talk in Task 1 and one index of lexical variety, plex lambda, which reflects the ratio of difficult or rare words used in the course of solving Task 1. Since the quantity of talk and the number of narrative clauses are strongly related (see Table 19), it should not be surprising that the number of narrative clauses used shows a similar pattern although the correlation with the Listening comprehension part of TOEFL-PBT is statistically not significant. Interestingly, it is only the Listening comprehension score of TOEFL-PBT which correlates with fluency, that is speech rate, on Task 1. Syntactic complexity and the ratio of narrative clauses is not related to any measure of language proficiency on Task 1, and the other measure of lexical variety, the d-index, which reflects type-token ratio, only moderately correlates with the C-test score ($r_s$=0.33) but not with any measure of TOEFL-PBT.

Similarly to Task 1, there are four task performance measures that seem to be strongly related to students' language proficiency in the case of Task 2 (see Table 29). Accuracy, as measured by the ratio of error-free clauses and one measure of lexical variety, the d-index correlate with all measures of English proficiency significantly. In the case of the other two variables: fluency and the quantity of talk produced, most of the correlations with the proficiency scores are positive and statistically significant as well. One exception is the TOEFL Structure and Written Expression score where the correlations are not significant for either of them, moreover the C-test score does not correlate with fluency. Despite the fact that there is a positive relationship between the quantity of talk produced and measures of language proficiency, the number of narrative clauses only correlates with the C-test score in the case of Task 2 ($r_s$=0.47). There seems to be no relationship between syntactic complexity on

Task 2, the other measure of lexical variety, plex lambda, and the ratio of narrative clauses and any of the English proficiency measures.

If we compare the two tasks with regard to how task performance measures relate to language proficiency, we can identify some similarities as well as some differences. The cognitively less and more complex tasks seem to be similar in the sense that accuracy and the quantity of talk produced are positively correlated with proficiency, while syntactic complexity and the ratio of narrative clauses are not related to proficiency in any way. The differences between the tasks concern the two measures of lexical variety, the number of narrative clauses, and fluency. One measure of lexical variety, the plex lambda indicating a higher ratio of rare words, is positively correlated with proficiency in the case of the cognitively less complex task, while the other measure, the d index reflecting type token ratio, has significant positive correlations with proficiency on the cognitively more complex task. Although this seems to be the general trend, the C-test score correlates moderately, but significantly with the d index on the cognitively less complex task as well ($r_s$=0.33).

When we look at correlations with the number of narrative clauses, we can see that the pattern displayed by C-test scores is somewhat unexpected again. Although it seems that the number of narrative clauses is only related to proficiency positively in the case of the cognitively less complex task, there is a strong positive correlation between the C-test score and the number of narrative clauses in the case of the cognitively more complex task as well ($r_s$=0.47). Despite the fact that there seems to be a general trend in the case of fluency as well, that is, fluency seems to be related to proficiency in the case of the cognitively more complex task, the picture is the least clear here. On the one hand, neither the C-test score nor the TOEFL Structure and Written Expression score correlates with fluency on either of the tasks. On the other hand, the TOEFL-PBT Listening comprehension score correlates positively with fluency in the case of the cognitively less complex task as well. ($r_s$=0.39)

Table 29
*Correlations between language proficiency and task performance measures for Tasks 1 and 2 (N=41)*

| | TOEFL Listening Comprehension | TOEFL Structure and Written expression | TOEFL Vocabulary and Reading Comprehension | Total TOEFL score | C-test score |
|---|---|---|---|---|---|
| Accuracy Task 1 | .48** | .52** | .48** | .54** | .57** |
| Accuracy Task 2 | .46** | .49** | .37* | .50** | .40** |
| Fluency Task 1 | .39* | .14 | .23 | .29 | .11 |
| Fluency Task 2 | .47** | .24 | .45** | .43** | .20 |
| Syntactic complexity Task 1 | -.01 | -.21 | -.05 | -.11 | -.07 |
| Syntactic complexity Task 2 | .21 | .12 | .18 | .22 | .09 |
| Quantity of talk Task 1 | .36* | .36* | .50** | .45** | .39* |
| Quantity of talk Task 2 | .38* | .27 | .33* | .38* | .47** |
| Lexical variety (d index) Task 1 | .28 | .15 | .18 | .21 | .33* |
| Lexical variety (d index) Task 2 | .40** | .45** | .50** | .49** | .60** |
| Lexical variety (plex lambda) Task 1 | .45** | .39* | .42** | .48** | .37* |

| | TOEFL Listening Comprehension | TOEFL Structure and Written expression | TOEFL Vocabulary and Reading Comprehension | Total TOEFL score | C-test score |
|---|---|---|---|---|---|
| Lexical variety (plex lambda) Task 2 | .22 | .27 | .24 | .30 | .17 |
| Number of narrative clauses Task 1 | .27 | .35* | .36* | .35* | .41** |
| Number of narrative clauses Task 2 | .28 | .26 | .20 | .29 | .47** |
| Ratio of narr clauses Task 1 | .03 | .01 | .04 | .01 | .11 |
| Ratio of narr clauses Task 2 | -.07 | -01 | -.21 | -.11 | .02 |

* $p < .05$. ** $p < .01$.

### *8.2.2 Correlations of aptitude and task performance*

Despite studies indicating that language aptitude largely determines the rate of progress when learning a foreign language (Carroll, 1981), its relationship with language proficiency in the case of intermediate and advanced learners, that is at a high level of ultimate attainment is not so straightforward, as it could be seen in section 7.3.1 of this dissertation. The question arises whether it is possible at all to find any relationship between language aptitude scores and specific task performance measures for the cognitively less and more complex tasks. Similarly to the previous section, the correlations between language aptitude and task performance measures for the

two tasks (see Table 30) are analysed separately before comparing them along the different measures.

Out of the four parts of HUNLAT hypothetically measuring four different aspects of language aptitude, the Hidden Sounds part reflecting the phonetic coding ability of the respondents seems to be most closely related to task performance measures on Task 1, the cognitively less complex task (see Table 29). Scores on the Hidden Sounds part are significantly positively correlated with accuracy ($r_s$=0.35), the quantity of talk produced ($r_s$=0.47), one measure of lexical variety, the d index, reflecting type token ratio ($r_s$=0.34), and the number of narrative clauses ($r_s$=0.43). Scores on the Vocabulary Learning task reflecting rote learning ability correlate negatively with two measures of task performance: the number of narrative clauses in Task 1 ($r_s$=-0.35) and the ratio of narrative clauses in Task 1 ($r_s$=-0.43). Interestingly, scores achieved on the Words in Sentences part measuring grammatical sensitivity do not correlate with any measure of task performance on Task 1. The remaining component of language aptitude, the Language Analysis score, and the total language aptitude score correlate with one measure of task performance each. The Language Analysis score correlates with the other measure of lexical variety, the plex lambda, reflecting a higher ratio of difficult words used ($r_s$=0.33), while the total language aptitude score correlates negatively with the ratio of narrative clauses in the cognitively less complex task ($r_s$=-0.39).

Interestingly, there are fewer correlations between language aptitude scores and task performance measures in the case of the cognitively more complex task, Task 2. It is again scores on the Hidden Sounds task that have the highest number of correlations with task performance measures: they positively correlate with fluency, that is speech rate ($r_s$=0.33), and with the quantity of talk produced ($r_s$=0.31). There are two other parts of the language aptitude test that are related to one task performance measure each: the Words in Sentences score reflecting grammatical sensitivity is positively correlated with one measure of lexical variety, the d index, while scores in the Vocabulary Learning task measuring rote learning ability correlate with syntactic complexity on the cognitively more complex task ($r_s$=0.35). However, the Language Analysis score appears to be independent of measures

of task performance on Task 2. Similarly to the Words in Sentences score, the total language aptitude score positively correlates with the d index, a measure of lexical variety ($r_s$=0.34).

If we try to compare the relationship of language aptitude scores and task performance measures on the cognitively less and more complex tasks, the first conclusion that can be drawn is that the number of statistically significant correlations is double in the case of Task 1, the cognitively less complex task, compared to Task 2, the cognitively more complex one (see Table 30). Despite this difference, a similarity between the two tasks is that scores on the Hidden Sounds subtest measuring phonetic coding ability seem to be the most strongly related to task performance measures on both tasks though the specific measures are not the same in the case of the two tasks. Besides the variable of the quantity of talk produced which correlates with the Hidden Sounds score on both tasks, phonetic coding ability seems to be related to accuracy, one measure of lexical variety, the d index, and the number of narrative clauses produced in Task 1, and to fluency in Task 2. Scores on the Vocabulary Learning part of the aptitude test seem to relate to task performance measures on the cognitively less and more complex tasks in different ways. Whereas on the cognitively less complex task they have negative correlations with the number of narrative clauses and the ratio of narrative clauses, on the cognitively more complex task rote learning ability correlates positively with syntactic complexity. The Language Analysis scores appear to be related to performance on the cognitively less complex task only, while the Words in Sentences score only correlates with a performance measure of the cognitively more complex task. The language aptitude total score also correlates with different measures of task performance on the less and more complex tasks: positively with accuracy and negatively with the ratio of narrative clauses on Task 1, and positively with one measure of lexical variety, the d index, on Task 2.

Table 30
*Correlations between language aptitude and task performance measures on Tasks 1 and 2*

| | Hidden Sounds | Language Analysis | Words in Sentences | Vocabulary Learning | Total Language Aptitude score |
|---|---|---|---|---|---|
| Accuracy Task 1 | .35* | .26 | .25 | .21 | .32* |
| Accuracy Task 2 | .10 | .23 | .12 | -.08 | .06 |
| Fluency Task 1 | .23 | .09 | -.01 | .29 | .17 |
| | | | | *p*=0.066 | |
| Fluency Task 2 | .33* | .09 | .05 | .24 | .20 |
| Syntactic complexity Task 1 | -.11 | -.17 | .10 | -.14 | -.13 |
| Syntactic complexity Task 2 | .11 | .05 | .16 | .35* | .26 |
| Quantity of talk Task 1 | .47** | .19 | .20 | -.08 | .22 |
| Quantity of talk Task 2 | .31* | .04 | .05 | .14 | .22 |
| Lexical variety (d index) Task 1 | .34* | .24 | .10 | .06 | .20 |
| Lexical variety (d index) Task 2 | .20 | .11 | .33* | .21 | .34* |
| Lexical variety (plex lambda) Task 1 | .18 | .33* | .22 | .02 | .24 |
| Lexical variety (plex lambda) Task 2 | -.22 | .23 | -.12 | .25 | .07 |
| Number of narrative clauses Task 1 | .43** | .17 | .09 | -.35* | .03 |

| | Hidden Sounds | Language Analysis | Words in Sentences | Vocabulary Learning | Total Language Aptitude score |
|---|---|---|---|---|---|
| Number of narrative clauses Task 2 | .14 | -.06 | -.02 | -.01 | .07 |
| Ratio of narr clauses Task 1 | -.18 | -.23 | -.11 | -.43** | -.39* |
| Ratio of narr clauses Task 2 | -.23 | -.24 | -.22 | -.14 | -.26 |

* $p < .05$. ** $p < .01$.

### *8.2.3 Correlations of creativity and task performance*

This section examines whether it is possible to find any relationship between an individual variable like creativity and various measures of task performance. In a similar fashion as it was done previously, creativity scores deriving from the four tasks of the creativity test are examined separately first, followed by the correlations of composite scores and task performance measures.

#### 8.2.3.1 Correlations with subtests of creativity

Table 31 presents correlations of one of the verbal tasks of the creativity test, Unusual Uses, and task performance measures on the two tasks. Examining the two tasks separately, it becomes obvious that in the case of Task 1, the cognitively less complex task, there is only one statistically significant relationship in the table, and that is between average originality and fluency ($r_s$=0.31). There are altogether four correlations between different measures of creativity and the task performance measures of Task 2, the cognitively more complex task. Syntactic complexity on Task 2 correlates with creative fluency ($r_s$=0.37), the quantity of talk with average originality ($r_s$=0.31), one measure of lexical variety, the d index with relative flexibility ($r_s$=0.31), and the ratio of narrative clauses with creative fluency ($r_s$=0.41).

Table 31
*Correlations between the creativity test task Unusual Uses and task performance measures on Tasks 1 and 2*

| | Average originality | Creative fluency | Relative flexibility |
|---|---|---|---|
| Accuracy Task 1 | -.01 | -.02 | .01 |
| Accuracy Task 2 | -.03 | -.04 | .02 |
| Fluency Task 1 | .31* | .14 | .26 |
| Fluency Task 2 | .21 | .02 | .18 |
| Syntactic complexity Task 1 | -.06 | .07 | .04 |
| Syntactic complexity Task 2 | .17 | .37* | .28 |
| Quantity of talk Task 1 | .07 | .06 | .17 |
| Quantity of talk Task 2 | .31* | .10 | .21 |
| Lexical variety (d index) Task 1 | -.03 | -.12 | -.12 |
| Lexical variety (d index) Task 2 | .08 | .10 | .31* |
| Lexical variety (plex lambda) Task 1 | .17 | -.13 | .01 |
| Lexical variety (plex lambda) Task 2 | -.03 | -.15 | -.04 |
| Number of narrative clauses Task 1 | .02 | .10 | .20 |
| Number of narrative clauses Task 2 | .14 | .22 | .25 |
| Ratio of narrative clauses Task 1 | .03 | .28 | .29 |
| Ratio of narrative clauses Task 2 | .07 | .41** | .28 |

* $p < .05$. ** $p < .01$.

Table 32 presents correlations between measures of the other verbal task, Remote Associations, and task performance measures on the two tasks. Interestingly, it seems that there are no statistically significant correlations between any of the creativity and task performance measures on this particular verbal creativity task.

Table 32
*Correlations between the creativity test task Remote Associations and task performance measures on Tasks 1 and 2*

| | Average originality | Creative fluency | Relative flexibility |
|---|---|---|---|
| Accuracy Task 1 | .17 | .14 | -.01 |
| Accuracy Task 2 | -.01 | .18 | .17 |
| Fluency Task 1 | -.06 | .05 | -.09 |
| Fluency Task 2 | -.10 | .03 | -.13 |
| Syntactic complexity Task 1 | -.13 | .18 | .21 |
| Syntactic complexity Task 2 | .12 | -.01 | -.20 |
| Quantity of talk Task 1 | -.02 | .09 | .25 |
| Quantity of talk Task 2 | -.04 | .02 | .04 |
| Lexical variety (d index) Task 1 | -.12 | .17 | .10 |
| Lexical variety (d index) Task 2 | .22 | .10 | .13 |
| Lexical variety (plex lambda) Task 1 | .12 | -.12 | .01 |
| Lexical variety (plex lambda) Task 2 | .27 | -.04 | -.01 |
| Number of narrative clauses Task 1 | .02 | .07 | .23 |
| Number of narrative clauses Task 2 | -.13 | .14 | .10 |
| Ratio of narrative clauses Task 1 | .10 | -.06 | -.05 |
| Ratio of narrative clauses Task 2 | -.02 | .15 | -.01 |

Table 33 presents correlations in the case of one of the drawing tasks, Circles. It we examine the two story-telling tasks separately, it can be seen that the number of significant correlations is lower in the case of the cognitively less complex task, Task 1. On this task, the average originality aspect of creativity is negatively related to syntactic complexity ($r_s$=-0.37), while creative fluency correlates positively with fluency, that is speech rate ($r_s$=0.32).

On Task 2, the cognitively more complex task, accuracy is negatively correlated with relative flexibility ($r_s$=0.40), while one measure of lexical variety, the plex lambda, seems to be significantly positively related to both average originality ($r_s$=0.36) and relative flexibility ($r_s$=0.40).

Table 33
*Correlations between the creativity test task Circles and task performance measures on Tasks 1 and 2*

| | Average originality | Creative fluency | Relative flexibility |
|---|---|---|---|
| Accuracy Task 1 | -.04 | .14 | -.20 |
| Accuracy Task 2 | -.15 | .21 | -.40** |
| Fluency Task 1 | -.04 | .32* | -.08 |
| Fluency Task 2 | .01 | .22 | -.05 |
| Syntactic complexity Task 1 | -.37* | .01 | -.09 |
| Syntactic complexity Task 2 | .02 | -.01 | .03 |
| Quantity of talk Task 1 | -.05 | .07 | -.01 |
| Quantity of talk Task 2 | .11 | -.07 | .17 |
| Lexical variety (d index) Task 1 | -.24 | .14 | -.07 |
| Lexical variety (d index) Task 2 | .08 | .08 | -.01 |
| Lexical variety (plex lambda) Task 1 | .25 | -.09 | .12 |
| Lexical variety (plex lambda) Task 2 | .36* | -.13 | .40** |
| Number of narrative clauses Task 1 | -.08 | .01 | -.08 |
| Number of narrative clauses Task 2 | .10 | .02 | .17 |
| Ratio of narrative clauses Task 1 | -.17 | -.02 | -.17 |
| Ratio of narrative clauses Task 2 | -.01 | .06 | -.07 |

* $p < .05$. ** $p < .01$.

Table 34 presents correlations between measures of the other drawing task, Picture Completion, and task performance measures on the two story-telling tasks. Similarly to the second verbal task, Remote Associations, it seems that there are no statistically significant correlations between any of the creativity and task performance measures on this particular creativity task.

Table 34
*Correlations between the creativity test task Picture Completion and task performance measures on Tasks 1 and 2*

| | Average originality | Creative fluency | Relative flexibility |
|---|---|---|---|
| Accuracy Task 1 | .12 | -.12 | -.09 |
| Accuracy Task 2 | -.01 | .01 | -.13 |
| Fluency Task 1 | .24 | -.05 | -.03 |
| Fluency Task 2 | .09 | -.21 | -.16 |
| Syntactic complexity Task 1 | .09 | .09 | -.07 |
| Syntactic complexity Task 2 | .13 | -.07 | -.04 |
| Quantity of talk Task 1 | .05 | -.17 | .19 |
| Quantity of talk Task 2 | .18 | -.04 | .12 |
| Lexical variety (d index) Task 1 | -.15 | -.28 | -.08 |
| Lexical variety (d index) Task 2 | .26 | -.22 | -.21 |
| Lexical variety (plex lambda) Task 1 | .06 | -.24 | .26 |
| Lexical variety (plex lambda) Task 2 | .01 | -.10 | .01 |
| Number of narrative clauses Task 1 | .16 | -.24 | .18 |
| Number of narrative clauses Task 2 | .13 | -.04 | .05 |
| Ratio of narrative clauses Task 1 | .23 | .11 | .07 |
| Ratio of narrative clauses Task 2 | -.04 | -.15 | .24 |

### 8.2.3.2 Correlations with composite scores of creativity

Although the use of composite scores is not justified on the basis of the results of the present study (see section 5.2.1.), in order to ensure comparability with our previous study (Albert & Kormos, 2004) correlations were calculated with composite scores as well (see Table 35). The trends witnessed in the case of the four creativity tasks can also be detected here, that is, performance on the cognitively more complex task seems to be somewhat more affected by creativity, whereas the cognitively less complex task seems to be independent of it.

On the level of composite measures, no statistically significant relationship can be detected between measures of creativity and task performance as far as Task 1, the cognitively less complex task is concerned (see Table 35). Average originality has a statistically significant relationship with the d index, one of the measures of lexical variety on Task 2, the cognitively more complex task ($r_s$=0.33). Besides this, average originality has a tendency level ($p$<0.1) relationship with the other measure of lexical variety, the plex lambda, also on the cognitively more complex task ($r_s$=0.30). The composite measure of verbal creativity is positively related to the ratio of narrative clauses in the story in the case of Task 2, the cognitively more complex task ($r_s$=0.33).

Table 35
*Correlations between composite scores of creativity and task performance measures on Tasks 1 and 2*

| | Average originality | Creative fluency | Relative flexibility | Total creativity score | Verbal creativity score | Figural creativity score |
|---|---|---|---|---|---|---|
| Accuracy Task 1 | .11 | .02 | -.12 | .01 | .04 | .03 |
| Accuracy Task 2 | -.02 | .08 | -.05 | .04 | .07 | .06 |
| Fluency Task 1 | .21 | .16 | .12 | .18 | .12 | .13 |

| | Average originality | Creative fluency | Relative flexibility | Total creativity score | Verbal creativity score | Figural creativity score |
|---|---|---|---|---|---|---|
| Fluency Task 2 | .14 | -.01 | .08 | .01 | .04 | -.04 |
| Syntactic complexity Task 1 | -.25 | .12 | .07 | .09 | .17 | -.01 |
| Syntactic complexity Task 2 | .13 | .09 | -.12 | .08 | .20 | -.10 |
| Quantity of talk Task 1 | -.02 | -.01 | .25 | .01 | .11 | -.01 |
| Quantity of talk Task 2 | .18 | -.07 | .22 | .04 | .05 | .03 |
| Lexical variety (d index) Task 1 | -.24 | -.12 | -.01 | -.11 | .01 | -.13 |
| Lexical variety (d index) Task 2 | .33* | .01 | .22 | .09 | .19 | .01 |
| Lexical variety (plex lambda) Task 1 | .25 | -.21 | .07 | -.13 | -.12 | -.06 |
| Lexical variety (plex lambda) Task 2 | .30 *p*=.054 | -.13 | .08 | -.07 | -.08 | -.03 |
| Number of narrative clauses Task 1 | -.01 | -.02 | .18 | -.01 | .13 | -.05 |

| | Average originality | Creative fluency | Relative flexibility | Total creativity score | Verbal creativity score | Figural creativity score |
|---|---|---|---|---|---|---|
| Number of narrative clauses Task 2 | .06 | .05 | .27 | .16 | .16 | .11 |
| Ratio of narrative clauses Task 1 | .03 | .15 | -.02 | .11 | .08 | .01 |
| Ratio of narrative clauses Task 2 | -.01 | .20 | .08 | .18 | .33* | -.10 |

* $p < .05$.

## 8.3 Discussion

### *8.3.1 Correlations of proficiency and task performance*

It sounds plausible that participants' level of proficiency in a foreign language determines the way they solve a task in that language. This commonsense conclusion, however, might not hold equally true for every single measure of task performance, and it might even differ from task to task. In order to shed light on the possible relationships or on the lack of them, different measures of language proficiency, a TOEFL-PBT total score and subscores and a C-test score, were correlated with the following measures of task performance: accuracy, fluency, syntactic complexity, quantity of talk, two indices of lexical variety (d index and plex lambda), and the number and ratio of narrative clauses. Findings of the correlational analyses are discussed in the paragraphs below.

There are four task performance measures that seem to be influenced by language proficiency in the case of Task 1, the cognitively less complex task (see Table 29). These are accuracy, the ratio of error-free clauses; the quantity of talk, the number of words uttered; one measure of lexical diversity, the plex

lambda reflecting the ratio of difficult words used; and the number of narrative clauses, that is the number of events in the story. Since all the above correlations were positive, it seems that the more proficient participants tended to make fewer errors while telling the story, they also talked more, used a higher ratio of difficult words and incorporated more events into their stories. The relationship between proficiency and accuracy sounds self-explanatory: they made fewer mistakes because they are more proficient, but it is an interesting question why they talked more. It is equally possible that they talked more because they felt more confident as a result of their high level of proficiency, or that they became proficient precisely because they usually tend to talk more while solving a task thus get more practice in these situations, which results in a higher level of proficiency (Swain, 1985). A third possibility is that they are merely more talkative or more willing to communicate (MacIntyre, Clément, Dörnyei, Noels, 1998), which because of the increased practice opportunities was an advantage. This could have resulted in a higher level of proficiency, and also manifested itself while solving this task. Since correlational analyses do not provide information about the direction of the relationship either of the listed options is possible.

The higher number of narrative clauses produced by the more proficient respondents might be a consequence of the fact that they talked more since there is no significant correlation between the ratio of narrative clauses produced per AS units and proficiency. This means that more proficient students did not incorporate more events into their stories if we take the length of the stories into account as well (see Table 29). The correlation with plex lambda, a measure of lexical variety, indicates that more proficient participants used a higher ratio of difficult or rare words in English, which is not surprising. Nevertheless, the fact that they used a higher number of narrative clauses, that is, they incorporated more events into their stories suggests that they produced better stories, and the higher ratio of rare words is likely to show that more proficient participants were able to find the right words, the appropriate vocabulary for the items depicted in the pictures that were needed for the story. Therefore, the more proficient the respondents were, the better they seemed to

cope with the task requirements which called for the reciting of a specific story, that is, a chain of events depicted by specific pictures.

There are two more relationships shown by the correlations, but unlike the previous ones discussed, these only manifest themselves for one measure of proficiency (see Table 29). One such relationship exists between the TOEFL Listening Comprehension score and the speech rate indicating fluency in Task 1. This relationship suggests that more proficient listeners in the sample tended to produce a higher number of syllables per minute as if their listening skills were connected to the fluency aspect of their oral skills in the case of the cognitively less complex task. The other such relationship exists between the C-test score and a measure of lexical variety, the d index reflecting type-token ratio. Therefore, those participants of the sample who had higher C-test scores used more varied vocabulary while solving the task. The co-occurrence of a higher level of proficiency and a wider range of vocabulary should not be surprising. Moreover, since Kontra and Kormos (2006) found that C-tests mainly measure vocabulary, this might explain why this relationship is not manifested for the TOEFL-PBT score.

When examining correlations between language proficiency and task performance measures in the case of the more complex task, Task 2, it seems that there are again four performance measures that are closely related to language proficiency, but these are somewhat different from the ones that are significant in the case of Task 1 (see Table 29). The overlapping variables are accuracy and the quantity of talk, that is, more proficient learners tended to be more accurate and they also talked more while solving the cognitively more complex task. It is plausible that the explanations offered for Task 1 might hold true here as well, that is, greater accuracy is a consequence of higher proficiency, while causality is difficult to determine in the case of the quantity of talk.

Interestingly, there seems to be a closer relationship between fluency, that is, the number of syllables uttered per minute and language proficiency in the case of the cognitively more complex task, as it manifested itself not only for one part of the TOEFL-PBT, the Listening comprehension score, but also for another part, the Vocabulary and reading comprehension score as well as

the total TOEFL-PBT score (see Table 29). It is quite likely that in the case of the cognitively more complex task extra attentional resources were needed to maintain high fluency, and these resources were only available for the more proficient speakers. The fourth task performance measure that is closely related to language proficiency on Task 2 is one measure of lexical variety, the d index reflecting type token ratio. It seems that more proficient participants were able to use a wider range of vocabulary on the cognitively more complex task. It is again possible that more proficient learners had more resources to direct them towards this aspect of the task.

If we want to compare the relationships of task performance measures with English proficiency on the two tasks differing in cognitive complexity, it might be important to briefly review task characteristics first. Task 1, the cognitively less complex task called for narrating a ready-made story depicted by pictures. The plot and characters were given, but the narration at some points required the knowledge of specific words. In Task 2, the cognitively more complex task, only some story ingredients were provided with the help of pictures, but the plot and characters had to be invented. Since the pictures were somewhat ambiguous, this made it possible for the speaker to avoid difficult words. These task characteristics might provide an explanation of the findings with regard to the differences found between the two tasks.

Besides similarities indicating that a higher level of proficiency coincides with greater accuracy and more talk on both types of task, there are certain differences as well, indicating that more proficient respondents seemed more able to cope with task requirements (see Table 29). On the cognitively less complex task, they used more difficult words, as shown by the plex lambda, probably to name the specific items shown on the pictures. Moreover, despite the fact that the length of the plot was given as different stages were shown by different pictures, they still managed to create longer plots, as shown by the number of narrative clauses. As the cognitively more complex task was less structured, and the respondents were given almost complete freedom with respect to the length of the story and the vocabulary used, higher proficiency manifested itself in more general measures: faster speech rate and a wider range of vocabulary used. Interestingly, syntactic complexity is not related to

any measure of proficiency on either of the tasks; it is more likely to be determined by other task characteristics not exploited by the narrative task or task content.

### *8.3.2 Correlations of aptitude and task performance*

Since language aptitude does not seem to be strongly related to language proficiency in the case of the advanced learners comprising the sample (see section 7.3.1), it was questionable whether any relationship can be identified between language aptitude scores and task performance measures. Moreover, if such relationships exist, it is not clear whether they would display the same pattern in the case of the cognitively less and more complex tasks.

Although the number of statistically significant correlations is considerably fewer than in the case of proficiency and task performance measures, the total language aptitude score seems to be related to two measures of task performance in the case of Task 1, the cognitively less complex task (see Table 30). Participants with higher aptitude scores tended to be more accurate, that is, they made fewer errors while narrating the cartoon strip task. It can be hypothesised that in this case greater language aptitude might have led to a higher level of proficiency and thus might have resulted in greater accuracy. The other relationship is more difficult to interpret: higher aptitude scores coincided with a lower ratio of narrative clauses in Task 1, that is irrespective of the total length of the story, these participants used a lower ratio of events. Since events provide the backbones of stories, this finding poses questions about the quality of the stories produced by these participants. It is possible that the explanation of the phenomenon lies in the way students characterised by higher aptitude allocate their attentional resources. It might be that these students mainly concentrate on the linguistic aspects of their performance but do not or cannot devote enough attention to creating stories. This hypothesis is supported by the fact that they are more accurate as stated above, and also by other characteristics listed in the subsequent paragraphs.

Out of the four parts comprising the language aptitude test, the Hidden Sounds section measuring phonetic coding ability seems to be most closely

related to task performance measures (see Table 30). Hidden Sounds scores positively correlate with accuracy, the quantity of talk produced, one measure of lexical variety, the d index, and the number of narrative clauses. Therefore, participants with better phonetic coding ability tended to make fewer errors, and they talked more while solving the task. They displayed more varied vocabulary in the sense that they used a higher number of different words relative to the number of words, and they included more events while telling their stories. Their task performance is quite similar to that of more proficient learners (see Table 29) despite the fact that there is no statistically significant relationship between scores on the Hidden Sounds task and language proficiency tests (see Table 24).

The correlation between the Language Analysis score and the other measure of lexical variety, the plex lambda (see Table 30), seems to point in the same direction, suggesting similarities between learners with high proficiency and high aptitude scores. This correlation means that participants with better inductive language learning abilities tended to use a higher ratio of difficult words, that is, they were probably able to use the specific vocabulary required by the task. This is precisely the way more proficient learners solved the cognitively less complex task.

The correlations between a third a component of aptitude, rote learning ability, measured by the Vocabulary Learning task is more difficult to interpret (see Table 30). There are negative correlations between the number and ratio of narrative clauses and the Vocabulary learning score. It seems that respondents with better rote learning abilities used quantitatively fewer and also relatively fewer events in their stories, which poses questions about the quality of their stories. In fact, it is probably this relationship which lies in the background of the negative relationship of the total aptitude score and the ratio of narrative clauses discussed at the beginning of this section. The explanation that these learners probably devote more attention to linguistic factors at the expense of other task requirements might also hold true here although the significant positive correlations present in the case of the total aptitude score are missing here. This however might simply be a consequence of the low number of participants, as a tendency level positive relationship can be detected between

the Vocabulary Learning scores and fluency on Task 1. If we assume that such a pattern of correlations signals that these students concentrate on the linguistic aspects of the task, then these relationships seems to highlight the crucial role of attention and of the way it is allocated.

When examining the relationship of aptitude scores and task performance measures in the case of Task 2, the cognitively more complex task (see Table 30), it becomes evident that the number of significant correlations dropped drastically. It seems as if with the increased cognitive load, language aptitude lost some of its importance, although the similarity with the pattern displayed by proficiency and performance measures remained. The language aptitude total score correlates positively with the d index, a measure of lexical variety, reflecting the type token ratio; therefore learners characterised by a higher level of aptitude used a higher ratio of different words; they displayed a wider range of vocabulary. The Hidden Sounds score correlates with fluency and the quantity of talk in the case of the cognitively more complex task, that is better phonetic coding ability coincided with higher speech rate and more speech produced overall. Similarly to the total score, the Words in Sentences part positively correlates with the d index; thus, participants with higher levels of grammatical sensitivity used a wider range of vocabulary, as well.

As in the case of the cognitively less complex task, the correlation with the Vocabulary Learning score defies the interpretation emphasising the similarity with the language proficiency scores. It seems that those participants who had better rote learning abilities used syntactically more complex sentences, that is more subordination, while performing the cognitively more complex task. Although no such relationship was detected in the case of the more proficient students, this finding is in line with the hypothesis that these learners seem to devote their attentional resources to the linguistic aspects of the task probably at the expense of other factors, such as the plot in the case of Task 1. It seems that the cognitively more complex task urged these respondents to use more subordination, that is, syntactically more complex sentences.

When trying to compare the two tasks at a more general level, the following conclusions can be drawn: first of all, there seems to be a dramatic

drop in the number of significant correlations from Task 1 to Task 2. This suggests that performance on a cognitively less complex task might be more directly influenced by language aptitude, while in the case of the cognitively more complex task other factors might have come into play as well. Secondly, despite the fact that there are a much higher number of significant correlations between language proficiency and task performance measures, the trends witnessed there are similar to what could be seen here. There are correlations with the quantity of talk and one measure of lexical variety, the d index, in both tasks; accuracy and the other measure of lexical variety the plex lambda reflecting a higher ratio of rare words only correlates in Task 1; while fluency seems to be more important in Task 2. Despite these similarities, correlations of the Vocabulary Learning task reflecting rote learning ability defy these trends. They can be interpreted, however, if we hypothesise that learners with better rote learning abilities tend to concentrate on linguistic aspects of tasks at the expense of other factors, for example the plot.

### *8.3.3 Correlations of creativity and task performance*

The thorough analysis of the different measures of the creativity test (see section 5.2.1) revealed that using composite scores might not be justifiable as the respondents' performance varies considerably across the four subtasks of the test. Although the reasons for this might be purely motivational, that is, students might have liked or disliked certain tasks or they might have got tired of them, it is also possible that the four different tasks of the creativity test measure slightly different things, somewhat different aspects of creativity. Despite this finding, I opted for using composite scores besides the scores of individual tasks in order to make the results of this study comparable with a previous exploratory study (Albert & Kormos, 2004).

Findings of the correlational analyses seem to support the approach taken when examining correlations between creativity and task performance measures separately on the four subtasks. It was revealed that on two out of the four tasks of the test, creativity and task performance measures are unrelated. Contrary to what might have been expected, this division is not along the

verbal versus drawing tasks. Creativity and task performance measures correlate on one verbal (Unusual Uses) and one drawing (Circles) task, but not on the other verbal (Remote Associations) and other drawing (Picture Completion) task. This result is not surprising in the light of the fact that results of the multitrait-multimethod analysis (see Table 12 in section 5.2.1) already suggested that the four tasks of the test probably measure slightly different aspects of creativity.

On the basis of the characteristics of the creativity test tasks, it might be hypothesised that the Unusual Uses subtest draws on respondents' capacity to come up with a large number of novel solutions, whereas the Remote Associations subtest attempts to map the number and quality of associations a person has. Considering the two drawing tasks, it might be argued that the Picture Completion subtest probably taps visual creativity to a greater extent that the Circles task, as abstract shapes need to be developed into interesting pictures when performing Picture Completion. The Circles task seems to be similar to the Unusual Uses task in the sense that the emphasis is placed on a large number of novel solutions although these need to be drawn.

Examining the two verbal tasks of the test (see Tables 31 and 32), it can be seen that creativity measures calculated on Remote Associations, the last task on the test, are not related to task performance measures on any of the story-telling tasks. Creativity measures calculated from the first task, Unusual Uses, however display some moderate correlations on both story telling tasks. The only correlation that can be detected in the case of the cognitively less complex task, Task 1, is between average originality and fluency, that is, speech rate. This indicates that those respondents who came up with a higher number of statistically rare solutions on the task uttered a higher number of syllables per minute while telling the story. Since the cognitively less complex task did not involve planning only formulation, it might be that higher average originality helped increase speech rate through a greater range of vocabulary, that is, by making it quicker for the creative person to find the right words.

As hypothesised, it seems that learner creativity played a greater role in the case of solving Task 2, the cognitively more complex task. In the case of this task, which also involved the invention of the story, that is planning,

besides linguistic formulation, students characterised by higher average originality tended to talk more. The ability to come up with a higher number of statistically rare solutions on the Unusual Uses task (unusual uses for different objects) coincided with using more words and creating longer stories. It is possible that these students had some unconventional ideas that they incorporated into their stories in the planning phase, which resulted in longer stories as the novelty of ideas required more explanation. The other significant correlation which is between relative flexibility and one measure of lexical variety, the d index, reflecting type token ratio suggests that creativity might be linked to vocabulary. The students who can be characterised with higher relative flexibility, that is, they selected their answers from a wide range of categories displayed greater lexical variety, used a higher ratio of different words. Therefore, these students either knew more words, or they had better access to them.

The third aspect of creativity measured by the test is creative fluency, which reflects the number of responses provided. This measure of creativity is related to two task performance measures on Task 2: complexity and the ratio of narrative clauses. Although it was expected that students characterised by a higher level of creative fluency would talk more, this was not the case. It was revealed that these students used a higher ratio of narrative clauses, that is, relatively more events in their stories. Since the task involved creating the stories themselves, creative fluency probably acted at the idea generation stage, when the students made decisions about the events taking place, and invented the plot of the story. The significant correlation between creative fluency and syntactic complexity is somewhat surprising, as it was hypothesised that creativity would not affect this variable. Further investigations would be needed to determine whether this relationship also exists in other samples and to prove that it was not caused by chance. A possible explanation of the phenomenon might be that since students characterised by a higher level of creative fluency can produce a large number of ideas in the planning phase with ease, they have extra resources that can be allocated to formulating syntactically complex sentences.

When looking at the two drawing tasks of the test (see Tables 33 and 34), it can be seen that creativity measures calculated from Picture Completion, the third task of the creativity test are not related to any task performance measures either on Task 1 or on Task 2. There are altogether five correlations between different aspects of creativity and task performance measures on the Circles task, the second task of the creativity test. In Task 1, the cognitively less complex task, creative fluency correlates with fluency, that is speech rate. Therefore, it seems that those students who came up with a large number of solutions, that is, generated a large number of ideas on the Circle task (involving creating pictures from circles) uttered a higher number of syllables per minute. Since this story-telling task did not provide a chance for inventing extra events as the stories were given, creative fluency might have manifested itself in increased speech rate for the stories probably through better access to words.

The other correlation that can be found in the case of Task 1 is a negative relationship between average originality and syntactic complexity. As it was hypothesised that creativity would not affect this variable, further investigations would be needed to determine whether this relationship also exists in other samples. A possible explanation for this phenomenon can be that when telling a story, syntactic complexity, that is the ratio of subordination, is probably used for providing detailed information about characters, locations and events in a linguistically sophisticated way. Participants characterised by higher average originality might have been concentrating too much on inventing unusual ideas, which was not easy or possible at all with a ready-made story; therefore, they had no available resources for either adding many details, or expressing them in a linguistically sophisticated way, that is, using subordination.

There is also an unexpected negative relationship between relative flexibility and accuracy in the case Task 2, the cognitively more complex task. Although it was hypothesised that no relationship would exist between accuracy and creativity, it can be hypothesised that coming up with a wide range of ideas leaves less attention for accurate performance. An interesting parallel could be drawn here with language aptitude. It seems that while

students characterised by a higher level of language aptitude seemed to devote less attention to the non-linguistic aspects of the task in some cases, for example they used fewer events in their stories (see section 8.3.2), more creative participants devoted less attention to certain linguistic aspects of their performance, such as syntactic complexity and accuracy. The ID variables of language aptitude and creativity seem to have antagonistic effects in this sense.

The other two correlations on Task 2, the cognitively more complex task, involve two measures of creativity: average originality and relative flexibility, and one measure of task performance: the plex lambda. It seems that those students who either produced a higher number of statistically rare solutions, or their solutions derived from a wider range of categories used more difficult words on Task 2; therefore, they displayed greater lexical variety.

When examining correlations with composite scores (see Table 35), the findings seem to be more straightforward in the sense that the unexpected correlations between creativity and task performance measures such as syntactic complexity and accuracy do not appear at this level. There is one statistically significant and one tendency level positive relationship between average originality and the two measures of lexical variety, and one statistically significant positive relationship between verbal creativity and the ratio of narrative clauses; all of them in the case of the cognitively more complex task, Task 2. Students characterised by higher average originality, who produced more statistically rare, unusual ideas, displayed greater lexical variety in terms of using a higher ratio of different words and also in terms of using more difficult words in their stories. Students who are characterised by higher verbal creativity produced a higher ratio of narrative clauses in their stories, in other words they invented more events. As apart from the average originality and relative flexibility scores all other scores are heavily influenced by fluency, we can hypothesise that respondents displaying high verbal creativity were probably good at idea generation at least on the two verbal tasks.

Based on findings from the individual tasks and the composite scores, the following can be concluded: it seems that more creative students tended to have a slight advantage on the story-telling tasks. As hypothesised, this advantage was more tangible in the case of the cognitively more complex task,

where the task itself called for the invention of a story. Findings suggest that the way divergent thinking or as Carroll (1993) labelled it the ability of idea production or general retrieval ability worked was through an easier retrieval of ideas and words and this is the reason why it correlated with measures like fluency, quantity of talk, ratio of narrative clauses and lexical diversity.

## 8.4 Conclusion

This section of the dissertation discussed the main research question of the study, the relationship of a potentially important individual variable, creativity and task performance measures. Before analysing this, however, it might be useful to summarize how the level of proficiency and aptitude affected students' performance on the cognitively less and more complex tasks. Perhaps not surprisingly, the level of English proficiency seemed to determine task performance to a great extent, but in the case of the two tasks with different levels of cognitive complexity slightly different factors seemed to play a greater role. Besides being more accurate and talking more, more proficient speakers seemed to be able to cope with task requirements better: they used the difficult words that the task called for and managed to include more events in the story in case of the cognitively less complex task, whereas they used more varied vocabulary and talked faster in the case of the cognitively more complex task than their less proficient counterparts.

Interestingly, although the relationship between language aptitude and proficiency was not strong at the level of general measures (see section 7.2.1), students with a higher level of aptitude tended to behave in a way very similar to proficient participants. Although the relationships were more moderate, the pattern is similar: greater aptitude correlated with greater accuracy, more talk, greater lexical variety and fluency. This mainly holds true for the cognitively less complex task, while the cognitively more complex task seems to be less affected by language aptitude.

Having examined the relationships of creativity and task performance measures on the four tasks comprising the creativity test separately, it became obvious that there are only two tasks out of the four where relationships can be

detected. Surprisingly, these are not the two verbal tasks, but one verbal and a drawing task. Despite some inconsistencies in the findings that necessitate further research for clarification, the main trends seem to suggest moderate relationships between oral task performance and some aspects of creativity, especially the fluency-free components such as average originality and relative flexibility. As expected, creativity seemed to have a greater effect on the cognitively more complex task that was less structured thus provided an opportunity for using one's imagination. It is hypothesised that the reason why creativity is mainly in connection with lexical diversity is that it helps in an easier retrieval of unusual concepts as suggested by Carroll (1993).

# Chapter 9: Conclusions and Pedagogical Implications

## 9.1 Introduction

This chapter of the dissertation contains the main conclusions of the study. First, in a brief summary of findings answers to the research questions posed in section 1.2 are provided. Next, I discuss the limitations of the study which impose constraints with regard to the generalizability of my results. The chapter is concluded by outlining the pedagogical implications of my findings and pointing out future research directions.

## 9.2 Summary of findings

The main aim of the dissertation was to describe the relationships between a possibly important ID variable, learner creativity, and oral narrative task performance. In order to provide a context for the study and ease the interpretation of findings, two further variables were introduced in the research design: the ID variable of language aptitude and the level of language proficiency. Since the cognitive complexity of tasks was hypothesised to have a differential effect on the results, two oral narrative tasks with different levels of cognitive complexity were used. Eventually, seven research questions were formulated in connection with the characteristics of the sample, differences in task performance on the tasks differing in cognitive complexity, and the relationships of the variables examined.

The first research question concerned characteristics of the sample. It seems that 1st year English major students at ELTE can be characterised by a relatively high level of language aptitude, while their English proficiency ranges from intermediate to advanced. A possible factor that might account for this discrepancy is the years spent studying English, that is, it can be hypothesised that those who have devoted more time to learning the English language are probably more proficient in it. As far as creativity is concerned, it can be argued that although the average level of creativity seems to be higher

than the national average in the sample, distribution figures show that individuals with high and low creativity can also be found among them. Therefore, the sample seems to be a suitable population for demonstrating the possible effects of creativity.

The second research question referred to identifying differences in the participants' performance on the cognitively less and more complex tasks. Findings of the study can be interpreted in the framework of Robinson's (2001c, 2003, 2005b) Cognition Hypothesis and his Triadic Componential Framework for task classification (Robinson 2001b, 2005b, 2007a). The cognitively more complex task was believed to differ from the cognitively less complex one along resource dispersing dimensions: it was less structured and it also involved the creation of a story besides linguistic formulation. The Cognition Hypothesis claims that if a task is made more complex along resource dispersing dimensions, it results in less fluent, less complex and less accurate performance. In line with these claims, performance on the more complex task was less complex lexically as shown by both measures of lexical diversity although syntactic complexity was not affected. A tendency level ($p$<0,1) decrease in fluency could also be detected on the cognitively more complex task. However, the cognitively more complex task resulted in more accurate performance which was contrary to our previous expectations.

A possible explanation of this seeming contradiction is that the cognitively more complex task probably differed from the less complex one along resource directing dimensions as well (Robinson 2001b, 2005b, 2007a). The fact that in the case of the cognitively more complex task only some story ingredients were given but the story itself was not depicted by the pictures probably urged participants to concentrate more on differentiating protagonists and marking changes of location. Since an increase in cognitive complexity brought about by resource directing features usually result in more accurate and more complex performance, the fact that participants' performance was more accurate on the more complex task no longer contradicts the claims of the Cognition Hypothesis (Robinson, 2001c, 2003, 2005b), it merely demonstrates synergic effects. A further claim of the Cognition Hypothesis that individual

differences probably play a greater role in the case of cognitively more complex tasks, also seems to be supported by the findings.

The third research question was formulated about the relationship of the individual variables language aptitude and creativity with language proficiency. It seems that in the case of learners whose proficiency is between the intermediate and advanced levels, the relationship between aptitude and proficiency is rather weak. This finding is not surprising in the light of the fact that language aptitude is believed to predict the rate of progress but not ultimate attainment. Correlations between learner creativity and English proficiency also tend to be rather low. Findings of the present study indicate that there might be a direct link between the average originality component of creativity and the participants' vocabulary knowledge. Since this result somewhat contradicts findings of an earlier exploratory study (Albert & Kormos, 2004) where relative flexibility was found to be related to C-test scores, further research is needed to clarify the relationship of creativity and language proficiency. However, there seems to be some support for the assumption that relative flexibility might be advantageous when solving certain types of test, C-tests for example.

The fourth research question concerned the relationship of the ID variables themselves; it implied investigating the connections between creativity and language aptitude. Although recent theories of language aptitude, for example the CANAL-F theory (Grigorenko, et al. 2000), suggest that a positive relationship might exist between aptitude and creativity, results of the study provided marginal support for this assumption. The fact that the aptitude test used in the study conveys a traditional view of language aptitude on which the MLAT (Carroll, & Sapon, 1959) is based, therefore, it fails to measure components that are believed to be central in the CANAL-F theory, can probably be held accountable for this discrepancy. The majority of the relationships identified in this study are negative ones: the creative fluency component of creativity seems to be negatively related to the phonetic coding ability and the inductive language learning ability. Interestingly, a similar phenomenon is observable in dyslexics, and dyslexia is believed to be partially caused by a relative right-hemisphere dominance (West, 1997). Since right-

hemisphere dominance is also hypothesised as one of the biological bases of creativity (Martindale, 19999, this might the common background responsible for the relationship.

The fifth research question was posed about the relationship of language proficiency and task performance measures on the cognitively less and more complex tasks. Although, perhaps not surprisingly, the level of proficiency strongly determines task performance, different factors seem to play a greater role in the case of the two tasks. In the case of the cognitively less complex task, more proficient learners talked more, they displayed greater accuracy, and they also seemed to be able to cope with task requirements better. This means that they invented stories containing more events, and they used more difficult words, that is, they probably used the specific words that the task called for. In the case of the cognitively more complex task, more proficient learners displayed a greater range of vocabulary, and they were more fluent. Therefore, proficiency seemed to have a differential effect in the case of the two tasks differing in cognitive complexity.

The sixth research question referred to the relationship of language aptitude and task performance measures in the case of the cognitively less and more complex tasks and it brought interesting results. Although the relationship between language aptitude and English proficiency was not strong at the level of general measures, participants characterised by a higher level of aptitude tended to behave in a manner similar to more proficient students. Despite the fact that the relationships were more moderate, the pattern was similar: greater aptitude seemed to be related to greater accuracy on the cognitively less complex task, greater fluency on the cognitively more complex task and more talk, and greater lexical variety on both tasks. Contrary to the claims of the Cognition Hypothesis (Robinson, 2001c, 2003, 2005b), participants' performance on the cognitively more complex task seemed to be less affected by their language aptitude.

The last research question concerned the relationship of learner creativity and task performance measures on the cognitively less and more complex tasks. First of all, it was revealed that out of the four parts of the creativity test, there are only two subtests in which significant correlations can

be detected between creativity and oral narrative task performance. Although one of these is a verbal while the other is a drawing task, they seem to be similar in the sense that they both involve the quick generation of a large number of novel ideas. Another important finding is that the relationships between creativity and task performance measures tend to be moderate; moreover, performance on the cognitively more complex task seems to be more strongly affected by creativity. Although this finding is line with the claims of the Cognition Hypothesis (Robinson, 2001c, 2003, 2005b) stating that individual differences play a greater role in the case of cognitively more complex tasks, an alternative interpretation is also possible here. It can be argued that the reason why there is a stronger relationship between creativity and task performance measures in the case of the cognitively more complex task is that this particular task provided a greater opportunity for participants to use their imagination.

The majority of the moderate positive relationships that can be detected between learner creativity and oral narrative task performance exist between fluency-free components of creativity, that is, average originality and relative flexibility, and the task performance measures of fluency, quantity of talk and lexical variety. It was hypothesised that either a larger lexicon containing more unusual or rare items, or better access to its elements could be in the background of the relationships detected. This finding is also in line with Carroll's (1993) interpretation of creativity, or as he labelled it general retrieval ability, which is conceived as an ability to retrieve unusual concepts with ease.

## 9.3 Limitations

When drawing conclusions on the basis of the findings, certain limitations must be kept in mind that warrant caution with regard to the interpretation of results. One aspect of the research that restricts the generalizability of the findings concerns the sample. The sample used in the study is not representative of the Hungarian population, and not even of that portion of the population who study languages. It is a specific sample as it comprises English majors, students who are expected to conduct their

university studies in English and some of whom will probably become English teachers, translators or interpreters. Conclusions of the study can only be regarded valid for this population.

Another concern that should be pointed out in connection with the sample is its size. Although the labour-intensive nature of the analysis of the tests and tasks used made working with a larger sample unfeasible, and task-based research is known to employ similar or even smaller samples, having more participants could have resulted in more statistically significant findings. There were a few tendency-level relationships identified in the course of research, and I expect that these might have been significant on a larger sample.

Besides limitations that must be observed because of characteristics of the sample, others must be kept in mind as a consequence of certain features of the instruments. The test of creativity turned out to be a particularly problematic instrument, as it was revealed that the four sub-tests comprising it do not measure the same construct. The lack of a unified construct operating in the background poses questions about the validity of the instrument and makes the usage of composite scores unjustifiable. Although it is possible that out of the four slightly different aspects of creativity only two, which are believed to trigger the generation of a large number of novel and interesting responses, are related to performance on oral narrative tasks, this hypothesis needs to be substantiated by further research.

An additional limitation concerns possible mismatches of aptitude theory and aptitude test. As tests are known to reflect the theory behind the construct they intend to measure, it might be that certain relationships between constructs remain hidden because the tests used for assessment measure only a limited aspect of the given construct. Therefore, it is possible that the positive relationship which was hypothesised to exist between creativity and language aptitude on the basis of the literature went undetected because the aptitude test used in the study failed to measure those aspects of language aptitude that would have been relevant for creativity. However, since the aptitude test used in the study is the only such test available for native speakers of Hungarian, it would have been impossible to choose a better instrument for this purpose.

Furthermore, this research conducted on the relationship of learner creativity and oral narrative task performance probably would have benefited from using a proficiency test designed to test oral skills specifically. Since the administration and evaluation of such tests theoretically involves at least who examiners, the testing of oral skills was regarded unfeasible in this study.

Finally, despite the fact that figures still seem to suggest that the three facets of creativity examined are related to task performance in a differential way, which is the same conclusion that was drawn from the exploratory study, results of the exploratory study are not totally compatible with findings of this study. Although some of the discrepancies can probably be ascribed to previously undetected problems of the measuring instruments and also to differences in some of the measures used in the two studies, the possibility that some of the results are artefacts cannot be ruled out entirely either.

## 9.4 Pedagogical implications and directions for future research

Direct pedagogical relevance of this study would have been assured by the fact that an oral narrative task very similar to the cognitively more complex task used here used to be part of the comprehensive language examination of first-year English majors at ELTE. Although because of recent changes in the system of training, the first-year comprehensive exam no longer contains this task, findings of the study might still be applicable for other language examinations employing a similar task type.

For those intending to use oral narrative tasks as part of a test of oral skills, it would probably be important to know that performance on the cognitively less complex version seems to be determined by proficiency to a greater extent, whereas ID variables probably play a greater role while solving the cognitively more complex task. The cognitively more complex task used in the study is likely to result in more accurate performance, while the cognitively less complex task urges learners to be more fluent and use a wider range of vocabulary. Moreover, even the knowledge of specific items of vocabulary can be tested with the help of a cartoon strip task.

On a theoretical level, findings of the study can be considered important as they seem to support claims of the Cognition Hypothesis (Robinson, 2001c, 2003, 2005b) as opposed to Skehan's (1998) framework of task difficulty. While Skehan predicts that due to an increase in processing load, greater cognitive complexity automatically brings about a drop in accuracy, fluency and complexity, Robinson believes that certain ways of increasing cognitive complexity along resource directing dimensions result in greater accuracy and complexity. The mixed results of the study can be interpreted in a way that they were brought about by the synergic effects of both resource-directing and resource-dispersing features of the task.

Since Robinson's (2001c, 2003, 2005b) Cognition Hypothesis as well as his Triadic Componential Framework (Robinson 2001b, 2005b, 2007a) are pedagogically motivated, that is, their aim is to aid pedagogical decisions concerning the sequencing of tasks in syllabi, my research bears indirect consequences regarding syllabus design as well. Through providing empirical support for the Cognition Hypothesis and the Triadic Componential Framework, my findings seem to substantiate that these theories can and probably should be used for making pedagogical decisions when designing syllabi.

Besides the above described implications which are applicable to testing and syllabus design, the relationships of creativity and proficiency and creativity and task performance have some pedagogical relevance as well. First of all, results of this study are not entirely consistent with findings of an earlier exploratory study (Albert & Kormos, 2004); therefore, further research would be needed to clarify the relationship of variables before arriving at far reaching conclusions. As a general trend, it can be stated that the three components of creativity: average originality, relative flexibility and creative fluency, seem to relate to language measures in a differential manner.

As regards the connections of creativity with language proficiency, it seems that average originality might be linked to language proficiency, while relative flexibility might be advantageous when being tested by certain types of tasks. Considering the relationships of creativity with task performance measures, it is again greater average originality and relative flexibility that

tend to be connected to superior task performance. Although the magnitude of connections seems to be moderate, they still suggest an existing relationship. Since creative students seem to be at an advantage when performing certain types of tasks, especially less structured and cognitively more complex ones, we should bear this in mind when using such tasks in the classroom and especially when using them for testing purposes. In order to ensure the fair treatment of students, emphasis should be laid on also employing tasks that are more structured, therefore, do not draw so heavily on learner creativity.

The research reported here certainly leaves many questions open and calls for further research. Besides replication studies that would shed light on which specific facet of creativity is linked to language proficiency and oral narrative task performance, the relationships of language aptitude and creativity should also be further explored. Although based on the literature one common point was certainly identified, that is the ability to cope with novelty, the results of the study suggest that some aspect of the two constructs might be incompatible. The relative right hemisphere dominance reflected by a visual-spatial mode of thinking characteristic of creative (Martindale, 1999) and also of dyslexic individuals (West, 1997), for example, appears to be less unfavourable when it comes to learning languages.

# References


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

Alderson, J. C. (2000). *Assessing reading*. Cambridge, UK: Cambridge University Press.

Alderson, C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge, UK: Cambridge University Press.

Alpert, R., & Haber, R. N. (1960). Anxiety in academic achievement situations. *Journal of Abnormal Psychology, 61*, 207-215.

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

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

Armbruster, B. B. (1989). Metacognition in creativity. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 177-182). New York: Plenum Press.

Aston, G. (1986). Trouble shooting in interaction with learners: The more the merrier? *Applied Linguistics, 7*, 128-143.

Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.

Baer, J. (1993). *Creativity and divergent thinking: A task specific approach*. Hillsdale, NJ: Lawrence Erlbaum.

Barkóczi, I. (1991). Átstrukturálást sugalmazó ingerek intuitív hatása [The intuitive effect of stimuli suggesting restructuring]. *Pszichológia, 11*, 483-502.

Barkóczi, I. (1994a). Az analógiás transzfer szerepe belátásos problémamegoldásban, a pszichometriai kreativitás függvényében [The role of analogic transfer in solving problems requiring insight as a function of psychometric creativity]. *Magyar Pszichológiai Szemle, 34*, 321-332.

Barkóczi, I. (1994b). A rendelkezésre álló információ felhasználása belátásos problémák megoldásában a pszichometriai kreativitás függvényében [Using available information for solving problems requiring insight as a function of psychometric creativity]. *Magyar Pszichológiai Szemle, 34*, 27-38.

Barkóczi, I., & Zétényi, T. (1981). *A kreativitás vizsgálata* [The examination of creativity]. Budapest, Hungary: Országos Pedagógiai Intézet.

Brock, C. (1986). The effects of referential questions on classroom discourse. *TESOL Quarterly, 20*, 47-59.

Brown, G., Anderson, A., Shilcock, R., & Yule, G. (1984). *Teaching talk: Strategies for production and assessment*. Cambridge, UK: Cambridge University Press.

Brown, G., & Yule, G. (1983). *Teaching the spoken language: An approach based on the analysis of conversational English*. Cambridge, UK: Cambridge University Press.

Brown, H. D. (1987). *Principles of language learning and teaching*. Englewood Cliffs, NJ: Prentice Hall.

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.

Bygate, M. (1999). Quality of language and purpose of task: Patterns of learners' language on two oral communication tasks. *Language Teaching Research, 3*, 185-214.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Candlin, C. (1987). Towards task based language teaching. In C. Candlin & D. Murphy (Eds.), *Language learning tasks*. Englewood Cliffs, NJ: Prentice Hall.

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 119-155). Rowley, MA: Newbury House.

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. S. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11-29). Englewood Cliffs, NJ: Prentice Hall.

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

Carroll, J. B. & Sapon, S. M. (1959). *Modern Language Aptitude Test. Form A. Manual + Test.* New York: The Psychological Corporation.

Carson, S. H., Peterson, J. B., & Higgins, D. M. (2003). Decreased latent inhibition is associated with increased creative achievement in high-functioning individuals. *Journal of Personality and Social Psychology, 85*, 499-506.

Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1997). Direct approaches in L2 instruction: A turning point in communicative language teaching? *TESOL Quarterly, 31*, 141-152.

Clark, H. H., & Clark, E. (1977). *Psychology and language: An introduction to psycholinguistics.* New York: Harcourt, Brace, Jovanovich.

Cohen, A. D. (1998). *Strategies in learning and using a second language*. Harlow, UK: Longman.

Cromer, R. (1974). The development of language and cognition: The cognition hypothesis. In B. Foss (Ed.), *New perspectives in child development* (pp. 184-252). Harmondsworth, UK: Penguin.

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

Czigler, I. (1992). *A figyelem kognitív pszichológiája* [The cognitive psychology of attention]. Budapest: Tankönyvkiadó.

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.

Csölle, A., & Károly, K. (1998). Learning from errors: A task-based approach. *NovELTy, 5,* (3), 19-36.

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

Dewaele, J.-M., & Pavlenko, A. (2003). Productivity and lexical diversity in native and non-native speech: A study of cross-cultural effects. In V. Cook (Ed.), *The effects of the second language on the first* (pp. 120-141). Clevedon, UK: Multilingual Matters.

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

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

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research, 4,* 275-300.

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

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

Duff, P. (1986). Another look at interlanguage talk: Taking task to task. In: R. Day (Ed.), *Talking to learn* (pp. 147-181). Rowley, MA: Newbury House.

Ehrman, M. E., & Leaver, B. L. (2003). Cognitive styles in the service of language learning. *System, 31,* 391-415.

Ehrman, M. E., & Oxford, R. L. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal, 79*, 67-89.

Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp.287-314). Ann Arbour, MI: The Univerisity of Michigan Press.

Ellis, R. (1994). *The study of second language acquisition*. Oxford, UK: Oxford University Press.

Ellis, R. (2004). Individual differences in second language learning. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 525-551). Oxford, UK: Blackwell.

Espinosa, S. M. (2005). Can P_Lex accurately measure lexical richness in the written production of young learners of EFL? *Porta Linguarum, 4*, (June) 7-21.

Flowers, J. H., & Garbing, C. P. (1989). Creativity and perception. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 147-162). New York: Plenum Press.

Foster, P. (1998). A classroom perspective on the negotiation of meaning. *Applied Linguistics, 19*, 1-23.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*, 299-323.

Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research, 3*, 215-247.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*, 354-375.

Freed, B. F. (2000). Is fluency, like beauty, in the eyes of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp.243-265). Ann Arbour, MI: The Univerisity of Michigan Press.

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

Gardner, H. (1993). *Creating minds*. New York: Basic Books.

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

Gardner, R. C., & MacIntyre, P. D. (1992). A student's contributions to second language learning: Part 1: Cognitive variables. *Language Teaching, 25,* 211-220.

Gardner, R. C., & MacIntyre, P. D. (1993). A student's contributions to second language learning: Part 2: Affective variables. *Language Teaching, 26,* 1-11.

Gass, S. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics, 9,* 198-217.

Gass, S., & Varonis, E. (1985). Task variation and NNS/NNS negotiation of meaning. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 134-156). Rowley, MA: Newbury House.

Gáspár, M. (2001). A divergens gondolkodás faktorainak interindividuális stabilitása 9-14 éves korban [The interindividual stability of factors of divergent thinking between the ages 9-14]. *Magyar Pszichológiai Szemle, 56,* 411-426.

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

Ghiselin, B. (1952). *The creative process*. New York: Bantam.

Givón, T. (1989). *Mind, code and context: Essays in pragmatics*. Hillsdale, NJ: Erlbaum.

Gough, H. G. (1979). A creative personality scale for the Adjective Check List. *Journal of Personality and Social Psychology, 37,* 1398-1405.

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.

Hargreaves, D. J., & Bolton, H. (1972). Selecting creativity tests for use in research. *British Journal of Psychology, 63,* 451-462.

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.

Hatch, A., & Lazaraton, E. (1991). *The research manual. Design and statistics for applied linguistics.* Boston: Heinle and Heinle.

Hatch, E. (1992). *Discourse and language education*. Cambridge, UK: Cambridge University Press.

Hayes, J. R. (1989). Cognitive processes in creativity. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 135-145). New York: Plenum Press.

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

Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information processing approach to task design. *Language Learning, 51,* 401-436.

Jarvis, S. (2002). Short texts, best-fitting curves, and new measures of lexical diversity. *Language Testing 19,* 57-84.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of ESL. *Cognitive Psychology, 21,* 60-99.

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.

Kárpáti, A. (1996). A Leonardo program pedagógiai hatásvizsgálata [Studying the pedagogical effects of the Leonardo program]. *Magyar Pedagógia, 96*, 3-34.

Kellerman, E. (1991). Compensatory strategies in second language research: A critique, a revision, and some (non)implications for the classroom. In R. Phillipson, E. Kellerman, L. Selinker, M. Sharwood-Smith, & M. Swain (Eds.), *Foreign/second language pedagogy research* (pp. 142-161). Clevedon, UK: Multilingual Matters.

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

Kolb, D. A. (1984). *Experiential learning: Experience as a source of learning and development.* Englewood Cliffs, NJ: Prentice Hall.

Kontra, E., & Kormos, J. (2006). Strategy use and the construct of C-tests. In R. Grotjahn (Ed.), *Der C-test: Theorie, Empirie, Anwendungen* (pp. 121-138). Frankfurt am Main, Germany: Peter Lang.

Kontráné Hegybíró, E. (2004) Egyéni különbségek a nyelvtanulásban [Individual differences in language learning]. In E. H. Kontra & J. Kormos (Eds.), *A nyelvtanuló: Sikerek, kudarcok, stratégiák* (pp. 8-24). Budapest, Hungary: Okker Kiadó.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*, 145-164.

Kormos, J., & Sáfár, A. (2006). A munkamemória és nyelvérzék szerepe az intenzív nyelvtanulásban [The role of working memory and language aptitude in intensive language learning]. *Magyar Pszichológiai Szemle, 61*, 557-580.

Krashen, S. (1982). Accounting for child-adult differences in second language rate and attainment. In S. Krashen, R. Scarcella, & M. Long (Eds.), *Child-adult differences in second language acquisition* (pp 202-226). Rowley, MA: Newbury House.

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

Kubie, L. S. (1958). *Neurotic distortion of the creative process*. Lawrence: University of Kansas Press.

Kürti, J. (1985). *Kreativitásfejlesztés kisiskoláskorban* [Developing creativity in primary school children]. Budapest, Hungary: Tankönyvkiadó.

Labov, W. (1972). *Language in the inner city: Studies in the Black English vernacular*. Oxford, UK: Basil Blackwell.

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. New York: Longman.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*, 255-271.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*, 387-412.

Liskin-Gasparro, J. E. (1996). Narrative strategies: A case study of developing storytelling skills by a learner of Spanish. *The Modern Language Journal, 80*, 271-286.

Long, M. (1985). A role for instruction for second language acquisition: Task based language training. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition*. Clevedon, UK: Multilingual Matters.

Long, M. (1989). Task, group and task-group interaction. *University of Hawaii Working Papers in English as a Second Language, 8,* 1-26.

Long, M. H., & Crookes, G. (1993). Units of analysis in syllabus design: The case for task. In: G. Crookes & S. Gass (Eds.), *Tasks in a pedagogical context: Integrating theory and practice* (pp. 9-54). Clevedon, UK: Multilingual Matters.

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

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* (pp. 24-45). Boston: McGraw-Hill.

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

Malvern, D. D., & Richards B. J. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp. 58-71). Clevedon, UK: Multilingual Matters.

Martindale, C. (1999). Biological bases of creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp.137-152). New York: Cambridge University Press.

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

McCarthy, M., & Carter, R. (1994). *Language as discourse: Perspective for language teaching*. New York: Longman Publishing.

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 Web site: http://www.grads.vt.edu/igss/faculty_staff/toefl_interp.html

Meara, P. (2001). *Plex v1.1*. Retrieved April 21, 2006, from University of Wales Swansea Web site: http://www.swan.ac.uk/

Meara, P., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect, 16*, 323-337.

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

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition, 20*, 83-108.

Meisel, J. (1987). Reference to past events and actions in the development of a natural second language acquisition. In C. W. Pfaff (Ed.), *First and second language acquisition processes* (pp. 206-224). Cambridge, MA: Newbury House.

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge, UK: Cambridge University Press.

NYEK Kutatócsoport (2004). *Jelentés a nyelvi előkészítő évfolyamos tanulók körében a 2004/2005. tanév őszi félévében elvégzett felmérésről angol és német nyelvből* [Report of the study conducted in the academic year 2004/2005 autumn semester with students attending language preparatory courses in English and German]. Retrieved May 3, 2006, from the Oktatási Minisztérium [Hungarian Ministry of Education] Web site: http://www.om.hu/letolt/vilagnyelv/om_nyek_jelentes_2004_osz.pdf

Oláh, A. (1987). Creativity and personality variables. In L. Kardos, Cs. Pléh, & I. Barkóczi (Eds.), *Studies in creativity* (pp. 87-108). Budapest: Akadémiai Kiadó.

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition.* New York: Cambridge University Press.

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* [Uniform Hungarian Language Aptitude Test]. Kaposvár, Hungary: Mottó-Logic.

Ottó, I., & Nikolov, M. (2003). Magyar felsőoktatási intézmények elsőéves hallgatóinak nyelvérzéke [The language aptitude of first year university students in Hungary]. *Iskolakultúra, 13,* (6-7), 34-44.

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know.* New York: Newbury House.

Oxford, R. L., & Ehrman, M. (1993). Second language research on individual differences. *Annual Review of Applied Linguistics, 13*, 188-205.

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.

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

Peterson, J. B., Smith, K. W., & Carson, S. H. (2002). Openness and extraversion are associated with reduced latent inhibition: Replication and commentary. *Personality and Individual Differences, 33,* 1137-1147.

Phillips, D. (1989). *Longman Practice Tests for TOEFL*. New York: Longman.

Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communication tasks for second language instruction and research. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 9-34). Clevedon, UK: Multilingual Matters.

Pica, T., Young, R., & Doughty, C. (1987). The impact of interaction on comprehension. *TESOL Quarterly, 21*, 737-758.

Pimsleur, P. (1966). *Pimsleur Language Aptitude Battery and Manual*. New York: Harcourt, Brace, Jovanovich.

Plough, I., & Gass, S. M. (1993). Interlocutor and task familiarity: Effects on interactional structure. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 35-56). Clevedon, UK: Multilingual Matters.

Read, J. (2005). Applying lexical statistics to the IELTS speaking test. *Research Notes, 20*, 12-16.

Reid, J. M. (1987). The learning style preferences of ESL students. *TESOL Quarterly, 21*, 87-111.

Reid, J. M. (1995). Preface. In J. M. Reid (Ed.), *Learning styles in the ESL/EFL classroom* (pp. viii-xvii). Boston: Heinle and Heinle.

Richards, B. (1987). Type/token ratios: What do they really tell us? *Journal of Child Language, 14*, 201-209.

Riding, R., & Rayner, S. G. (1998). *Cognitive styles and learning strategies: Understanding style differences in learning and behaviour*. London: David Fulton.

Riggenbach, H. (1991). Towards an understanding of fluency: A microanalysis of nonnative speaker conversation. *Discourse Processes, 14*, 423-441.

Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning, 45*, 99-140.

Robinson, P. (2001a). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research, 17*, 368-392.

Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interaction in a componential framework. *Applied Linguistics, 22*, 27-57.

Robinson, P. (2001c). Task complexity, cognitive resources and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 287-318). Cambridge, UK: Cambridge University Press.

Robinson, P. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies, 21,* 45-107.

Robinson, P. (2005a). Aptitude and second language acquisition. *Annual Review of Applied Linguistics, 25,* 46-73.

Robinson, P. (2005b). Cognitive complexity and task sequencing: A review of studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching, 43* (1), 1-33.

Robinson, P. (2007a). Criteria for grading and sequencing pedagogic tasks. In M. P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7-27). Clevedon: Multilingual Matters.

Robinson, P. (2007b). Task complexity, theory of mind, and intentional reasoning: Effects on L2 production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching, 45,* 237-257.

Robinson, P., & Gilabert, R. (2007). Task complexity, the cognition hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching, 45,* 161-176.

Rogers, C. R. (1954). Toward a theory of creativity. *ETC: A review of general semantics, 11,* 249-260.

Runco, M. A. (1985). The reliability and convergent validity of ideational flexibility as a function of level of academic achievement. *Perceptual and Motor Skills, 61,* 1075-1081.

Runco, M. A. (1986). The discriminant validity of gifted children's divergent thinking test scores. *Gifted Child Quarterly, 30,* 78-82.

Runco, M. A. (1989). Parents' and teachers' ratings of the creativity of children. *Journal of Social Behavior and Personality, 4,* 73-83.

Runco, M. A., & Albert, R. S. (1985). The reliability and validity of ideational originality in the divergent thinking of academically gifted and no gifted children. *Educational and Psychological Measurement, 45,* 483-501.

Runco, M. A., Okuda, S. M., & Thurston, B. J. (1987). The psychometric properties of four systems for scoring divergent thinking tests. *Journal of Psychoeducational Assessment, 5*, 149-156.

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

Salkind, N. J. (2004). *Statistics for people who (think they) hate statistics*. Thousand Oaks, CA: Sage.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. New York: Newbury House/Harper and Row.

Seliger, H. W., & Shohamy, E. (1989). *Second language research methods*. Oxford, UK: Oxford University Press.

Skehan, P. (1989). *Individual differences in second language learning*. London: Edward Arnold.

Skehan, P. (1996). A framework for the implementation of task based instruction. *Applied Linguistics, 17*, 38-62.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, UK: Oxford University Press.

Skehan, P. (2001). Tasks and language performance assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 167-185). London: Pearson Education.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*, 185-211.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*, 93-120.

Snow, C. (1983). Age differences in second language acquisition: Research findings and folk psychology. In K. M. Bailey, M. Long, & S. Peck (Eds.), *Second language acquisition studies* (pp. 141-150). Rowley, MA: Newbury House.

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

Spielberger, C. D. (1966). *Anxiety and behavior*. New York: Academic Press.

Stein, B. S. (1989). Memory and creativity. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 163-176). New York: Plenum Press.

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

Sternberg, R. J. (1985b). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge, UK: Cambridge University Press.

Sternberg, R. J. (1997). A triarchic view of giftedness: Theory and practice. In N. Coleangelo & G. A. Davis (Eds.), *Handbook of Gifted Education* (pp. 43-53). Boston: Allyn and Bacon.

Sternberg, R. J. (2002). The theory of successful intelligence and its implications for language aptitude testing. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 13-43). Amsterdam: John Benjamins.

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. (1995). *Defying the crowd: Cultivating creativity in a culture of conformity*. New York: Free Press.

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

Sternberg, R. J., & Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 3-15). New York: Cambridge University Press.

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 Hall.

Torrance, E. P. (1966). *Torrance tests of creative thinking: Research edition*. Princeton, NJ: Personnel Press.

Tóth, P. (2007). A problémamegoldó gondolkodás fejlesztése az informatikaórán [Developing thinking to enhance problem solving in informatics classes]. In A. Benedek (Ed.), *Új tendenciák a képzők képzésében: Konferenciakiadvány 2006 november 18* (pp. 57-66). Budapest, Hungary: BME Alkalmazott Pedagógia és Pszichológia Intézet.

VanPattern, B. (1990). Attending to content and form in the input: An experiment in consciousness. *Studies in Second Language Acquisition, 12*, 287-301.

Wagner, R. K., Muse, A. E., & Tannenbaum, K. R. (2007). *Vocabulary acquisition: Implications for reading comprehension*. New York: The Guilford Press.

Ward, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology, 27*, 1-40.

West, T. G. (1997). *In the mind's eye: Visual thinkers, gifted people with dyslexia and other learning difficulties, computer images and the ironies of creativity*. New York: Prometheus Books.

Witkin, H. A., Goodenough, D. R., & Karp, S. A. (1967). Stability of cognitive style from childhood to young adulthood. *Journal of Personality and Social Psychology, 7*, 291-300.

Zétényi, T. (1989). *A kreativitástesztek tesztkönyve I-II* [Creativity test booklet I-II]. Budapest, Hungary: Munkaügyi Kutatóintézet.

# Appendix A -Single picture task A

Your task is to tell your partner a story about this picture. The picture can be the beginning, the middle, or the end of your story. You have five minutes to think before you start.

# Appendix B - Single picture task B

Your task is to tell your partner a story about this picture. The picture can be the beginning, the middle, or the end of your story. You have five minutes to think before you start.

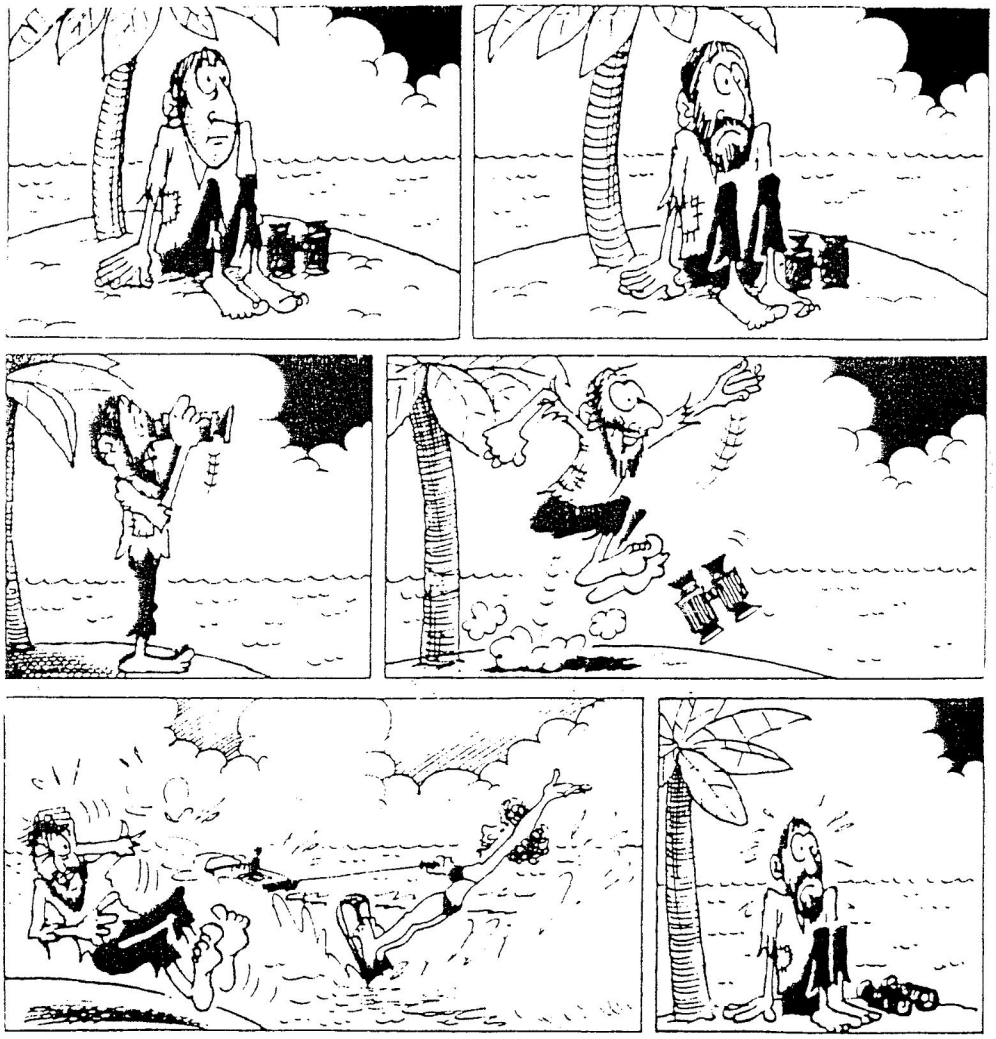
# Appendix C - Picture sequence task A

You will find some pictures in the envelope. Your task is to arrange the pictures in any order you like, then to tell your partner the story shown by them. You must use all the pictures from the envelope, but you may also add extra stages. You have five minutes to think before you start.

# Appendix D - Picture sequence task B

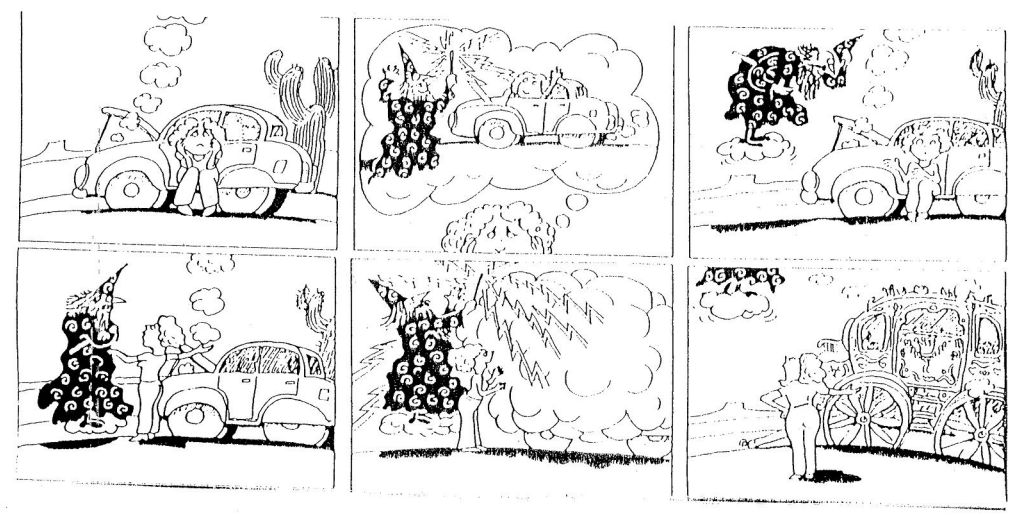
You will find some pictures in the envelope. Your task is to arrange the pictures in any order you like, then to tell your partner the story shown by them. You must use all the pictures from the envelope, but you may also add extra stages. You have five minutes to think before you start.

# Appendix E - Cartoon strip task A



Your task is to tell your partner the story of this cartoon strip. You may also add stages not shown by the pictures. You have five minutes to think before you start.

# Appendix F - Cartoon strip task B



Your task is to tell your partner the story of this cartoon strip. You may also add stages not shown by the pictures. You have five minutes to think before you start.

# Appendix G - Post-task interview

1. Egészében véve milyennek találtad a feladatot?
2. Mi az, ami leginkább tetszett benne?
3. Mi az, amit könnyűnek találtál?
4. Mi az, ami legkevésbé tetszett benne?
5. Mi az, amit nehéznek találtál?
6. Csináltál már hasonló feladatot?

ahol releváns: Kitalált történeted mondtál el vagy egy megtörtént esetet?

7. Mi volt az instrukció? Mire szólított fel? Foglald össze néhány szóban!
8. Világos volt számodra az instrukció megfogalmazása? Könnyen érthetőnek találtad?
9. Mi a véleményed a felkészülésre szánt időről? Számodra ez túl sok/ túl kevés/ éppen megfelelő volt?
10. Miből állt maga a feladat? Mit kellett csinálni?
11. Szerinted mit szeretettem volnal én, a kutató elérni, mi volt az én célom ezzel a feladattal?

1. What do you think of the task in general?
2. What did you like most about it?
3. What did you find easy?
4. What did you like least about it?
5. What did you find difficult?
6. Have you ever done a similar task?

where relevant: Was your story an imaginary or a true one?

7. What was the instruction? What did it ask you to do? Summarize it in a few words!
8. Was the wording of the instruction clear for you? Was it easy to understand?
9. What do you think of the preparation time? Was it too much / too little / just enough for you?
10. What was your task? What did you have to do?
11. What do you think my aim as a researcher was with this task? What did I want to achieve?

# Appendix H - Evaluation criteria for the oral narrative task

on the basis of Hatch (1992), McCarthy and Carter (1994), and Liskin-Gasparro (1996) a narrative should consist of the following parts:

1. **Abstract** - What is the story going to be about?

- can be missing
- serves as a title of the story

2. **Orientation** - Who are the participants? When and where did the action take place? In which circumstances?

- narratives can have more than one layer of orientation
- introduces time, spatial setting of the story, the characters and their roles
- in English narratives copula sentences (use of be), presentatives (there is/there are sentences), and identifying or descriptive relative clauses are often used for this purpose
- nothing happens here, no action is passing from one person to another, the verbs are statives or intransitives most of the time

3. **Complicating action** - Then what happened? What problems occurred? (goal and problem, steps to resolve the problem)

- narratives can have more than one layer of complicating action
- after completing the story world setting, the story line is set up: a hero with a goal, who is prevented from the easy attainment of that goal by some problems, thus the hero develops a plan for solving these problems
- focus is on the hero (actor) and on the actions the hero uses to solve problems so that the goal can be achieved, shows how the hero works out the problem to reach the goal
- usually consists of a set of action clauses arranged in a temporal order
- the actions in the story line are typically highly transitive verbs, actions are taken and completed, the action is strongly transmitted from the agent to the objects

4. **Resolution** - How did events sort themselves out? What finally happened?

- shows the goal attained

5. **Evaluation** - What is the point of the story? So what?

- can usually be found throughout the narrative, woven into the story line
- may be summarised in the moral (coda)
- can be phrased as bracketed asides
- they serve to involve the audience more fully into the story

**Evaluation devices in Hatch (1992)**

a. non-verbal gestures, expressive intonation
b. lexical intensifiers
c. repetitions
d. mimicking or direct quotes
e. direct evaluative pointers
f. rhetorical questions to the listener
g. relative clauses or other embedded asides

**Evaluation devices in Liskin-Gasparro (1996)**

1. comment on action - brief or long
2. reference to previous action
3. ironic aside
4. retarding narrative actions: gerunds, progressive constructs
5. retarding narrative action: juxtaposition of narrative and descriptive clauses
6. direct speech
7. repetition: lexical and/or syntactic
8. contrast
9. expressive phonology
10. lexical choice

6. **Coda** - What is the bridge between the events in the story and the present situation of the narration?

- can be missing
- contains a moral that summarises or evaluates the story's relevance

# Appendix I - Creativity test

Név: férfi nő

Születési év:

A most következő feladatok jellemzője az, hogy nincs egyetlen helyes megoldásuk. Itt az a jó, ha minél többféle, minél érdekesebb válaszokat adnak. Próbáljanak olyan válaszokat kitalálni, amelyek feltehetően senki másnak nem jutnának eszébe. Fogják fel játékosan, legyenek ötletesek, eredetiek.

Jó szórakozást a munkához!

Kérjük fejezze be az alábbi megkezdett mondatokat. Írjon annyi befejezést, amennyit csak tud. Lásson munkához, 3 perc áll rendelkezésére!

AZ ÖREGEMBER AZ UTCÁN VÉGIG .....

A KÖRÚTI HÁZ SÖTÉT KAPUJÁBAN .....

AZ AUTÓ KEREKEI AZ ORSZÁGÚTON .....

Kérjük, képzeljen el és írjon le az alábbi tárgyakkal kapcsolatban annyi, a megszokottól eltérő használati lehetőséget, amennyit csak tud.

Lásson munkához, 5 perc áll rendelkezésére.

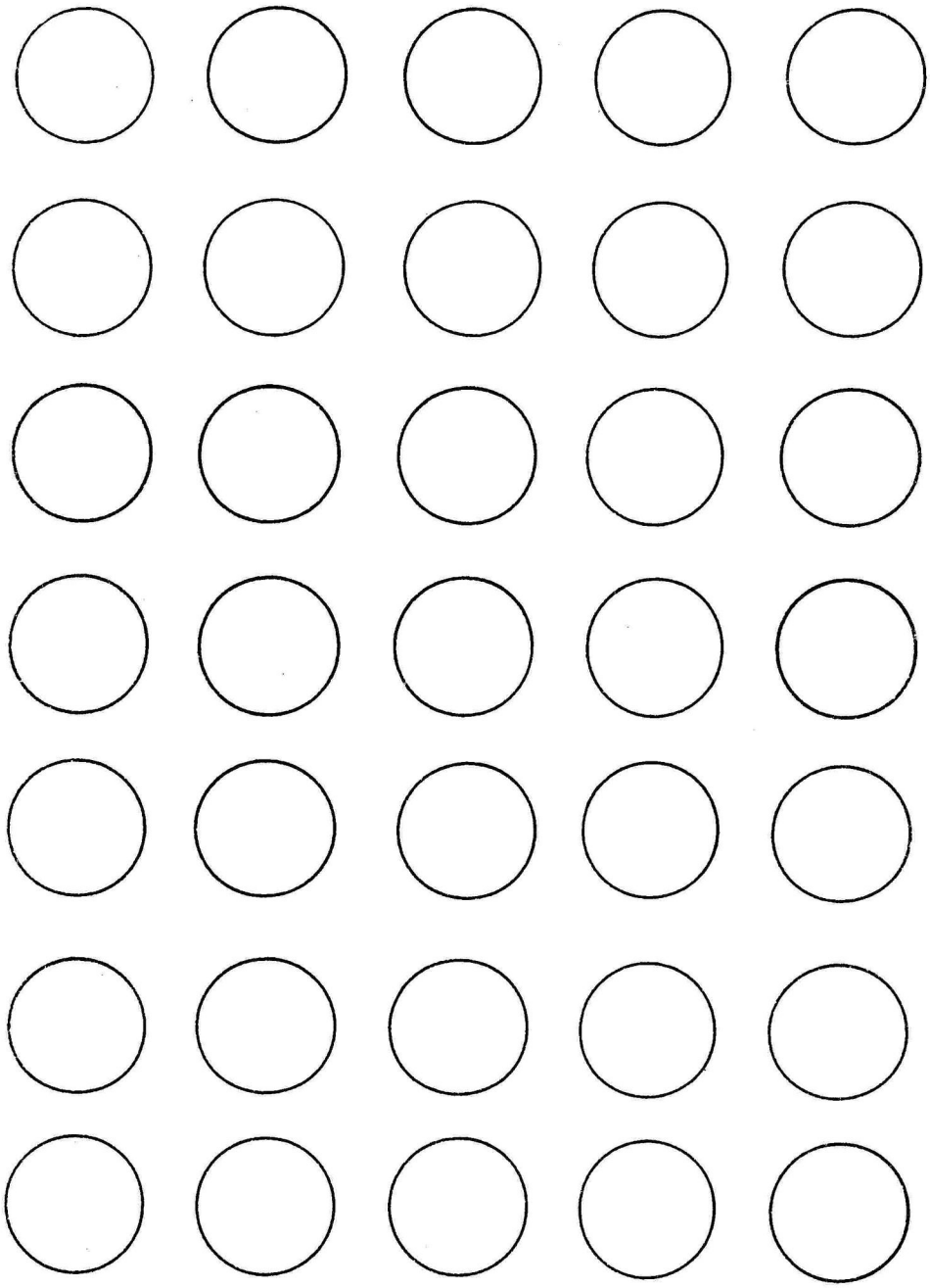
TÉGLA

CERUZA

GYUFA

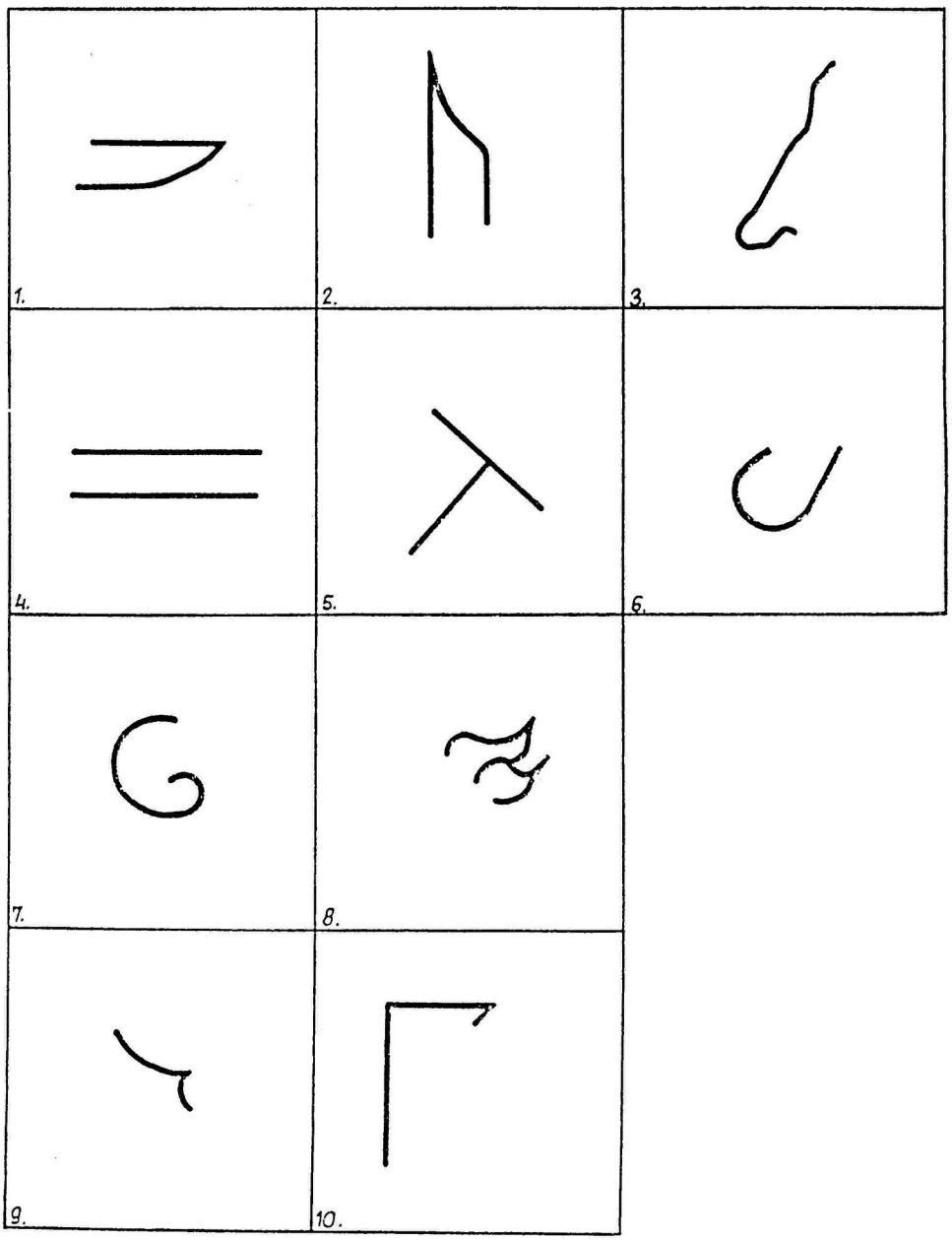
Az itt található körökből készítsen ábrákat, alakokat, rajzokat úgy, hogy a kör a készítendő rajz lényeges elemét képezze. Húzhat kiegészítő vonalakat a körön kívül és belül is. Ha egy-egy rajzzal elkészült, akkor írja alá, hogy mit ábrázol. Igyekezzen érdekes rajzokat készíteni.

Most lásson munkához, 8 perc áll rendelkezésére.



Az alábbi vonalas rajzokból készítsen érdekes ábrákat, alakokat, stb. Ha egy-egy rajzzal elkészült, próbáljon neki érdekes címet adni. Lehetőleg az összes figurát egészítse ki, erre 10 perc áll rendelkezésére.

Most lásson munkához!



Kérjük írjon az alábbi két-két szó közé egy olyan harmadikat, amely jelentésbelileg kapcsolódik mindkettőhöz. Írjon minél több változatot, amennyi csak eszébe jut.

Lásson munkához, 6 perc áll rendelkezésére.

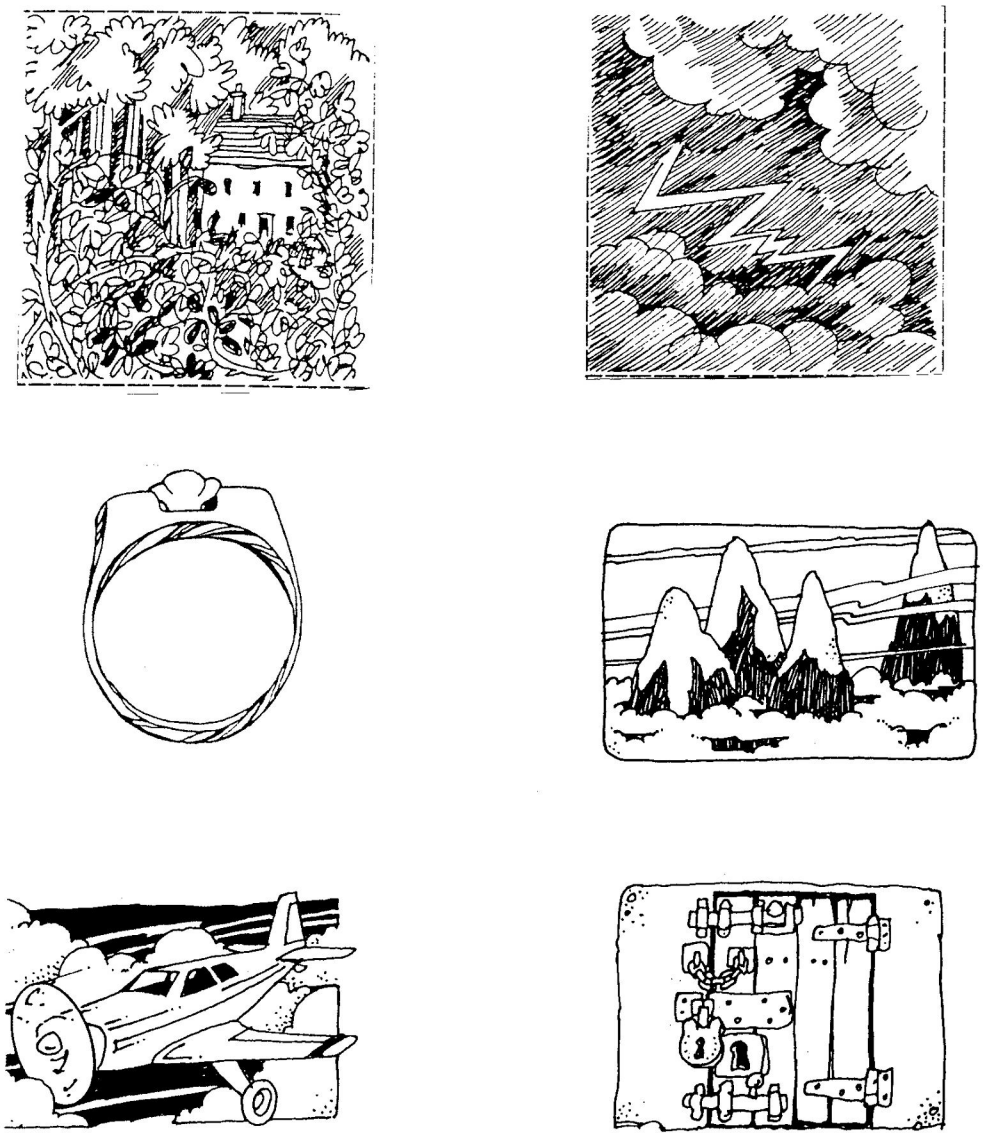
| | | | |
|---|---|---|---|
| Például: | ÁGYÚ | dörög<br>légvédelem<br>repülőgép | ÉG |

TÜKÖR ESŐ

GYÜMÖLCS CSÓK

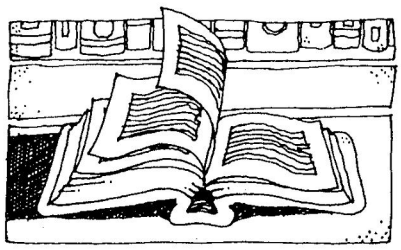
TORONY CERUZA

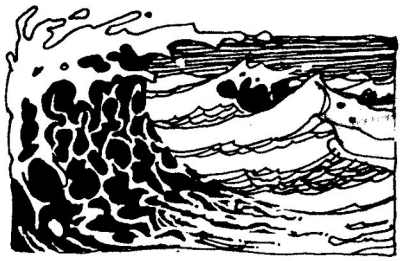
# Appendix J - New picture sequence task A

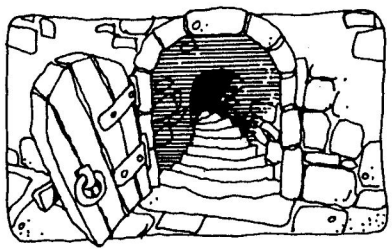


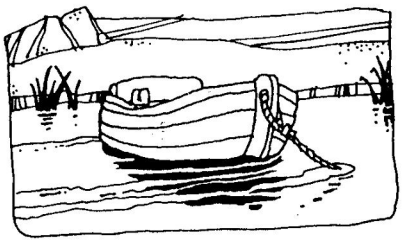
You will find six pictures in the envelope. Your task is to narrate a story which includes all the elements depicted by the pictures. You must use all the pictures from the envelope, but you may also add extra information if you wish. You have five minutes to think before you start.

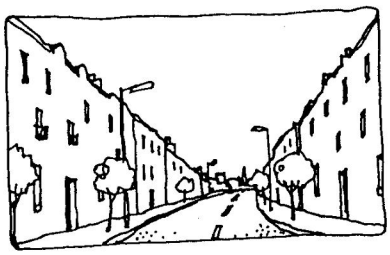
# Appendix K - New picture sequence task B











You will find six pictures in the envelope. Your task is to narrate a story which includes all the elements depicted by the pictures. You must use all the pictures from the envelope, but you may also add extra information if you wish. You have five minutes to think before you start.

# Appendix L - C-test used in the study

**Text 1**

One cool autumn evening, Bob L., a young professional, returned home from a trip to the supermarket to find his computer gone. Gone! All sorts of crazy thoughts raced through his mind: Had it been stolen? Had it been kidnapped? He searched his house for a clue until he noticed a small piece of printout paper stuck under a magnet on his refrigerator door. His heart sank as he read this simple message: CAN'T CONTINUE, FILE CLOSED, BYE.

**Text 2**

There is a third factor besides farming and herding in the spread of man-made deserts: deforestation: The progressive destruction of the Third World's stock of trees is damaging not only in dry regions: everywhere it occurs it can accelerate the decay of the soil and reduce its capacity to feed people. It can reduce rainfall and lead to drought.

**Text 3**

There are certain things which no student can do without and others may not be as necessary as you thought. It may be worth considering some small hints. You may find yourself in need of electrical appliances such as light bulbs, adapters or plugs. These can be obtained from many places. GILL is a good hardware shop and trying to find it is a challenge. It is hidden in a little alley leading off High Street called Wheatsheaf Yard.

# Appendix M - Sample tasks from HUNLAT

| Feladat elnevezése | Mérni kívánt készség/képesség | Példa (Ottó & Nikolov, 2003) |
|---|---|---|
| Rejtőző hangok | Fonetikai kódolás | [tik] hangsort halljuk; a tanulási szakasz alatt hallott hangsorok és átírásuk alapján válasszuk ki a következő öt lehetőségből a helyes átírást:<br>A. thik, B. dik, C. dhik, D. tiik, E. egyik sem. |
| Nyelvi elemzés | Szabálykivonás | adottak a következő mesterséges nyelvi és magyar nyelvi szópárok: „dant = ház”, „dantim = házban” „gup = pohár”; a négy lehetséges válasz közül azt kell kiválasztani, amelynek jelentése „pohárban”:<br>A. dantim, B. dant, C. gupim, D. gup. |
| Szavak szerepe a mondatban | Nyelvtani érzékenység | adott a következő mondat: „London Anglia fővárosa.” Válasszuk ki a következő mondat öt megjelölt szavából azt, amelyik ugyanazt a szerepet tölti be a második mondatban, mint az első mondatban a LONDON szó: „(A) Tamás (B) imádott (C) horgászni a (D) város melletti (E) kis patakban”. |
| Szótanulás | Asszociatív memória | A tanulási fázis alapján válasszuk ki a „simba” szó jelentését az öt lehetséges válaszból: A. oroszlán, B. sárkány, C. villany, D. doboz, E. egér. |