

Thinking outside the Box:

Processing Instruction and Individual Differences in Working Memory Capacity

Stephanie Andrea Peter M.A.

A Thesis submitted in partial fulfilment of the
requirements of the University of Greenwich
for the Degree of Doctor of Philosophy

May 2016

ACKNOWLEDGEMENTS

For my family.

You're my nature and my nurture.

I'll be forever grateful.

To my supervisor, Professor Alessandro Benati: Thank you for your unwavering support and enthusiasm throughout the whole project and for treating me like a colleague rather than a student – I appreciated that. Thanks go also to my second supervisor, Dr Cécile Laval.

I would like to take this opportunity to thank the academic and administrative members of the Department of Applied Linguistics as well. Thank you for your encouragement, your interest in the project, and for every informal chat in a hallway or on the stairs. Special thanks go to Dr Sarah Ann Liska for starting our PG Reading Group and to Dr Corinne Palmer-Brown for giving me purpose.

To my dear fellow PhD students: Thanks for the memories and for your friendship. You are all very special to me and it has been a privilege to share the ups and downs of this journey with you. I am particularly grateful for Benedetta Basile's help, perspective, and sense of humour, all of which helped me through the last stages of this project. Grazie di cuore, mia cara!

To my participants: This project would not have been possible without you and your commitment. Thank you for your help, time, and interest. I wish you all the best for your future endeavours!

To Mark: Thanks for keeping me sane and for reminding me of the essentials in life by somehow always asking the right questions. And for making my day, every day.

Last, but not least, I am indebted to the University of Greenwich. Without my VC Scholarship, I would not have been able to embark on this adventure.

ABSTRACT

Processing Instruction is a pedagogic intervention that manipulates the L2 input learners are exposed to in the classroom. Proponents of this intervention claim that it poses a minimal strain on learners' processing resources. While there has been extensive research on the benefits of Processing Instruction in general and the role of individual differences in particular, no conclusive evidence has been found regarding the role of individual differences in Working Memory Capacity.

To explore the question whether Processing Instruction is equally beneficial for learners at different points of the Working Memory Capacity spectrum, a case study on the effects of computer-delivered Processing Instruction has been conducted. German switch prepositions were the target feature and students' instructional gains were evaluated through sentence- and discourse-level tasks in a pre- and post-test design. Additionally, students' on-task performance was recorded during instruction. The Working Memory Capacity scores were supplemented with questionnaire data on potential mediating variables such as motivation, anxiety, personality, and aptitude. The analysis of individual learner profiles addressed yet another gap in the literature:

Robinson's (2001) work, Snow's (1989) aptitude-treatment interaction concept, and Dörnyei & Skehan's (2003) perspective on individual differences all demand a look at the bigger picture. Yet much of the Second Language Acquisition research to date has operationalised Working Memory according to Baddeley & Hitch's (1974) model, using quasi-experimental research designs – which usually fail to capture the complex and dynamic nature of Working Memory. This study addressed this gap with attention to the operationalisation of Working Memory, the analysis of task demands as well as perceived difficulty, and a focus on the interplay of several learner variables. Results seem to support the importance of Working Memory for Second Language Acquisition, at least in the short run. However, they also show a clear impact of participant-treatment interactions which might not have become evident in a group-comparison study.

CONTENTS

TABLES9
FIGURES	11
TERMS AND DEFINITIONS	13
INTRODUCTION	15
Aims of the study	16
Thesis organisation	17
1 THE IMPORTANCE OF INPUT FOR SECOND LANGUAGE ACQUISITION	21
1.1 Input Processing and Second Language Acquisition.....	21
1.2 The MOGUL framework.....	24
1.3 Input Processing according to MOGUL.....	27
1.4 Second Language Acquisition within MOGUL	29
1.5 A new definition of <i>noticing</i>	32
1.6 What gets processed and how	33
1.7 Input Processing according to VanPatten.....	37
1.8 Input Processing Principles	39
2 PREVIOUS RESEARCH ON PROCESSING INSTRUCTION	41
2.1 Processing Instruction.....	42
2.2 Critique regarding Processing Instruction	45
2.3 Extending the Processing Instruction framework.....	49
New target languages and structures.....	49
The components of Processing Instruction	51
Enhanced Processing Instruction	52
Discourse-level assessment	53
Comparison with meaningful output-based instruction.....	54
Secondary effects.....	56
Token frequency effects	57
Long-term effects.....	59
Learning rate	59
Computer-based Processing Instruction.....	61
2.4 Processing Instruction and Individual Differences	63
Age.....	65
Gender	69

First Language	71
Aptitude	72
Working Memory.....	74
2.5 Summarising the gaps: Implications for this study	78
Conclusion	83
3 WORKING MEMORY: DEFINTION AND OPERATIONALISATION	85
3.1 Working Memory.....	86
3.2 Working Memory Models.....	88
3.3 Individual Differences in Working Memory Capacity.....	93
3.4 Working Memory and Second Language Acquisition	95
3.5 Operationalisation of Working Memory and Phonological Short-Term Memory Capacity .99	
Auditory Digit Span Procedure	104
Automated Reading Span Procedure.....	105
4 LEARNER VARIABLES: SELECTION AND OPERATIONALISATION.....	108
4.1 Variables of interest.....	109
4.2 Attitude, Interest, and Motivation.....	113
4.3 Personality	115
4.4 Grammatical Sensitivity	117
4.5 Language Analytical Ability	118
5 EXPERIMENTAL TREATMENT	122
5.1 Target feature: German switch prepositions.....	122
5.2 Instructional Materials	128
Accusative and Dative Recap.....	130
Picture-matching activity.....	131
Sentence-level comprehension	132
Sentence-level listening comprehension	133
Discourse-level interpretation	134
Location versus direction.....	136
Aural sentence puzzle.....	136
5.3 Assessment materials	137
Picture description.....	137
Fill-in-the-blanks.....	138
Interpretation/ translation.....	139
Essay	141
5.4 Pilot Study data collection	141
Findings.....	144

Implications for the main study	147
5.5 Main Study data collection.....	148
Experimental Design.....	151
Timeline of the data collection.....	153
Participant sample	155
5.6 Steps undertaken to maximise validity and reliability.....	157
6 RESULTS	159
6.1 Memory Span Tasks	159
Digit Span Scores.....	160
Reading Span Scores.....	161
6.2 Test scores before, during, and after instruction.....	166
On-line performance during Structured Input Activities (session 1)	166
On-line performance during Structured Input Activities (session 2)	170
Pre- and Post-test scores.....	172
6.3 Learner Variables.....	179
Observational data	185
Potential effects of the materials	187
6.4 Putting the findings in perspective	189
Reading Span and Digit Span scores	189
Processing Instruction.....	190
Instructional gains.....	193
Learner variables	197
6.5 Recap of methodology and context	201
7 DISCUSSION.....	204
7.1 Overview of research questions and hypotheses	204
7.2 Summary of theoretical and research contexts.....	206
Working Memory Capacity and Phonological Short-Term Memory.....	208
Attitude, Interest, and Motivation.....	210
Anxiety.....	211
Extraversion.....	212
Grammatical Sensitivity and Language Analytical Ability.....	212
7.3 Discussion of the study's contribution to the existing body of research.....	214
7.4 Interpretation of results	216
MgrYyVvU.....	217
YKX5Bwj	218
74YuGGCRD.....	220

nLQ6aVus	222
8 CONCLUSION	224
8.1 Summary of the findings	224
8.2 Evaluation of significance of the findings	227
8.3 Limitations of this study and indications for future research	232
8.4 Further suggestions for teaching German switch prepositions.....	239
8.5 Practical applications.....	243
REFERENCES	245
APPENDIX A: Overview of Processing Instruction Studies.....	259
APPENDIX B: Digit Span task instructions	261
APPENDIX C: Sentences to be judged on the Reading Span task	264
APPENDIX D: Instructions for the Reading Span task.....	266
APPENDIX E: Attitude and motivation questionnaire	269
APPENDIX F: Personality questionnaire	272
APPENDIX G: Grammatical sensitivity test.....	274
APPENDIX H: Language analytical ability test	275
APPENDIX I: Input sentences and on-task performance.....	277
APPENDIX J: Pre- and post-tests.....	281
APPENDIX K: Material testing session.....	289
APPENDIX L: On-task performance during instruction (Pilot Study)	290
APPENDIX M: Exit questionnaire	294
APPENDIX N: Flyer and participant information	295
APPENDIX O: UREC application form (excerpt).....	298
APPENDIX P: Overview of the research process	299

TABLES

Table 2.1:	Participant distribution in Santamaría (2007)	76
Table 3.1:	Second Language Acquisition stages according to Dörnyei & Skehan (2003: 597)	97
Table 4.1:	Summary of variables of interest and their potential impact on Working Memory	121
Table 5.1:	Definite articles in different cases	125
Table 5.2:	Pronouns in nominative, accusative, and dative case, respectively	126
Table 5.3:	Schedule of the data collection for the pilot study	143
Table 5.4:	AMTB questionnaire results in percentages	144
Table 5.5:	Working Memory and Phonological Short-Term Memory Capacity scores	145
Table 5.6:	Grammatical sensitivity and language analytical ability scores	145
Table 5.7:	Instructional gains from pre-test to immediate post-test	146
Table 5.8:	Overview of variables and instructional gains for the pilot study	147
Table 5.9:	Research questions and operationalisation of variables for main study data collection	150
Table 5.10:	Experimental design of the main study	152
Table 5.11:	Timeline of data collection (main study)	154
Table 5.12:	Demographic data of the participant sample (main study)	156
Table 6.1:	Maximum Forward Digit Span scores and composite Forward Digit Span scores	160
Table 6.3:	Reading Span scores for all perfectly recalled letters vs all perfectly recalled sets	161
Table 6.4:	Measuring all perfectly recalled letters/sets vs accounting for accuracy of judgement	162
Table 6.5:	Storage and processing scores in combination with participants' sex	163
Table 6.6:	Pre- and post-test scores for the fill-in-the-blanks task	173
Table 6.7:	Pre- and post-test scores for the picture description task	174
Table 6.8:	Pre- and post-test scores for the translation task	175
Table 6.9:	Instances of switch prepositions in the pre- and post-test sessions	177
Table 6.10:	Overall scores for the Essay task in the pre- and post-test sessions	178
Table 6.11:	Essay overall error rates for the pre- and post-test sessions	179
Table 6.12:	Interest and attitude scores	180
Table 6.13:	Use anxiety and class anxiety scores	180
Table 6.14:	Motivational intensity and instrumental orientation scores	181

Table 6.15:	Grammatical sensitivity and language analytical ability scores and processing times	182
Table 6.16:	Extraversion scores	182
Table 6.17:	Position of non-sense word within the sentence and respective error rates	187
Table 6.18:	Position of non-sense word within the sentence and respective letter recall error rates	188
Table 6.19:	Participant rankings for test scores and overall Working Memory Capacity scores	197
Table 8.1:	Contributions of this study to the field of Processing Instruction research	231

FIGURES

Figure 1.1:	The Mogul architecture after Sharwood Smith & Truscott (2014: 17)	25
Figure 1.2:	The MOGUL architecture, by Sharwood Smith (2009)	26
Figure 1.3:	Figure 1.3: Processing Principles after VanPatten (2002: 758)	39
Figure 3.1:	Baddeley's Working Memory model after Ricker, AuBuchon & Cowan (2010: 574)	89
Figure 3.2:	The Embedded Processes Model of Working Memory (after Cowan, 1999)	91
Figure 3.3:	The shared characteristics of Working Memory Capacity measures	100
Figure 3.4:	Baddeley's Working Memory model after Ricker, AuBuchon & Cowan (2010: 574) with highlights	106
Figure 4.1:	Overview of selected learner variables in context	109
Figure 4.2:	Gaps previously identified in the literature	112
Figure 4.3:	Words in Sentences example	117
Figure 4.4:	The artificial language Pip	120
Figure 5.1:	German switch prepositions at a glance	123
Figure 5.2:	Example slide for the accusative Structured Input Activity	131
Figure 5.3:	Example slide for the picture-matching task	132
Figure 5.4:	Example slide for the sentence-level comprehension task	133
Figure 5.5:	Example slide for the sentence-level listening comprehension	134
Figure 5.6:	Example slide for the discourse-level interpretation task	135
Figure 5.7:	Example slide for the sentence-level task with stative and directive verbs	136
Figure 5.8:	Example slide for the aural sentence-level task with stative and directive verbs	137
Figure 5.9:	Picture prompt for the picture description task	138
Figure 5.10:	Example sentences from different test versions	138
Figure 5.11:	Example of Structured Input interpretation task	139
Figure 5.12:	Example sentence of main study translation task	140
Figure 6.1:	Overview of the absolute span scores by measure	163
Figure 6.2:	Relative Span Scores by Learner (in percentages)	164
Figure 6.3:	Performance on the accusative SIA	166
Figure 6.4:	Performance on the dative SIA	167
Figure 6.5:	Performance on the picture-matching SIA	167
Figure 6.6:	Performance on the sentence-level comprehension SIA	168
Figure 6.7:	Performance on the sentence-level listening comprehension SIA	169
Figure 6.8:	Performance on the discourse-level comprehension SIA	169

Figure 6.9:	Performance on the picture-matching SIA in session 2	170
Figure 6.10:	Performance on the sentence-level comprehension SIA in session 2	171
Figure 6.11:	Performance on the direction vs location SIA in session 2	171
Figure 6.12:	Performance on the aural sentence puzzle SIA in session 2	172
Figure 6.13:	Pre- and post-test scores for the fill-in-the-blanks task	173
Figure 6.14:	Pre- and post-test scores for the picture description task	174
Figure 6.15:	Pre- and post-test scores for the translation task	176
Figure 6.16:	Instances of switch prepositions in the pre- and post-test sessions	177
Figure 6.17:	Overall scores for the Essay task in the pre- and post-test sessions	178
Figure 6.18:	Essay overall error rates for the pre- and post-test sessions	179
Figure 6.19:	Overview of learner variables	183
Figure 6.20:	Overview of learner profiles using relative scores	184
Figure 6.21:	Variables across learners	185
Figure 7.1:	Example sentence in Words in Sentences section	213
Figure 7.2:	MgrYyVvU's learner profile	217
Figure 7.3:	YKX5Bwj's learner profile	218
Figure 7.4:	74YuGGCRD's learner profile	220
Figure 7.5:	nLQ6aVus's learner profile	222
Figure 8.1:	Suggested classification of locative prepositional verbs	241

TERMS AND DEFINITIONS

Digit Span	Memory Capacity span as elicited in a Digit Span task, see section 3.5
ERP	Event-Related Potential, measure of brain activity in a research paradigm using electroencephalograms
Form	Auditory or visual (i.e. textual) surface features of language
Input	Linguistic data that are available for processing, can be external or internal to the individual
IP	Input Processing, see section 1.1
Intake	Features of the input that are actually being processed by the individual and can be integrated in the developing system
L1	First Language
L2	Second Language, whether learnt or acquired in the Krashenian sense
LTM	Long-Term Memory
MOGUL	Modular On-line Growth and Use of Language framework, see section 1.2
MOI/MOBI	Meaningful Output-Based Instruction, see section 2.3
NI	Experimental group that receives no instruction (NI) on the target feature and that serves as a control group in some experiments
Noticing	Detection of a feature in the input that then becomes available for further processing, see section 1.5
Output	L2 speech, signs, or writing that learners produce
OV	Object – Verb sentence structure
OVS	Object – Verb – Subject sentence structure
PI	Processing Instruction, for a description see Chapter 2
PSTM	Phonological Short-Term Memory, see section 3.5
Reading Span	Working Memory Capacity span as elicited in a Reading Span task, see section 3.5
SLA	Second Language Acquisition: the process of acquiring a language in addition to one's first language, simultaneously or consecutively, naturally or through instruction
SIA	Structured Input Activity: manipulated L2 input designed in accordance with the Processing Instruction guidelines
SVO	Subject – Verb – Object sentence structure

Target feature	L2 feature that is the object of investigation/subject of instruction and on which the participant's proficiency is measured before and after instruction
TI	"Traditional" Instruction, defined by Processing Instruction researchers as output-oriented instruction that starts with meaningless drills and moves to communicative output practice
Treatment	Instructional intervention that participants or groups of participants receive in an experimental setting
UG	Universal Grammar: perspective on second language acquisition that presupposes innate language learning abilities that are different from the underpinnings of other cognitive processes
V2	Restriction on the placement of the finite verb, which comes in second place in German main clauses
WM	Working Memory, for a definition see Chapter 3
WMC	Working Memory Capacity as elicited by complex span tasks, see section 3.3

INTRODUCTION

The point of departure for VanPatten's (1996) Input Processing Theory is that learners of a second language are limited-capacity processors of information who have to employ certain selective processing strategies to cope with the L2 input they are exposed to (VanPatten, 1990). While some of these strategies might be adequate, inappropriate strategies can lead to incorrect form-meaning connections and therefore hinder successful language learning. Processing Instruction is a pedagogic intervention predicated on the Input Processing Theory as it seeks to optimise learners' input processing strategies by systematically manipulating the input that learners receive in the classroom.

To this end, the learners are confronted with aural and textual input that has been reduced in complexity, that is unambiguous, and that requires them to process grammatical forms in order to derive meaning. Inappropriate processing strategies will consequently lead to incorrect interpretations and learners will be forced to adapt more appropriate strategies to avoid continued frustration. Proponents of the Input Processing Theory (e.g., VanPatten, 2007, 2004, 1996; Benati, 2013; Lee & Benati, 2007) have claimed that Processing Instruction poses only a minimal strain on the individual's processing resources. If this is correct, then learners of different Working Memory Capacity spans should be performing equally well during Processing Instruction.

While there has been extensive research on the benefits of Processing Instruction in general (e.g. VanPatten & Cadierno, 1993; Cadierno, 1995; Benati, 2001, 2005; Benati & Lee, 2008; Uludag & VanPatten, 2012) and on the role of individual differences such as age (Benati, 2013; Laval, 2013; Mavrantoni & Benati, 2013), gender (Agiasophiti, 2013), and linguistic background (Lee & McNulty, 2013), only one study (Santamaría, 2007) has considered the role of individual differences in Working Memory Capacity. The results of this study have been inconclusive. Yet, Working Memory is a construct which has been hypothesised to play a role in Processing Instruction under particular circumstances:

Perhaps there is a relationship between grammatical sensitivity and the ability to use EI [explicit information; S.P] during a processing task, a relationship that may involve a third variable such as working memory. Only future research that uses a multitude of measures of individual differences could address this issue.

VanPatten & Borst (2012: 104)

This is exactly what the current study attempts to do.

Working Memory has been investigated within several disciplines interested in the exploration of higher-level human cognition, including psychology and Second Language Acquisition research, but the respective operationalisations of the construct have been rather different. While the less modular Working Memory models (e.g. Cowan, 1999) have recently enjoyed a lot of attention in psychometric experiments and have received support from Event-Related Potential research paradigms, much of the Second Language Acquisition research to date has operationalised Working Memory according to Baddeley & Hitch's (1974) seminal model, using quasi-experimental group comparisons – which usually fail to capture the complex and dynamic nature of Working Memory. Robinson's (2001) understanding of aptitude complexes, based on Snow's (1989) aptitude-treatment interaction concept, and Dörnyei & Skehan's (2003) perspective on individual differences all demand a look at the bigger picture, however.

Aims of the study

To explore whether Processing Instruction is equally beneficial for learners at different points of the Working Memory Capacity spectrum, a case study on the effects of computer-delivered Processing Instruction has been conducted. German switch prepositions were the target feature and the participants' performance before, during, and after two 50-minute sessions of Processing Instruction was evaluated through sentence- and discourse-level tasks in a pre- and post-test design. It was hypothesised that the participants would perform equally well on the Processing Instruction tasks, but display performance differences on the more challenging assessment tasks.

The participants' Working Memory and Phonological Short-Term Memory Capacity scores were supplemented with questionnaire data on potential mediating variables such as learner background, motivation, L2 anxiety, extraversion, grammatical sensitivity, and language analytical ability. It was hypothesised that these variables would have an effect on the participants' processing resources because they had the potential to affect the way the participants would interact with the materials. The analysis of individual learner profiles addressed the aforementioned gaps in the literature with attention to the operationalisation of Working Memory Capacity, the analysis of task demands as well as perceived difficulty, and a focus on the interplay of several selected learner variables.

The results seem to support the importance of Working Memory Capacity in second language learning, at least in the short run. While the participants showed similar

performance levels during the Processing Instruction treatment, they displayed performance differences on the immediate and the delayed post-tests. However, the results also show a clear impact of participant-treatment interactions, which might not have become evident in a group-comparison study. While Working Memory Capacity has traditionally been operationalised as a relatively stable construct, these findings point to a more dynamic processing resource pool that emerges from and depends on the interactions between learner, input, and context.

This dissertation is called Thinking outside the Box because it favours a process-oriented perspective on Second Language Acquisition, meaning that it is interested in what might be going on in the learner's proverbial "black box" of cognition. It also subscribes to a notion of Working Memory that is fluid as opposed to a strictly modular multi-component model that seems somewhat static (and whose illustrations are often box-like in appearance). While the concept of modularity, especially in connection with linguistic processing, is not at all abandoned, its working definition differs from traditional understandings of modularity: "When a neuroscientist uses the word "module", s/he is usually referring to the fact that brains are structured, with cells, columns, layers, and regions which divide up the labor of information processing in various ways" (Elman, 1997: 36). This could be due to synchronous or diachronous development and some, more specialised, modules might display more plasticity than other, less specialised, modules. Future cross-disciplinary research will have to explore to which extent this modularity is responsive to the exposure to the actual environment (nurture) or fixed in its genetic predisposition (nature).

Thesis organisation

In a field that is not defined by unanimity, one statement seems to find general support in Applied Linguistics: Input is necessary for Second Language Acquisition. Based on the importance of this construct, Chapter 1 will review the role of input and input processing for Second Language Acquisition (1.1) and discuss the Modular On-line Growth and Use of Language (MOGUL) framework (1.2 – 1.6) as a theoretical underpinning for VanPatten's (1996) Input Processing Theory (1.7 and 1.8) and the resulting pedagogical intervention termed Processing Instruction. It will argue that MOGUL is a much-needed complement for Processing Instruction research that deals with individual differences because these stem from the processes inside the learner's mind as well as from the features of the input.

Processing Instruction as a form of explicit grammar instruction will then be described in more detail in Chapter 2, especially in terms of the research foci that have generated questions and issues that are of relevance for this study. Previous research on individual differences in Processing Instruction research will be presented in Chapter 2.4. It will be argued that Working Memory is a particularly relevant variable to measure in this context because Processing Instruction is based on the assumption that second language learners are limited-capacity processors of information. The language learning tasks used in Processing Instruction are consequently designed to put as little strain as possible on the processing resources of the learners and should therefore be beneficial to the whole range of possible Working Memory Capacity spans. While one previous study addresses Working Memory limitations as an individual differences variable, it does so with regards to the assessment rather than the instruction. The present study addresses this issue.

Chapter 3 presents a definition of Working Memory and introduces two of the most seminal Working Memory models (section 3.2). While it would go beyond the scope of this dissertation to examine the whole spectrum of Working Memory models and conceptualisations (from such diverse disciplines as philosophy, cognitive linguistics, cognitive physiology, psychology, medicine, neuro-sciences, etc.), it seems necessary to discuss at least the very basic cognitive foundations which underlie the notion of Working Memory before attempting to make claims about the role of Working Memory in a particular instructional treatment. The remainder of the chapter then explains the operationalisation of the Working Memory and Phonological Short-Term Memory constructs for this study as well as the instruments and procedures that have been chosen to measure them.

Chapter 4 introduces the learner variables that have been selected for exploration as potential mediating variables of Working Memory Capacity limitations: attitude towards the second language and its speakers, interest in learning the language, use anxiety inside and outside of the classroom, extraversion, grammatical sensitivity, and language analytical ability. Once the selection of these variables is justified, the individual constructs will be presented and their respective operationalisation will be explained together with the questionnaires and tests that have been adapted for data collection in this study. It is the researcher's belief that theory and application cannot be separated as they inform each other, which is why Chapters 3 and 4 adapt a process-oriented perspective on this research project and present the theoretical constructs as well as their operationalisation for this study and the concrete instruments of measurement that follow from these deliberations.

Chapter 5 then focuses on the experimental treatment. Section 5.1 introduces the target feature, German switch prepositions, and explains how the Processing Instruction approach differs from conventional ways of teaching this feature of German grammar. Section 5.2 presents the instructional materials which have been designed following Farley's (2005) guidelines for Structured Input Activities. The aim of Processing Instruction is to force the participants to process case inflections rather than semantic information in order to derive meaning from the input. This is reflected in the design of the instructional materials. The next section of the chapter then describes the assessment materials used in the pre- and post-tests. In section 5.4, the pilot study and its findings will be presented, the implications for the main study will be briefly discussed, and changes in the materials from pilot study to main study will be highlighted again. Section 5.5 summarises the experimental design as well as the circumstances and the timeline of the data collection. It also introduces the participant sample. The chapter ends with a summary of the steps that have been undertaken to maximise the validity and reliability of the data.

Chapter 6 presents the results of the main study. First, different scoring techniques for the Working Memory Capacity measures will be compared in order to find the one that is most sensitive to individual differences in Working Memory Capacity (6.1). Then participants' performance scores before, during, and after instruction will be presented (6.2) and the questionnaire data on potentially mediating variables will be described and supplemented with observational data (6.3). Section 6.4 will then attempt to put the results for the individual variables into perspective before section 6.5 provides a recap of the study's methodology and context.

Chapter 7 begins with a restatement of the research questions and hypotheses of this dissertation. The theoretical and practical context of the study will then be summarised (7.2) before the contributions of its results to the field of Processing Instruction research (7.3) will be discussed. The chapter ends with an interpretation of the results that is based on the MOGUL framework and that focuses on individual learner profiles, stressing the importance of the individual in individual differences research.

Chapter 8 summarises the findings of this study and evaluates their significance in terms of their theoretical and methodological contributions to the existing body of research (8.2). Following that, the limitations of this small-scale case study on Processing Instruction targeting German switch prepositions will be acknowledged and future research opportunities for addressing these gaps will be pointed out (8.3). Section 8.4 makes

suggestions for a more principled approach to teaching German switch prepositions. The dissertation ends with recommendations for practical applications of Processing Instruction in computer-based learning contexts in general and Blended Learning in particular (8.5).

1 THE IMPORTANCE OF INPUT FOR SECOND LANGUAGE ACQUISITION

This introductory chapter will stress the crucial role of input for second language acquisition. Following that, it will introduce a theoretical framework of input processing, namely the Modular On-line Growth and Use of Language (MOGUL) model devised by Sharwood Smith and Truscott (see Sharwood Smith & Truscott 2014 for the latest version). The MOGUL framework will be described in some detail as it provides the theoretical underpinning for the researcher's understanding of input processing and Second Language Acquisition. There is a clear link to the Input Processing Principles posited by VanPatten (e.g. 1996, 2002), which have been influenced by Sharwood Smith's earlier work. MOGUL, however, deals with the processing as well as the acquisition of first and second languages and addresses many issues and questions which have not been considered in VanPatten's Input Processing Principles, let alone in the area of Processing Instruction research. Throughout this dissertation, the MOGUL framework will serve as a theoretical foundation for the consideration of findings from previous and current research. It is therefore warranted to take a closer look at the architecture of the language system proposed by Sharwood Smith & Truscott (2011, 2014) and to describe how language processing leads to language growth. The re-conceptualisation of constructs such as noticing and the definition of consciousness within the MOGUL framework will offer a more concise explanation of the processes that are potentially being executed in the individual's brain when attending to linguistic input – processes which are traditionally of no interest to VanPatten and colleagues as they go beyond the tangible features of the input that are accessible to the experimenter's manipulation.

1.1 Input Processing and Second Language Acquisition

Input is the sine qua non of language learning. There is no theory in First or Second Language Acquisition research that does not refer to input in some way, despite the fact that different theories position themselves differently with respect to the role of input: In behaviourism, the repeated exposure to input is believed to lead to the development of conditioned stimulus-response patterns that are entirely dependent on the input that the individual receives. A similarly powerful role is ascribed to input in the context of connectionist theories, where all information is latent in the environment, meaning that

the input alone is responsible for changes in the learner's mind. Even in generative theories, which presuppose rich innate foundations for language-related abilities, the input interacts with the Universal Grammar and leads to parameter setting. It should be noted though that only primary linguistic data count as input for Universal Grammar accounts, feedback or explicit metalinguistic knowledge does not. Sanz (2005: 7) hence concludes that "generativism offers a widely accepted theory of language that addresses the *what* of language acquisition, but when it comes to the *how*, most SLA scholars turn to cognitive, information-processing models."

Other models of Second Language Acquisition have incorporated the role of input in terms of its social (e.g. interactional input, Gass, 1997) or informational (e.g. in the Competition Model, Bates & McWhinney, 1981) value. The importance of input has also been acknowledged for language teaching and led to empirical research on input manipulations and/or enhancement that strived to focus the learner's attention to form. Input processing, i.e. "the interaction between the learner's mind and instances of the language to which the learner is exposed" (Truscott & Sharwood Smith, 2011: 498), plays a key role in Second Language Acquisition because it provides the linguistic information that is (at least partly) converted to intake, which in turn can be integrated into the developing system. The importance of input for language acquisition (Gass 1988, 1997) is thus generally accepted. The question is rather how the conversion from input to intake is to be conceptualised.

The differentiation between input and intake dates back to Corder (1967) and is essential: Not all of the input provided actually interacts with the learner's mind. Even the information that the learner consciously attends to is not necessarily processed on a deeper level and can remain un-analysed, with no impact on processing or development (Truscott & Sharwood Smith, 2011: 499). Whether different features of language are processed in different ways and/ or stored in different places is not considered in this concept of input conversion, however. Closely linked to the conversion from input to intake is the concept of *noticing* as proposed by Schmidt (1995, and elsewhere). He argues in his Noticing Hypothesis that learners must be aware of a formal feature in the input in order to acquire it. However, they do not need to understand the form in order to notice it: Noticing refers to the "conscious registration of the occurrence of some event," not the "recognition of a general principle, rule or pattern" (Schmidt, 1995: 29).

While it is hard to determine what goes unnoticed, the distinction between noticing and understanding is even harder to operationalise as there is only a thin line between noticing

the sensory event of a surface feature being present in the input and classifying this feature as a noun, for example. If noticing is nothing more than being globally aware of input, without any further interpretation of the surface feature by the learner's processing resources, then this hypothesis is not applicable to a pedagogic framework that focuses on form and aims to make learners notice features of the input. The concepts of consciousness and noticing therefore need a theoretical foundation if they are to be integrated into a cognitive approach to input processing: "The central requirement for such an account is that it should say how consciousness fits into the cognitive system, including what can become conscious and under what conditions" (Truscott & Sharwood Smith, 2011: 507). The Modular On-line Growth and Use of Language (MOGUL) framework proposed by Sharwood Smith and Truscott (most extensively discussed in their 2014 book) offers just such an account.

It has been chosen as a theoretical underpinning for this study because its explanatory scope allows a fruitful integration of findings from Working Memory studies, individual differences research, and processing-oriented accounts of Second Language Acquisition. As a framework, MOGUL is general and abstract enough to be consistent with accepted theories and models in other fields, for example Cowan's (1999) Embedded-Processes Model of Working Memory, Eysenck's (1990) Arousal Theory of personality, and Dewaele's (2013) research on foreign language anxiety, to name just the most relevant for this project. In addition to that, it incorporates the most promising concepts from very different schools of thought within the field of Second Language Acquisition, as Sharwood Smith and Truscott (2014, Chapter 4.2) repeatedly emphasise. The ultimate goal of these authors is to provide a theoretical framework which can lead to a research agenda of principled, empirical testing of falsifiable hypotheses. The foundation for this has been laid with the incorporation of different types and sources of data, from behavioural observations to neuro-imaging findings, into one parsimonious account of how language is processed in the individual's mind and how language growth is a consequence of this processing. Once the architecture of this framework has been described, some reinterpretations of crucial concepts will be presented, such as noticing and consciousness, paving the way for a better explanation of how Processing Instruction might increase the chances of stable form-meaning connections being made. While VanPatten's Input Processing Principles describe the characteristics of the input, MOGUL offers a description of what happens in the individual's mind during input processing. Considering that individual differences affect the individual's processing rather than the external input, MOGUL is an important complement of the Input Processing Principles.

1.2 The MOGUL framework

While a communicative orientation towards language teaching seems to be the *modus operandi* of choice in most classrooms these days – at least on paper – some basic notions of generative linguistics, such as the existence of developmental stages in language acquisition and the relative systematicity of the developing system, have reached a wide acceptance in recent years (Sanz, 2005: 7). “Moreover, this branch of linguistics can be seen as coming closer to more cognitive understandings of language and language development through recent work that draws on developments in psycholinguistics” (Whong, 2007: 143). One example of this fusion of generative and cognitive approaches to language is the MOGUL framework proposed by Truscott & Sharwood Smith (2004):

For example, although aspects of MOGUL bear a resemblance to the Competition Model (MacWhinney 1987), the foundations of MOGUL are quite different in that any parsing activity, including competition, is tightly constrained by universal linguistic principles of a type not recognized in the Competition Model but nevertheless ones which have guided the largest existing body of experimental studies in L2 grammatical development.

Sharwood Smith & Truscott (2005: 238)

This means that the MOGUL framework presupposes a modularity of the mind in which core linguistic knowledge has a different status from other types of language-related knowledge. But because the authors attempt to give an account of both the content of linguistic knowledge and the on-line processing of this content, insights from information processing theories closer to connectionist and emergentist theories of language acquisition are incorporated (Sharwood Smith & Truscott, 2014: 99–102).

At the core of the MOGUL architecture is a language module that houses those systems that are specific to human language, i.e. the phonological and the syntactic modules. These two sub-systems are linked with each other via an interface system that also links them to adjacent systems outside the core language module, such as the auditory system, which feeds into the phonological module, but also deals with all other auditory input that is potentially meaningful, e.g. the creaking of a door or the noise generated by a vibrating mobile phone. The conceptualisation of the phonological loop in Baddeley’s (1986, 2000) Working Memory model has considerable overlap with this phonological module, which is not surprising as both constructs are based on behavioural evidence that leads researchers to make assumptions regarding the nature of phonological

processing. The phonological module is also linked via interfaces to the visual system (which is a connection that is relevant for reading texts) and the articulatory system, which is supposed to deal with both speech production and sign language (Truscott & Sharwood Smith, 2011: 508). Via another interface, the core language module is linked to the conceptual system, which is responsible for the encoding of meaning, corresponding with the traditional notions of semantics and pragmatics. It is important to note that the conceptual system is not part of the core language module, which makes it accessible for conscious thought. Related to these systems is the system of affective structures, containing affective representations connected with perceptual and conceptual items, but also general states of the individual, such as happiness or anxiety. Figure 1.1 below represents an overview of the different modules with a high degree of abstraction:

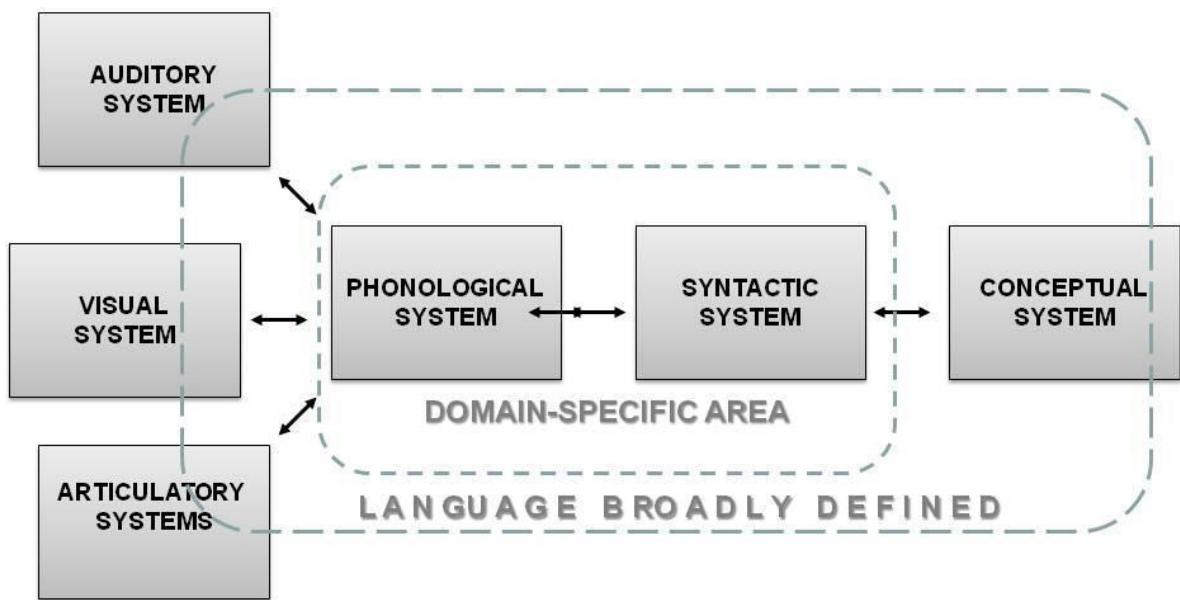
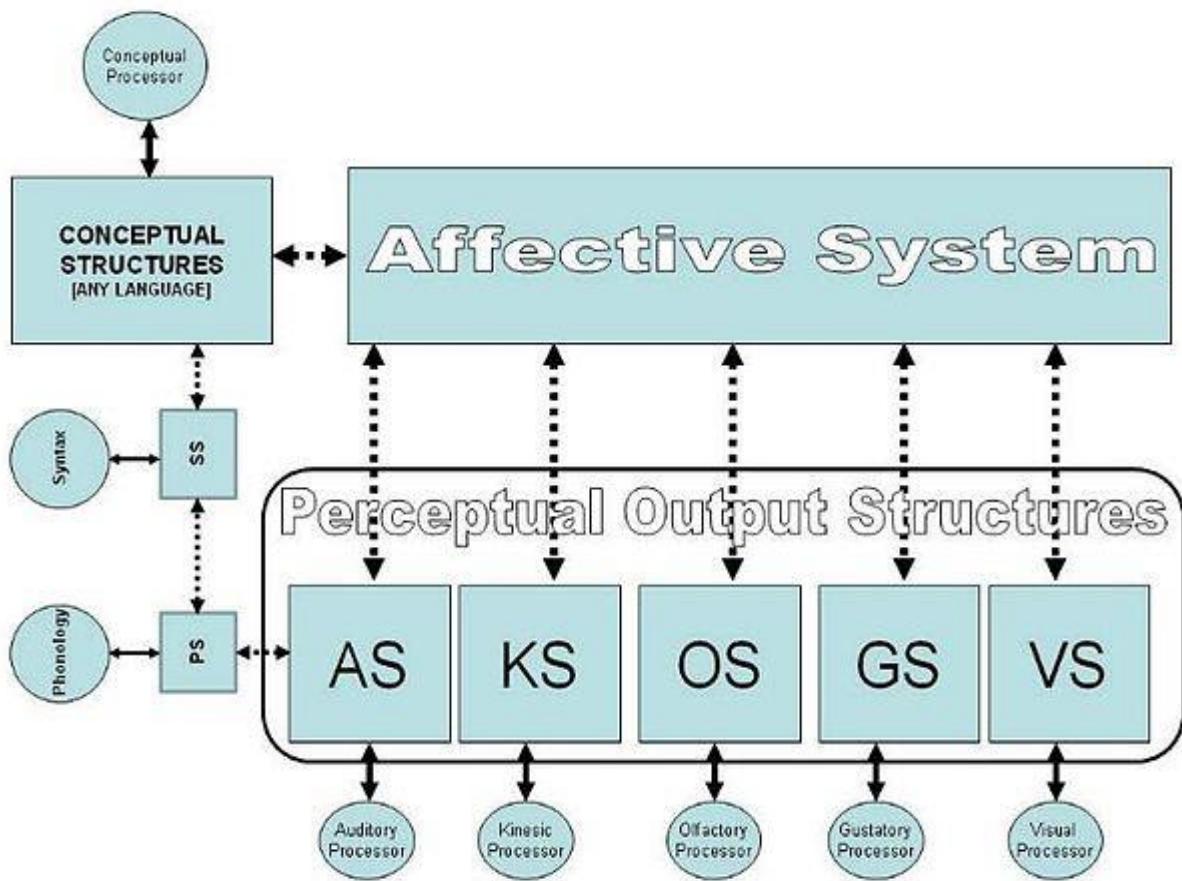


Figure 1.1: The Mogul architecture after Sharwood Smith & Truscott (2014: 17)

Another important module is the perceptual output structure (POpS), as shown in Figure 1.2 below, that “represents the ultimate output of the modality-specific processing systems, each dealing with input from one of the senses” (Truscott & Sharwood Smith, 2011: 509). The modules are connected with each other via interfaces, thus synchronising the response to sensory events. The activation levels of the affective structure and the perceptual output structure are exceptionally high, which is evidence of the importance attributed to sensory events and the influence of affect on behaviour – a remnant of their crucial role in human evolution (Truscott & Sharwood Smith, 2011: 511–2). It should be noted that this conceptualisation assumes independent stores for each of the five, and

potentially more, senses – a concept that has initially been criticised by some authors (e.g. Cowan, 1999), but is not incompatible with their empirical findings (see Cowan, 2015).



https://en.wikipedia.org/wiki/MOGUL_framework#/media/File:MOGULarchitecture.jpg

Figure 1.2: The MOGUL architecture, by Sharwood Smith (2009)

Each of the modules presented above consists of an information store (lexicon) that contains representations of module-specific structures and a processor that contains the module's rule system. The phonological module for example can only form and process phonological structures and it is the only module capable of this. The processors all work independently in parallel with those features of the input that are relevant to them. When exposed to linguistic input during conversation, for example, the auditory system will deal with the auditory aspects of that input and produce an auditory representation, while the visual system will process the visual input and produce a visual representation, and so on. Interfaces match items in the various modules that have high activation levels and are therefore available for processing, creating a consistent overall representation or message. "The internal interfaces between phonology and syntax display high

connectivity: some elements are in close correspondence (a phonological ‘word’ is very similar although not identical to an equivalent category in syntax, for example), whereas areas linked up by what are here called external interfaces are not so richly interconnected” (Sharwood Smith, 2013: 33).

1.3 Input Processing according to MOGUL

What happens during real-time input processing then? The processing of the item ‘green’, for example, begins when the processor of the auditory module constructs a representation for the utterance [‘gri:n]. This leads to the activation of the corresponding phonological structure /green/ in the phonological store and the co-indexed information (adjective) is activated in the syntactic store. Another interface serves to match the activated representation of ‘green’ in the conceptual store and quite probably also representations stored in the affective structures. With the co-activation of conceptual structures, meaning is activated – and as the conceptual store is not part of the core language system ruled by UG, conceptual structures are potentially accessible to conscious reflection and, more generally, to thought. It should be noted though that the output of the auditory system, /‘gri:n/, would have activated representations of phonetically similar representations in the phonological store, such as /‘grɪn/ or /‘bi:n/, as well as a multitude of connected representations in the conceptual store (e.g. COLOUR, CONCERNED WITH PROTECTING THE ENVIRONMENT, NOT YET RIPE) that are associated with green. Visual representations in the visual store (e.g. remembered or imagined pictures of trees, meadows, beans, etc.) would also have received activation. These co-activated representations can be rather complex: /‘gri:n/ could have activated composite representations, such as ‘green with envy’ or ‘the grass is always greener on the other side’, if the processing system has processed these collocations numerous times before. Additional characteristics of the input can also activate representations in the conceptual store, such as a specific register, a dialect, or the mood of the speaker (Sharwood Smith & Truscott, 2014: 188).

Several representations are active in each store and across modules at the same time, each representation competing with some and reinforcing others. “Coordinated activity over the entire system thus occurs as an emergent property, rather than through the efforts of a central agency” (Elman, 1997: 85). If the utterance has to do with green beans, for example, the convergent active representations in the various modules will reinforce

each other and achieve high activation levels which can be matched by the interfaces. Representations which receive less reinforcement, such as ‘green with envy’ lose the competition in this case. The selection of the most fitting representations can be made based on processing within the core language module, based on processing that happens outside of it, or based on a combination of both: “No utterance consists entirely of primitives, so in this sense every instance of processing involves the use of stored chunks. Considerable variation occurs, however, in the amount of on-line construction that can or must be done” (Sharwood Smith & Truscott, 2014: 78).

The more often the processing system has encountered a chunked representation, e.g. ‘green with envy’, the more likely is this chunk – that is stored outside the core language module – to win against output from inside the UG-driven modules. Retrieval of chunked representations is faster for frequently co-occurring items than is the on-line analysis of the input within the syntactical module. Yet, both kinds of processing routinely happen in parallel and the fastest processing route produces the winner of the competition (Sharwood Smith & Truscott, 2014: 78–82). This can have negative outcomes, especially when the more general learning strategies that shape processing outside the core language module are at odds with the UG-driven processing within it or when the appropriate processing for output depends on pragmatic knowledge rather than UG rules. Learners of a pro-drop language can have acquired the pro-drop feature in their core language module, but fail to drop the pronoun when pragmatics would require them to do so (Sharwood Smith & Truscott, 2014: 221–2). Evidence in support of this parallel processing has recently come from studies employing the Event-Related Potentials paradigm (Tanner & Van Hell, 2014; Tanner, et al., 2014). These studies have analysed individual participants’ waveforms in addition to the usual grand mean averages and have found that both L1 and L2 speakers of a language are distributed across a continuum ranging from memory-dominant to analysis-dominant processing, with both processing paths leading to high levels of proficiency.

Contrary to traditional models of Working Memory, there is no structural difference between Working Memory and Long-Term Memory in this framework. Working Memory simply consists of the activated representations in a store and is therefore distributed across the network of modules. Resting activation levels of a structure depend on how well it is integrated in Long-Term Memory, i.e. how frequently it is processed, and impact on how easily the structure can be activated and thus made available for further processing. As MOGUL makes no distinction between processing for input and processing for output, this suggests that frequent exposure to a structure makes it more

probable to be produced, but this ultimately depends on resting activation levels and competition, as we will see later. It is possible though that activation patterns emerge from frequent processing, linking particular structures or items with each other and thus decreasing competition between these items and other associated (i.e. co-indexed) items. Extended to sentence comprehension, it becomes clear that this is not a linear process. “Several candidate chains are temporarily maintained in working memory so a best-fit can be achieved. This means there are some repeated to-and-fro passes through the separate modules before a chain of representations are selected.” (Sharwood Smith, 2013: 34).

This new activation pattern, as well as its individual representations, will receive a slight increase in resting activation levels, making it more likely to be processed, as evidenced in semantic and syntactic priming studies. Subsequent and frequent use of the same representations will lead to a further increase in resting activation levels, whereas the resting activation level of rarely processed representations fades over time. This makes them less readily available for processing in both interpretation and production. It should be noted, however, that this decrease in activation levels can be very gradual and that the successful processing also depends on the activation levels of the competing representations. The MOGUL framework is linked with Truscott & Sharwood Smith’s (2004) Acquisition by Processing Theory (APT), thus based on the notion that language growth occurs simply through the on-line use of language. This notion rejects any kind of separate learning mechanism, but presupposes a language-specific mental system. The question as to whether the growth of this system is based on changes in the connections between neurons or changes in the neural substrates themselves, is not considered by the authors as the framework operates at a more abstract level (Sharwood Smith & Truscott, 2014: 342–4). MOGUL is, however, firmly grounded in and informed by linguistic theory and has the potential to inspire future cross-disciplinary research that addresses these questions.

1.4 Second Language Acquisition within MOGUL

What does this mean for the acquisition of an additional language? Crucially, the difference between First and Second Language Acquisition lies in the presence of an already established language in the mind of the individual by the time the L2 is being acquired. “The interesting thing about this phenomenon, from a developmental viewpoint,

is that it suggests that ***the ability to learn may change over time – not as a function of any explicit change in the mechanism, but rather as an intrinsic consequence of learning itself*** (Elman, 1997: 70; emphasis is not mine). This leads to the situation that the newcomer representations of the L2 start at a disadvantage and consequently need a lot of processing to build up their activation levels in order to be able to win the competition against L1 representations. “Initially, L2 input will be mapped on L1-based representations wherever possible. New representations will compete with the old ones until their strength is established via further exposure to the language” (Sharwood Smith, 2013: 36).

While the resting activation levels of linguistic items depend on exposure to input, there is also the understanding that certain activation patterns develop over time, making some associations more probable than others and involving meta- or extra-linguistic knowledge as well. “A matching parse receives reinforcement within the domains of both grammar and cognition” (Whong, 2007: 148). In terms of instructed (adult) Second Language Acquisition, it is possible that some linguistic representations are stored in systems outside the core language module only. This also touches the question as to “whether each and every one of these sub-areas [syntax, phonology, semantics and pragmatics; S.P.] can be acquired in the same way, i.e. whether the same learning mechanisms are involved” (Sharwood Smith, 2013: 28). For an understanding of language as something that is learnt in fundamentally different ways than other forms of knowledge, which is the notion that underlies Processing Instruction, this question is rather important.

MOGUL links language development to language processing, using concepts from psycholinguistics and information processing research. Learning, understood as shifts from L1-like grammar representations to increasingly L2-like grammar representations, can therefore be explained in terms of competition between representations. The resting activation levels of representations in the phonological, syntactic, and conceptual stores increase with the processing frequency of the respective item(s). This internal frequency does not necessarily equal the objective frequency of the item in the input that the individual is exposed to though: If the auditory processor does not provide auditory representations for the other modules’ input (e.g. because there is interfering noise that wins the competition), then the item will not be processed in the language systems and will have no impact on their development. The current activation level is the resting activation level plus additional activation by the module’s processor. As various modules are linked to each other via interfaces, “[i]ncreased activation of an item follows automatically from increased activation of the item it is associated with” (Sharwood Smith

& Truscott, 2005: 234). If the phonological processor activates /'gri:n/, the co-indexed representations in the syntactic module (*adjective*) and the conceptual system (the colour green) will also be activated. In addition to this co-activation, which especially applies to the contents of the core language module, spreading activation can add slight increases in activation levels. The activation of the conceptual representation GREEN, for example, will lead to spreading activation of the whole category COLOUR.

If growth is linked to activation levels, then it is gradual as activation can take on many levels. On the other hand, there will be noticeable shifts between two grammatical systems once certain activation thresholds are crossed. These are relative to the competition and not fixed though. Competition takes place because several alternative elements, belonging to both L1 and L2(s), are activated at once. In terms of language production, this can be witnessed in the form of slips of the tongue. In these cases, phonologically or semantically similar items with equally high activation levels compete with each other and the less appropriate representation wins the competition. It is consequently uttered because the production takes place before the overall message is parsed, making inhibition of the non-adequate alternative impossible. For language acquisition, competition means that items which win on a regular basis will be further strengthened as their resting activation levels will increase over time. After a (potentially very long) period during which several competing items have similar activation levels, producing optionality for the learner, one item will eventually win out and this will lead to the attrition of the rivalling item. Attrition can lead to the ‘forgetting’ of the L1 if the individual has been immersed in the L2 environment for a long time (Isurin, 2013). It does not mean that the less active representation vanishes, however. Under certain conditions, it could even occasionally win the competition again. In circumstances of regular L2 processing though, a fundamental change in the system can be noticed once the threshold is crossed and the L2 item will dominate the competition in the future. This kind of learning curve supposedly applies to all modules, fitting with the more general assumption from Working Memory research that “all types of features share a common set of abstract dynamic properties, such as the neural mechanism of activation, the persistence of activation, and the neural mechanism of interference caused by similar features” (Cowan, 1995: 37).

1.5 A new definition of *noticing*

From a MOGUL perspective, subliminal perception is possible. It occurs when the articulatory representation of an input string is built, but does not reach the activation threshold required for entering consciousness, letting information in the utterance go unnoticed. Once the input reaches the activation threshold, the individual becomes consciously aware of the overall articulatory representation of the utterance, but no portion of it is singled out as particularly interesting. This rather general awareness turns into noticing when a “follow-up POpS representation, consisting of one portion of the original representation, is constructed as the result of processing that treats it as an instance of a particular form, and it reaches an activation level sufficient for awareness” (Truscott & Sharwood Smith, 2011: 520). If this form-function connection is made, the individual can become aware of it. Conscious awareness would indicate high activation levels, which would make enduring changes in the information stores more probable – which is what traditionally is understood as learning. This does not mean, however, that an appropriate meaning is attached to the form detected in the input, or that the representation leads to conscious understanding beyond the noticing of this particular instance of form (e.g. in terms of the individual being able to formulate a rule). This would require additional representations in the conceptual store to reach an activation level that is high enough to make them available to conscious cognitive processes, thus creating explicit linguistic knowledge. The existence of the latter, on the other hand, does not lead to a growth of the language module.

For a re-interpretation of the Noticing Hypothesis within the MOGUL framework, we can summarise that noticing is not about the awareness of additional representations, but about recognising new information in the input as an instance of a particular form or representations already established. Noticing the third person singular –s in English present tense, for example, then involves an auditory representation that activates a phonological representation that, in turn, is co-indexed with the information ‘+present, +singular, 3rd person’ in the syntactic store. No noticing will occur if the sound does not register, thus not producing any output of the auditory module, or if no clear connection to the syntactical information has been established previously. Once a concept has a conscious expression though, either in spoken or written form, it is accessible to conscious processes, even if the concept itself never reaches awareness. For Second Language Acquisition, this means that input for acquisition can be defined as an either articulatory or visual perceptual representation of spoken or written language. The input

must contain specific linguistic information in order to qualify as potential intake and the module in question must be able to abstract this linguistic information from the input, requiring the processor to be able to deal with this kind of information and requiring the module's store to be ready for it. The extraction of information depends on the impact of the representation on the perceptual output structure: For the extraction of new information, the processor needs time and relatively high activation levels to deal with the information. If the information recurs frequently and the extraction increases in automaticity, activation thresholds can be a lot lower.

1.6 What gets processed and how

The role of consciousness has been a hot topic in Second Language Acquisition research and VanPatten has expressed his views on the matter with regards to Processing Instruction thus:

Note that the learner may or may not be aware of the product; it is sufficient for the learner to be aware of her attention. Since product (knowledge) and the process (of being aware) are separable within the construction of attention, conscious attention does not automatically suggest explicit or declarative knowledge. In short, one can attend and yet not be aware of the product or not be able to fully articulate it. This is worthy of emphasis since so often in SLA research explicit knowledge (a product) is equated with consciousness."

VanPatten (1994: 33)

MOGUL offers an explanation for the hypothetical relationships between consciousness, attention, and the growth of language as well as of metalinguistic knowledge. As we have seen, the sub-modules each process different aspects of linguistic input and are connected via interfaces. One of those interfaces links language processing to the individual's real word knowledge in the conceptual store. While modular linguistic knowledge in the core language module cannot be accessed consciously, representations outside it, i.e. metalinguistic knowledge and semantic representations, can be accessed consciously (Whong, 2007: 148). Thus a dual representation of knowledge is created. The individual can be conscious of representations associated with the input, but not of the input or the abstract concept of language itself:

"[T]he module is true to the traditions of learnability theory and reflects the standard principle that the domain-specific language module (or its equivalent) is inaccessible from outside. We have no direct knowledge or awareness of the workings of syntax and phonology and can only understand its mechanisms from outside, indirectly, via reflection, that is, 'metacognitively'."

Sharwood Smith (2004: 264)

But under which circumstances do representations become conscious? Within the MOGUL framework, the answer is: "A representation becomes conscious if and only if its activation level crosses a threshold" (Truscott & Sharwood Smith, 2011: 514). This threshold level is only reached in modules such as the perceptual output structure and the affective structure, modules which are hypothesised to have extremely high resting activation levels thanks to their important role in human evolution. What individuals are aware of then is the auditory or visual representation, i.e. the perceivable expression of language: sound, signs, or written form. Other representations in the conceptual or syntactic stores will never reach activation levels high enough for them to cross the consciousness threshold.

This implicates that very little linguistic knowledge can ever become conscious to the individual's mind and that learners can never become aware of grammar in the sense that they can never become aware of e.g. the syntactic representations that are at the basis of their language processing. Knowledge like this is built upon processing experiences and modular in nature (Sharwood Smith, 2004: 269), i.e. informationally encapsulated. Nevertheless, individuals can become indirectly aware of extra-modular linguistic knowledge under certain circumstances, i.e. when these representations are expressed in perceptual form (Truscott & Sharwood Smith, 2011: 514). Learners can also develop metalinguistic knowledge that is stored outside the core language system, namely in the conceptual module. This means that metalinguistic knowledge is in principle accessible to consciousness; it does not mean though that the individual is at all times aware of the processes and contents within the conceptual module. While unfamiliar problems at first require attention, they can become automated in time, decreasing the need for attention considerably. The same is supposed to be the case for the development of metalinguistic knowledge. But this system on its own does not lead to language acquisition: "For a metagrammar to operate and indeed for the rawest, simplest kind of metalinguistic awareness (without any knowledge of formal grammar) there still needs to be a link up of sorts with the core language system described above" (Sharwood Smith, 2004: 270).

As mentioned above, perceptual output structures have especially high resting activation levels and therefore have privileged processing status. This means that the sounds of language (and its visualisations) are accessible in terms of their surface features at least. In order to be “relatively aware of, say, intonational aspects, syllabic structure or the presence of sibilants” (Sharwood Smith, 2004: 272), the existence of a special interface processor is supposed. This interface processor makes phonological generic structures available for processing. These structures contain less information than the original auditory features of the input, but provide sufficient information for the individual to become aware of sound patterns and basic metalinguistic knowledge (e.g. categories such as syllable, rhyme, intonation). Once the phonological structures are associated with a conceptual representation (i.e. meaning), the activation of the conceptual structure leads to the activation of co-indexed items in the core language system, enabling parallel input processing in the core language system and the meta-grammar. Bearing in mind that the MOGUL framework conceptualises processing as bidirectional, it is thus almost impossible to say whether an L2 production of the learner is governed by the core language system or the meta-grammar. Either can lead to right or wrong processing, depending on the computations of the processors, based on appropriate/ inappropriate mappings or the application of correct/ incorrect metalinguistic information.

“This accords well with Felix’s notion of a competition between an in principle still active UG-driven system and ‘general problem-solving’ principles that older, cognitively more mature L2 learners, especially when exposed to formal classroom instruction, appear to resort to and which might affect their steady progress towards ultimate attainment as evidenced by very young acquirers.”

Sharwood Smith (2004: 274)

This means that it is possible to counterbalance at least some deficits in terms of modular linguistic knowledge by applying appropriate metalinguistic strategies. It also means that, while exposure to naturalistic input induces language growth in the core language module, there is also the possibility of consciously learning about a language, e.g. by learning explicit rules. This notion has received empirical support (see Whong, 2007: 149), but the degree to which it is possible for adult learners of a second language to acquire this language through exposure to input alone, i.e. by acquiring modular, unconscious linguistic knowledge without explicit rule-based instruction, remains uncertain. The point is that the MOGUL framework permits a modelling of both modular development and explicit instruction. Therefore, it is attractive for classroom-based

research on Second Language Acquisition and can serve as the underlying model of input processing for this study on Processing Instruction.

It might be difficult to test some of the presuppositions on which the current MOGUL framework is built. The activation of representations during processing, for example, is difficult to measure for several reasons. If activation is gradual, as assumed by Sharwood Smith & Truscott (2014), then instruments of measurement will have to be extremely sensitive in order to show differences in activation levels. Some representations in Long-Term Memory might receive spreading activation because they are relevant to the processing context, but not necessarily to the input to be processed. Sharwood Smith & Truscott assume that, once an activation threshold is crossed, representations enter consciousness automatically. “Alternatively, it could be that items in the focus of attention possess some neural status that cannot be encoded along the same level-of-activation dimension” (Cowan, 1995: 109).

Those are questions which will have to be addressed in future research if MOGUL (or any other model of language processing) wants to remain competitive. Recent neuroscientific research has stressed the importance of neural connectivity for successful language learning and provided evidence for the hypothesis that the development of language-related functional regions of the brain depends on the context in which the language is learnt as well as on the quality of the resting-state connections between regions (Berken, et al., 2016; Chai, et al., 2016). Efficient processing, in this view, leads to a decrease in activation levels, not an increase, turning the interpretation of thresholds and relative activation levels as well as the interpretation of the role of inhibition on its head. For now, however, the MOGUL framework provides a good basis for the description of VanPatten’s understanding of input processing. It should be noted though that processing in VanPatten’s definition is not concerned with the initial detection or noticing of features in the input. It focuses on the establishment of form-meaning connections (VanPatten, 2005: 277) and the likelihood of these connections being established depends mainly on the characteristics of the features involved.

1.7 Input Processing according to VanPatten

The learner does not have some unlimited supply of attention but instead attention is somewhat fixed and limited. In this sense, a learner's attention would make it difficult to do other similar mental processing at the same time [...]. Because of this limited capacity for processing, we can reframe the "consciousness issue" this way: What gets attended to in the input and what does not?

VanPatten (1994: 28)

Based on this notion, VanPatten begins his exploration of learners' processing behaviour to find out how learners link form and meaning when they process input with the aim of establishing comprehension. He recognises that input processing depends on what learners do with the input rather than how often a form is represented in the input objectively (VanPatten, 2002: 757). Within the Processing Instruction framework, input processing plays a crucial role and is clearly separated from processes related to accommodation and restructuring of intake or output production. Input processing deals with how learners convert input into intake, with intake being defined as those features of the input which are available for further processing in the learner's Working Memory (VanPatten, 2002: 757).

It is important to point out the role of working memory in this model since the first principles are predicated on a limited capacity for processing information; that is, learners can only do so much in their working memory before attentional resources are depleted, and working memory is forced to dump information to make room for more (incoming information).

VanPatten (2005: 268)

Working Memory is thus closely linked with on-line sentence comprehension. Considering that the underlying assumption of the Input Processing Principles is that second language learners are limited-capacity processors of information, "available attentional resources are limited and compete to be allocated to certain aspects of the input" (Leow, et al., 2008: 666). This leads individuals to the development of processing strategies to keep the input manageable while attending to its meaning. These strategies result in processing preferences. "As the learner processes an incoming input string, it must be tagged and coded in particular ways. If acquisition is to happen, the internal processors must eventually attend to how the propositional content is encoded linguistically" (VanPatten & Cadierno, 1993: 227). For English native speakers learning German for example, this means that initially, default processing strategies from the L1 will be applied to German

input strings and subject status will be assigned to the first (pro-) noun in the sentence. According to VanPatten, a shift from inappropriate L1 processing strategies to those adequate for the L2 only occurs if the default strategies lead to incorrect interpretations and a communicative breakdown. As a consequence, more reliable cues for the L2 will increasingly be attended, such as morphological case markers.

During this parsing process, the more salient linguistic surface features (i.e. forms) naturally receive more attention than others, which means that they are more likely to be converted into intake and thus more likely to be integrated into the developing system. One factor that determines the probability of a particular form in the input getting processed and converted into intake is the concept of communicative value: "Communicative value refers to the meaning that a form contributes to overall sentence meaning and is based on two features: [+/-inherent semantic value] and [+/- redundancy]" (VanPatten, 2002: 759). Following this concept, forms in the input are least likely to be attended if they have no inherent semantic value and encode redundant information, such as case inflections on German adjectives, for example:

Ich habe einen guten Platz zum Picknicken gefunden.

[I have found a good picnic spot.]

The –en marking the adjective as referring to a masculine singular noun in accusative case is redundant as the same information is already encoded in the inflectional ending of the indefinite article and can also be deduced from the semantics of the content words. Indeed, German adjective inflections are perceived as particularly difficult to acquire and sometimes even native speakers commit errors by assigning different case markings to the determiner and the adjective, respectively. This difficulty and similar processing problems are the object of VanPatten's (e.g. 1990, 1996, 2002) Input Processing Principles, which will be presented in the next section.

From a MOGUL perspective, the semantic value is important for the processing outside the core language module and for co-indexation of representations inside and outside of the UG-driven system. Redundancy is helpful once the individual modules' representations are established and reinforce each other. If only the salient features are processed, however, the system will not develop representations for the ones that do not become output of the perceptual modules. Processing Instruction supposes that materials that increase the communicative value and frequency of a target item, while eliminating its redundancy, help the learner to acquire it by making robust form-meaning connections more likely. The Input Processing Principles postulated by VanPatten describe the

processing preferences of L2 learners and their strategies of assigning argument structure to a sentence.

1.8 Input Processing Principles

The Input Processing Principles are based on behavioural findings that originate from VanPatten's (1990) study on attention to form and content in the input. There, he explores how much learners comprehend when listening to a text on the consequences of inflation for Latin countries and when listening to the same text while simultaneously trying to note either a key lexical item (*inflación*), the feminine definite article (*/a*), or a bound verb morpheme (-*n*). Attention to meaning and comprehension are measured by a free recall task in the students' first language, English. The different task conditions (i.e. types of attention) are compared based on the scores of an idea unit analysis. Results show that recall scores for listening to the text only and for listening to the text while noting the key lexical item are similar. They also show that recall scores significantly drop when participants listen to the text for content while trying to note a grammatical morpheme. On the basis of these findings, VanPatten assumes that attention to form competes with attention to meaning when L2 learners process input and that linguistic forms can only receive the learner's attention if the input is easily understood (Van Patten, 1990: 296). Considering that VanPatten (1990) finds differences between early-stage and advanced learners in terms of their attention to bound versus unbound morphemes, he posits that with increasing proficiency level, learners are better able to attend relatively non-salient or redundant features of the input (1990: 288–90). Based on this experiment as well as on previous research of other scholars, VanPatten posits the following processing principles:

P1) Learners process input for meaning before they process it for form.

- Learners process content words in the input before anything else.
- Learners prefer processing lexical items to grammatical items (e.g., morphology) for the same semantic information.
- Learners prefer processing “more meaningful” morphology before “less” or “nonmeaningful” morphology.

P2) For learners to process form that is not meaningful, they must be able to process informational or communicative content at no (or little) cost to attention.

P3) Learners possess a default strategy that assigns the role of agent (or subject) to the first noun (phrase) they encounter in a sentence/utterance. This is called the first-noun strategy.

- The first-noun strategy may be overridden by lexical semantics and event probabilities.
- Learners will adopt other processing strategies for grammatical role assignment only after their developing system has incorporated other cues (e.g., case marking, acoustic stress).

P4) Learners process elements in sentence/utterance initial position first.

- Learners process elements in final position before elements in medial position.

Figure 1.3: Processing Principles after VanPatten (2002: 758)

It should be noted that the Input Processing Principles only attempt to describe which features of the input are more or less likely to be processed by the learner, based on their saliency, redundancy, and overall communicative value. This framework does not make any predictions about the learnability of these features, which is also dependent on other factors, such as the complexity of the feature, the frequency in the input, etc. As we have seen in the presentation of the MOGUL framework, language growth depends on the continued re-activation of features through processing. VanPatten's Input Processing Principles do not account for this stage, they only describe how likely a feature is to be detected in the input, not how the output of the perceptual modules becomes input for the language-related modules or how they are stored for future use. The learners' default processing strategies may not always lead to rich intake, they can even lead to wrong mental representations of a second language, for example if the first noun of a sentence is taken to be the subject when it is not. To counter the application of inappropriate L1 processing strategies to the L2, VanPatten devised a pedagogic intervention called Processing Instruction. In this focus-on-form approach, the L2 input is manipulated in a way that makes otherwise neglected features of the input more salient. The next chapter presents Processing Instruction in more detail.

2 PREVIOUS RESEARCH ON PROCESSING INSTRUCTION

Language instruction, especially grammar instruction, for a long time has primarily focused on output practice of the target structure. VanPatten & Cadierno (1993: 225–6) remark that Second Language Acquisition research has been concerned with the reasons for instruction, the content and the timing of instruction, but not with the nature of instruction or the processes that it seeks to manipulate. They argue that grammar instruction should start with the initial stage of acquisition, i.e. at the input processing level, changing the processing strategies of learners and providing them with richer intake for their developing language system. In the Processing Instruction framework, this conversion from input to intake is termed making form-meaning connections and means that linguistic information drawn from the input is integrated into the developing system. Input processing thus “involves those strategies and mechanisms that promote form-meaning connections during comprehension” (VanPatten & Cadierno, 1993: 226).

The argument that input processing should be addressed before output practice is not entirely new, as evidenced by publications such as “The Comprehension approach to foreign language instruction” (Winitz, 1981). This book presents a precursor of Processing Instruction that focuses on the need for learners to understand linguistic structures without recourse to contextual information in order to develop them for productive use. There have also been several other approaches to enhancing the input that learners encounter in classroom settings, enriched input (which exposes the learner to a substantial number of instances of the target item) and enhanced input (in which the target structures are typographically enhanced or orally stressed) being just two examples. Processing Instruction, in contrast, is an explicit form of grammar instruction that aims to change learners’ input processing strategies. It is a focus-on-form approach that follows from VanPatten’s Input Processing Principles and includes a set of precise guidelines (see e.g. Farley, 2005). It must therefore be differentiated from structured-input instruction that works with materials which are not in accordance with the Processing Instruction guidelines and which do not seek to alter inadequate processing strategies (Erlam, 2003a: 560–1).

One of VanPatten’s presuppositions is that unless learners experience the inappropriateness of their default processing strategies, they will not change their assumptions about the target language. Processing Instruction therefore subscribes to the view that learning is failure-driven (VanPatten, 2005: 275); it is an interventionist

treatment that seeks to enrich learners' intake by manipulating the input in a way that forces them to abandon inappropriate processing strategies. While learning is not understood to be failure-driven within the MOGUL framework, Sharwood Smith & Truscott (2014: 300–307) allow for the possibility that instruction can facilitate acquisition if it offers opportunities for straightforward co-indexation of representations in the phonological, syntactic, and conceptual stores – and if it helps learners focus on otherwise non-salient features of the input. VanPatten & Cadierno claim to achieve exactly that through Processing Instruction in their seminal (1993) study.

2.1 Processing Instruction

In the original study, VanPatten & Cadierno (1993) test the impact of Processing Instruction on the learner's developing system (in terms of both input comprehension and output production) and compare the effects to the learning outcomes of Traditional Instruction (TI). The Processing Instruction (PI) group receives information on the target structure (Spanish word order and object pronouns) and on processing strategies that work in their L1 English, but not for the Spanish target structure. Participants in the PI group then engage in referential (picture-sentence matching) and affective (stating a personal opinion regarding the input) Structured Input Activities that require them to process the target structure (i.e. OVS and OV word orders) correctly in order to comprehend the meaning of the sentences. At no point does the PI group engage in output practice. The Traditional Instruction (TI) group receives information on object pronouns and a paradigm of all the forms, then engages in mechanical, then meaningful, then communicative activities. This group does not receive instruction on sentence interpretation and is not warned about inappropriate processing strategies. Materials for both experimental groups are balanced for tokens and vocabulary and the instruction is carried out by the same instructor over the course of two days. The control group meanwhile reads an essay and receives no instruction on either the target item or the processing strategies.

Performance is assessed on a sentence-level interpretation test and a sentence-level production test. Assessment consists of a pre-test, an immediate post-test, and two delayed post-tests (after two and four weeks, respectively). Different test versions are used in a split-block design. There are 20 items (10 target items and 10 distractors) on the interpretation test and 10 items (5 target items and 5 distractors) on the production test. The interpretation task is similar to the one the PI group has practised; the production

task is similar to materials the TI group has received. Correct interpretations are scored with one point. Two points are awarded for correct use and placement on the production task, while the use of either an incorrect pronouns or a wrong word order receives one point. Absence of pronouns is scored with 0. Results show that the PI group gains significantly on the interpretation test and maintains the gain over the 4-week period of testing, while gains of the TI and control groups do not reach significance. On the production test, both the TI and PI groups make significant gains and maintain those gains over the whole testing period. The control group does not make significant gains. Based on these results, VanPatten & Cadierno (1993) conclude that altering learners' processing strategies leads to an alteration of their developing systems and that these alterations can be seen in interpretation as well as production. In this the PI group differs from the TI group, which only gains significantly on the task that has been practised during the instructional treatment. According to the authors, this difference does not constitute a differential effect of training either comprehension or production, but is supposed to be the effect of manipulated input on the developing system (VanPatten, 2002: 772).

In a second experiment, VanPatten & Cadierno (1993) select six second-year university-level Spanish classes from the University of Illinois and, after administering a pre-test, assign them randomly to three groups (Processing Instruction, Traditional Instruction, or No Instruction). The experiment involves 80 participants and is spread over two consecutive days. The participants complete three post-tests this time. The Traditional Instruction group receives explanations of the form, function, and position of direct object pronouns, as well as a chart introducing all pronoun forms. Participants engage in practice activities that move from mechanical output drills and meaningful practice to communicative practice. Production of the target item, both oral and written, is required at all times (VanPatten & Cadierno, 1993: 230). The Processing Instruction group receives instruction on the grammatical concepts of object and subject, the possible word orders in Spanish, and on how to interpret OVS strings correctly. Participants in this group engage in activities that require them to correctly assign argument structure to the input string as well as activities that require them to attend and respond to the content of an utterance or written sentence – they are never required to produce the target item though (VanPatten & Cadierno, 1993: 231). Both groups receive information about object pronouns and where to place them, work with the same vocabulary, and engage in roughly the same amount of practice. What is different is the type of practice: The Traditional Instruction group practises producing the target structure, the Processing Instruction group practises interpreting the target structure.

Assessment consists of a pre-test and three post-tests, all of which involve aural interpretation (matching a sentence with one of two pictures) and written production (completing fragmentary sentences based on a visual clue). The first post-test takes place on the second day of instruction, the second post-test is administered one week after instruction, and the third one a whole month later. Raw scores are calculated for each individual and participants scoring higher than 8/10 on the pre-test (on either the interpretation or the production task) are excluded – as are those individuals that do not show an effect of L1 processing strategies on the pre-test. Participants receive one point each for matching the right picture with the sentence on the interpretation task (for a total of ten points) and two points for correct form and placement of the pronoun on the production task (for a total of ten points), they still receive one point for an incorrect form or placement on the production task (VanPatten & Cadierno, 1993: 234-235). The authors find that on the post-tests of the interpretation task, the Processing Instruction group is superior to the other groups (VanPatten & Cadierno, 1993: 236). On the post-tests of the production task, however, there are no significant differences between the Traditional Instruction group and the Processing Instruction group, both performing better than the No Instruction group (VanPatten & Cadierno, 1993: 237). The authors thus conclude that “processing instruction altered the way in which the subjects processed input, which in turn had an effect on the developing system and what the subjects could access for production” (VanPatten & Cadierno, 1993: 238).

Traditional Instruction, in contrast, apparently did little to improve the processing strategies of the participants. As the gains of the Traditional Instruction group on the interpretation task do not significantly differ from the gains of the No Instruction group on the same task, it is furthermore assumed that the gains of the Traditional Instruction group in terms of interpretation cannot be attributed to instruction at all (VanPatten & Cadierno, 1993: 238). VanPatten & Cadierno (1993) therefore argue that Traditional Instruction does not provide intake for the developing system as this system requires primary linguistic data instead of explicit practice. While they allow for the possibility that learnt knowledge can be acquired through output practice, they advocate Processing Instruction for its ability to affect the developing system of the learner (VanPatten & Cadierno, 1993: 238).

In sum, Processing Instruction is an approach to grammar instruction that focuses on form. Learners are steered away from inadequate processing strategies through confrontation with manipulated input that requires them to make correct form-meaning connections in order to comprehend the meaning of the input. The starting point of

Processing Instruction activities is the question why a target item or structure poses a problem (i.e. which processing principle causes misinterpretations). In the next step, learners are pushed away from this problematic strategy. It is crucial that learners experience the failure of their default processing strategies, so that they are forced to acquire more adequate strategies. “PI is designed to cause failure in interpretation at the beginning stages of activities so that the processors can begin to ‘readjust’” (VanPatten, 2002: 768). This makes Processing Instruction unique and separates it from other forms of input-oriented instruction, e.g. input flood or input enhancement (for a comparison of input enhancement and Processing Instruction, see Agiasophiti, 2011). Quite uniquely as well, learners are not required to produce the target item during the instructional phase. This does not mean, however, that the role of output for language acquisition, especially for the development of fluency and accuracy, is denied. It rather means that there should be a focus on input processing before output practice takes place.

In referential Structured Input Activities, learners must process the target item or structure correctly in order to comprehend the sentence meaning. Answers in this type of activity are objective; they are either right or wrong. Importantly, sentences including the target item/ structure are presented as well as sentences that contain distractors, so learners have to process each sentence for meaning and cannot rely on test-taking strategies. This is a significant difference to mechanical drills in traditional output practice. The next step for learners is to engage in affective Structured Input Activities that require them to express an opinion or belief, or generally to express affective responses while processing real-world information in combination with the target feature (VanPatten, 2002: 766–7). Both types of Structured Input Activities are meaningful and contain a “minimal information load” (VanPatten, 2002: 784), aiming to ensure that learners have sufficient attentional resources to notice and process the target form.

2.2 Critique regarding Processing Instruction

One of the major objections to Processing Instruction is that it seems to be narrowly defined and trivial. But, as VanPatten points out, the link between the Input Processing Principles and the form of instruction requires a precise description of the instructional approach – and while the definition seems narrow, the model supposedly has a wide-reaching impact in terms of predictive power regarding the saliency of particular items or structures (VanPatten, 2002: 768). Research has shown that Processing Instruction, if

compliant with the guidelines, can be a very effective approach to grammar instruction, both in classroom settings and in computer-based individual instruction. Depending on the operationalisation of the ‘Traditional Instruction’ that Processing Instruction is compared to, studies have found significant advantages for learners who received Processing Instruction over learners who received ‘traditional’ grammar instruction.

One of the first studies to scrutinise Processing Instruction is DeKeyser & Sokalski (1996). Their replication of VanPatten & Cadierno (1993) aims to prove that the original findings are not reliable. Seen from a skill-theory perspective, the Processing Instruction results are surprising indeed as either group should only have improved on the skill they had been training. However, while the Traditional Instruction group in VanPatten & Cadierno (1993) has significantly improved their scores on the production task that they have practised during the instructional treatment, the Processing Instruction group has also significantly improved on the production task – despite never having produced the target item during instruction. DeKeyser & Sokalski (1996) therefore compare input and output groups’ learning gains on Spanish object pronouns and word order in one experiment and Spanish on Spanish conditional tense in another. Treatment consists of either input or output practice and progresses from mechanical to meaningful activities for both groups. Performance is assessed in a comprehension test and a production test (translation and fill-in-the-blanks tasks, respectively). For the experiment with object pronouns as a target structure, results show that the input group performs better on the immediate interpretation post-test, while the output group performs better on the immediate production post-test. The delayed post-test does not elicit these performance differences. In the experiment with the Spanish conditional tense as the target structure, the output group has an advantage over the input group on the immediate interpretation and production post-tests, but cannot maintain this advantage in the delayed post-test. The input group, however, maintains the (relatively low) level of improvement over the whole period of assessment.

DeKeyser & Sokalski (1996) look at Spanish object pronoun acquisition from a skill-theory perspective and therefore focus on the different effects of input- and output-based instruction on the respective skills, meaning that their research interest differs from VanPatten & Cadierno’s (1993). Moreover, they do not seek to alter any processing strategies through instructional treatment; which is why they do not use Processing Instruction materials and/or methods. This means that DeKeyser & Sokalski (1996) is no conceptual replication of VanPatten & Cadierno (1993) and that the results obtained by DeKeyser & Sokalski do not challenge those of the original study. Another study that

produces contradictory results to the VanPatten & Cadierno (1993) study is Salaberry (1997). He compares input processing and output processing groups that receive instruction on Spanish object pronouns. Assessment involves a comprehension test and a production test consisting of a sentence-level translation task and a narrative task. Results for both groups show a significant gain from pre-test to post-test on all tasks, but no difference between groups. As in the DeKeyser & Sokalski (1996) study, Salaberry (1997) does not replicate the Processing Instruction treatments because he does not seek to change learners' inadequate processing strategies through the use of Structured Input Activities. In addition to that, the assessment formats differ from VanPatten & Cadierno's (1993) study, making a comparison futile.

Collentine (1998) compares Processing Instruction and output-oriented instruction, using the Spanish subjunctive in adjectival clauses as a target structure. The Processing Instruction group engages in picture-sentence matching tasks and participants are required to respond to sentences in either subjunctive or indicative mode. The output group engages in fill-in-the-blanks tasks that require learners to use the subjunctive or indicative mode, respectively. Assessment involves a picture-sentence matching task as an interpretation measure and a fill-in-the-blanks task as a production measure. Results show that both instructional groups make significant gains, but Processing Instruction does not prove to be superior to output-oriented instruction and gains are not as significant as the ones witnessed in VanPatten & Cadierno's (1993) experiments. While it is not possible to establish which role the explicit instruction on the target item for both groups has played in the outcome of instruction, one can criticise the materials used in Collentine (1998) as not appropriate for the Processing Instruction framework: "learners had to hold a good deal of information in working memory and process a significant amount of the L2 before indicating comprehension or matching a description to a sentence" (VanPatten, 2002: 784).

Collentine (1998) is therefore not accepted as a conceptual replication of VanPatten & Cadierno (1993) by veteran proponents of Processing Instruction. Other researchers, such as Morgan-Short & Wood Bowden (2006: 41), see Collentine's materials as comparable to those used in other Processing Instruction studies and stress the importance of a fair comparison between treatments, which means that all variables should be equal, but the output-based instruction should focus on output production where Processing Instruction focuses on input interpretation. Making Traditional Instruction as meaningless and drill-like as possible, then, amounts to hedging the bets for Processing Instruction.

In sum, the studies providing contradictory evidence to VanPatten & Cadierno (1993) are not accepted as conceptual replications of the original experiments and therefore do not challenge the results: “A conceptual replication study is one, then, that attempts to duplicate the same treatments, procedures, and so on of a study but may alter a variable because the researcher is examining generalizability vis-à-vis that one altered variable” (VanPatten, 2002: 779). As pointed out above, the differential results can be attributed to different research questions and/ or perspectives (especially DeKeyser & Sokalski, 1996; Salaberry, 1997) and to differential approaches to instruction and materials. As Erlam has stressed in her study on the relative effectiveness of different instructional types, there are significant conceptual and methodological differences between structured-input instruction in general and Processing Instruction in particular (Erlam, 2003b: 560–1).

Constructive feedback regarding the basic assumptions of Processing Instruction, on the other hand, has led to improvements of the theoretical and research frameworks. There has been substantial critique regarding the First Noun Principle, especially on the grounds that it seems anglo-centric. One should note though that the original studies (VanPatten & Cadierno, 1993) do not seek to make claims about the relevance of this principle for other constellations than the L2 Spanish acquisition by English native speakers – for which the principle is accurate. Moreover, VanPatten has subsequently reformulated the First Noun Principle and added strategies that may override word order as a strategy for the assignment of argument structure (see e.g. VanPatten, 2002: 758). In addition to that, research on sentence processing suggests that the language processor tends to make the earliest possible choice when assigning subject roles to keep the processing load manageable (e.g. Juffs & Harrington, 1996). Entertaining several potential interpretations of one and the same sentence over an extended period of time would be very costly in terms of processing resources – no matter how extensive or efficient those might be.

In the past decades of Processing Instruction research, many questions and critical points have been addressed, i.e. whether the results of the original (1993) study are generalisable to different target structures and across languages (Benati, 2001, 2005, 2007; Hikima, 2010; VanPatten & Uludag, 2011, among others). In his (2004) presentation of Processing Instruction, VanPatten integrates concepts such as noticing and Working Memory into the framework. While some of the claims still await empirical testing, as we will see, the merit of the Input Processing model lies in the fact that it attempts to explain at the input-level why some features are harder to learn than others and that it provides a practical approach to grammar instruction on the basis of these

insights. “Processing Instruction addresses the processing issues raised by Competition Model researchers, but brings testing into the classroom” (Santamaría, 2007: 32).

2.3 Extending the Processing Instruction framework

A substantial amount of the early Processing Instruction research is dedicated to the comparison of this pedagogical intervention with other types of instruction. While the superiority of Processing Instruction over traditional output-based instruction (VanPatten & Cadierno, 1993; Benati, 2001; Cheng, 2004) and in some cases even over meaningful output-based instruction (Farley 2004; Benati, 2005; Lee & Benati, 2007) is being established, a second research focus emerges with the exploration of what exactly makes Processing Instruction so successful: structured input, explicit information, or a combination of the two (Van Patten & Oikkenon, 1996; Benati, 2004; Farley 2004; Wong 2004). Other studies explore whether the effects of Processing Instruction are durative only or longitudinal (VanPatten & Fernández, 2004), whether Processing Instruction can be delivered via a computer and still be as effective and whether aural or textual enhancement can make it even more effective (Lee & Benati, 2007). Additional research foci concern the effects of Processing Instruction on discourse-level tasks (Cheng, 1995; Benati & Lee, 2010) and on the possibility of transfer-of-training effects on other target features and processing principles (Benati & Lee, 2008; Laval, 2008). The following section summarises the main research foci and their findings and points out the issues and questions generated by previous Processing Instruction research.

New target languages and structures

Cadierno (1995) replicates the original study using a different target structure, i.e. the Spanish preterit tense. “Cadierno (1995) mixed up tenses in her referential activities so that learners had to rely exclusively on the verb ending in each sentence to determine temporal reference (past, present, future)” (VanPatten, 2002: 782). Her results are in accordance with VanPatten & Cadierno (1993). Cheng (1995) uses the Spanish copula verbs *ser* and *estar* in her study and adds a written composition task to the assessment. She can replicate the results of the (1993) study using the established forms of assessment and finds that both PI and TI groups make significant gains on the composition task. Farley (2001) explores the effects of Processing Instruction on the

acquisition of the Spanish subjunctive and finds that significant gains in terms of interpretation and production are made by the Processing Instruction group.

Buck (2000) compares the effects of Processing Instruction and traditional instruction on the acquisition of the English present continuous by native speakers of Spanish. She finds that the PI group makes greater gains and is able to maintain those, while the TI group cannot maintain their gains on the delayed post-test. Benati (2001) also compares these two types of instruction, but uses the Italian future tense as target structure. His production post-test includes an additional oral task where students narrate a sequence of events, based on a series of pictures. Benati's results differ from those of the original study with regards to the fact that the TI group improves on the interpretation task as well, though not as significantly as the PI group. The TI group gains can be explained by the type of activities the group has engaged in during the instructional period, which are more meaningful than in other comparison studies. VanPatten & Wong (2004) replicate the original study using the French *faire* causative as the target structure. Their results are similar to the (1993) study, but due to sampling problems, they cannot find that the PI group maintains their gains on the delayed post-test.

Of specific interest for this study is White et al. (2015) though. The authors compare a Processing Instruction group to Traditional Instruction and No Instruction groups (PI = 24, TI = 24, NI = 8) and explore the effects of instruction on German dative definite articles. This target feature is interesting because all of the definite articles are marked for case and differ from the nominative case articles, hence allowing SVO as well as OVS word orders. White et al. (2015) develop German assessment materials that are similar to VanPatten & Cadierno's (1993) study involving the First Noun Principle, i.e. a picture-matching task where participants listen to 15 sentences and indicate which of the two pictures matches the respective sentence, and a sentence-level production task where participants provide the correct target form, in this case the definite article with the correct dative case inflection. Pre- and post-test performance is measured for five items on the interpretation task and for four items on the production task only. While the authors of this study find that the Processing Instruction group performs slightly better than the Traditional Instruction group on the interpretation measure and equally well on the production measure, they fail to demonstrate the clear superiority of Processing Instruction that has been found in previous studies on different target features and languages. Whether this is due to the small number of tokens that the participants are exposed to during the instructional treatment (35 instead of >120 in earlier studies), the

limited range of potential test scores, or the characteristics of German as the target language, remains unclear though and merits further investigation.

Overall, the initial findings that Processing Instruction groups outperform Traditional Instruction groups on the interpretation tasks and perform equally well as the TI groups on production tasks have been replicated in a substantial number of studies on different language combinations and for different processing problems (for an overview, see Appendix A). The performance gap between PI and TI groups, however, depends largely on the target feature and on how meaningless or drill-like the TI materials are. We will get back to this point when reviewing studies that compare PI to meaningful output-based instruction.

The components of Processing Instruction

The outcomes of the early studies imply that the results of the original study can be generalised to other target structures and other languages. Depending on the operationalisation of “Traditional Instruction”, Processing Instruction seems to be more or less superior to Traditional Instruction when it comes to the acquisition of structures regarding various aspects of the language system: word order, verbal morphology, lexical-aspectual choice, mood selection, etc. (VanPatten, 2002: 775). VanPatten & Oikkenon’s (1996) study shows furthermore that the effectiveness of Processing Instruction is due to the nature of the Structured Input Activities and that neither explicit information nor explicit feedback are as influential as the engagement in those activities. In this study, three groups of secondary-school students in their second year of Spanish are compared: a full PI group replicating the original study, a group receiving explicit information only, and a group receiving Structured Input Activities (SIAs) only. After three instructional sessions, the original assessment materials are used to measure the outcomes. The full PI group and the SIAs group show significant gains from pre-test to post-test, but the group which has received explicit information only does not. VanPatten & Oikkenon suggest that the monitoring function of explicit information does not contribute to the alteration of the developing system (VanPatten, 1996: 126), whereas the processing of Structured Input does. Benati comes to the same conclusion in his (2004) study, but finds minimal positive effects of explicit information. Sanz & Morgan-Short (2004) investigate the effects of explicit information and explicit feedback in a replication of the original VanPatten & Cadierno (1993) study, concluding that explicit information or

feedback is not necessary as task-essential practice is sufficient for development. A more fine-grained analysis of learning rate, however, suggests that explicit information can increase the speed of acquisition – if it is easily applicable and used in tandem with Structured Input Activities (Fernández, 2008; VanPatten & Borst, 2012; VanPatten et al., 2013).

Enhanced Processing Instruction

Based on the findings suggesting that Structured Input is the crucial component of Processing Instruction, Lee & Benati (2007a) attempt to find out whether the effects of Structured Input can be further increased by enhancing the target feature. To this end, they conduct two experiments. For the first one, twenty university students (L1 English) in their first year of learning Italian are randomly assigned to a Structured Input (SIA) group or an Enhanced Structured Input (SIAE) group. Both groups receive the same input sentences in the same order of presentation, but for the SIAE group, the adjective endings are enhanced. This means that the instructor stresses the gender-specific adjective endings when presenting aural input and that the written input is textually enhanced, with the adjective endings bolded and underlined. The participants receive two two-hour instructional sessions on Italian gender agreement on two consecutive days. For the pre- and post-tests, interpretation is measured on an aural task where participants have to indicate which picture is being described by the sentence presented and production is measured via a written fill-in-the-blanks task. The analysis of the pre- and post-test scores leads Lee & Benati (2007a) to conclude that further textual or aural enhancement of Structured Input does not provide any additional benefit since both groups improve significantly and improve in equal measures.

In their second experiment, involving 24 university students (L1 English) in an intermediate course of Italian, they compare Structured Input and Enhanced Structured Input groups that receive instruction on the Italian subjunctive of doubt and come to the same conclusion. Agiasophiti (2011), however, finds an effect of input enhancement in her study on the acquisition of V2 and accusative case marking in German: “Mean score performance reveals an apparent hierarchy, namely Combined >PI>IE>Control group, which is valid both in the short term, one week post-instruction, and in the long term, twelve to fourteen weeks post-instruction” (Agiasophiti, 2011: 174). This is interesting as the performance scores of the four different treatment groups point to several relevant

findings. In this study, the Structured Input Activities are receiving added value by textual enhancement and input enhancement on its own also leads to better gains than no instruction. As pointed out by several researchers (Henry et al, 2009; VanPatten & Borst, 2012; VanPatten et al., 2013), German case marking morphology seems to pose a special problem for learners with English as a first language. This target feature is the only one so far that produces robust findings that suggest that it benefits from additional help in the form of explicit information and/or input enhancement. The implications of these findings for the present study will be discussed at the end of this chapter.

Discourse-level assessment

Another limitation of the original study is the use of sentence-level tasks for assessment. Cheng (1995) addresses this by including discourse-level assessment tasks and can show that Processing Instruction leads to gains on a composition task. She cannot find a superiority of Processing Instruction over Traditional Instruction for this task though. VanPatten & Sanz (1995) address the same limitation and replicate the original study on Spanish object pronouns with 44 college students of Spanish (distributed across a PI and a control group, respectively). They measure the effectiveness of instruction on three different kinds of output tasks: the original (VanPatten & Cadierno, 1993) sentence-level task, a question-and-answer interview task, and a video narration task. All of these tests are administered in both oral and written modes. It should be noted that VanPatten & Sanz use this study to investigate the role of Processing Instruction on output. To that end, the final participant pool for analyses only contains participants who have demonstrated interpretation gains. Due to different item numbers per task and a potential open-endedness of object pronouns on the discourse-level production tasks, the researchers have to convert scores and decide how many objects are obligatory for each discourse-level output that the individual participant has produced. Results show once again that the PI group makes significant gains on the interpretation and on the sentence-level production test in both written and oral conditions.

The other tasks yield more complex results. The question-and-answer test is discarded quickly because it does not elicit as many object pronouns as expected (VanPatten, 1996: 107), but in the more complex video narration task, participants actually try to produce object pronouns. The PI group improves significantly on sentence- and discourse-level tasks in the written mode, but only significantly improves on the sentence-level task in the

oral mode (VanPatten, 1996:107–9). Sanz & Morgan-Short (2004) can replicate these results in their computer-delivered experiment. “It appears that the effects of PI obtain in different types of output tests and are not limited to sentence-level tests. As can be expected, written tests produce performance superior to that of oral tests” (VanPatten, 2002: 789). Hikima (2010) adds to this by showing beneficial effects of Processing Instruction for the interpretation of discourse-level assessment tasks.

Comparison with meaningful output-based instruction

As mentioned before, a fair comparison between input-oriented and output-oriented types of instruction should hold all other variables constant. This means that the materials for both instruction types should involve meaningful activities. In the early 2000s, researchers finally seek to explore the effectiveness of Processing Instruction compared to types of grammar instruction other than traditional output-based instruction. Farley (2001a) for example compares Procession Instruction to meaning-based output instruction (MOI) in the context of the acquisition of the Spanish subjunctive in expressions of doubt. The PI group engages in referential and affective Structured Input Activities and never produces the target structure during the instructional period. Structured Input Activities require the participants in the PI group to attend to subordinate rather than main clauses, thus changing their processing preferences. The MOI group hears or reads main clauses and creates appropriate subordinate clauses using subjunctive or indicative forms, respectively. They also express personal beliefs or doubts using subordinate clauses. Both groups receive two sessions of instructional treatment. Assessment involves an interpretation task based on PI materials and a production task based on MOI materials. The last post-test takes place one month after treatment. Results show that both instructional groups make significant gains and that those gains do not differ between groups.

The gains of the MOI group on the interpretation test are surprising for some, but they might be due to the fact that in this group, participants receive the meaningful output of other learners and the instructor as Structured Input. The MOI group does therefore not engage in output-based activities exclusively. Another explication might be that the Spanish subjunctive seems to be a problematic structure to acquire, as already seen in Collentine (1998). In his (2001b) study, Farley cannot replicate his earlier results as only the PI group is able to maintain its gains on the delayed post-test. “Within MOGUL, the

effectiveness of Focus on Form depends on the extent to which it primes the learner" (Whong, 2007: 152). Processing Instruction is then particularly effective as it involves both the metalinguistic store (explicit information on the target structure, strategies that try to make sense of the language) and the core language module stores (facilitation of form-meaning connections). Constructs such as the subjunctive mood, which is existent, but not often expressed in the learner's L1 English, might then take many processing attempts to be fully acquired as the forms as well as the concept are novel to most students.

Morgan-Short and Wood Bowden also address the question of the relevance of meaningful output in their (2006) study. Their point of departure is the following observation: The more meaningful the Traditional Instruction materials were in previous studies, the smaller was the gap between Processing Instruction groups and Traditional Instruction groups in terms of performance, suggesting that meaningfulness of materials could be one variable to manipulate (Morgan-Short & Wood Bowden, 2006: 37). The authors conduct an experiment using the same target feature as the original VanPatten & Cadierno (1993) study, but they choose to deliver the materials computer-based to have more methodological control. 45 first-semester students of Spanish (L1 English) are randomly assigned to a Processing Instruction group (PI), a meaningful output-based group (MOBI), or a control group (CG). The instructional groups receive the same explanations, but the PI group practises interpretation while the MOBI group practises meaningful production. The control group is exposed to the target form, but is engaged in reading comprehension only.

It should be noted that the cut-off score for inclusion in the study is 33%, which is significantly lower than the usual 60% or even 70% common in Processing Instruction research. Post-tests are administered immediately following instruction and one week after instruction. While the results indicate that both types of meaningful instruction lead to significant improvements in interpretation, the production results indicate that the MOBI group has an advantage over the PI group on the immediate post-test. The significant losses from immediate to delayed post-test for the MOBI group on both interpretation and production tasks, however, lead the authors to consider the possibility that Processing Instruction might induce more stable gains in the longer run (Morgan-Short & Wood Bowden, 2006: 52–3). Due to the time constraints in this study, the longer run is still a sprint, considering that the delayed post-test is administered one week after instruction. The stability of the Processing Instruction gains therefore is still an under-researched field. In addition to that, even this relatively controlled study has another important

limitation: the “MOBI group was not a pure output group because these participants were provided input in the instruction as well as in the feedback. As a consequence, our study is unable to reveal the effects of output alone” (Morgan-Short & Wood Bowden, 2006: 58). Whether a clean-cut separation between the two modes is possible at all or not, it has not been successfully demonstrated yet.

Secondary effects

Testing Lee’s hypothesis that learners who have received Processing Instruction on one target feature will be able to transfer their improved processing abilities to other target features (Lee, 2004: 319), Laval (2008) measures primary effects of Processing Instruction on the French imperfect tense, secondary transfer-of-training effects on French subjunctive mood (same processing principle: Lexical Preference) as well as cumulative transfer-of-training effects on the French causative with *faire* (different processing principle: First Noun Principle). For the experiment, 28 undergraduate students of intermediate-level French (L1 English) complete 6 sentence-level pre-tests (2 per feature, i.e. comprehension and production, 50% cut-off) two weeks prior to instruction, then receive two hours of instruction on the French imperfect tense (PI=13, TI=9, C=6). The immediate post-test measures the participants’ performance on tasks that require the accurate interpretation and production of the primary target feature (imperfect tense), the secondary target feature (subjunctive mood), and a third feature (causative with *faire*). Participants in the instructional groups only receive instruction on the primary target feature, the control group receives no instruction on any of the features. The TI and PI materials are balanced for vocabulary, activity types, number of activities, and practice time.

The results of Laval (2008) show that the Processing Instruction group outperforms the Traditional Instruction and control groups on the interpretation of imperfect tense and that both the PI group and the TI group perform equally well and better than the control group on the production task involving the imperfect tense. For the subjunctive of doubt and the causative with *faire*, the PI group outperforms both other groups in terms of interpretation and production. Laval emphasises that individuals that have received Processing Instruction are more efficient learners due to their ability to make forms available for interpretation and production (instead of Traditional Instruction learners that have been trained on output production). This supposedly leads to an increased ability to transfer

strategies learnt in the context of one verbal morphological marker (imperfect) to another (subjunctive) and even to an entirely different structure (causative). The main reason for TI groups underperforming on PI interpretation tasks though is that TI participants rely on other strategies (lexical, contextual, and/or probabilistic cues). If those cues are taken away, as is the case in PI, they have disadvantages compared to PI participants that have been trained to pay attention to the only available cue. It should also be noted that the Processing Instruction group is far from mastering the secondary feature (mean post-test scores: 3.69 out of 10 for interpretation and 1 out of 20 for production) or the causative construction (mean post-test scores: 4.61 out of 10 for interpretation and 1 out of 20 for production).

Benati & Lee's (2008) study on secondary transfer-of-training effects explores whether learners who have received Processing Instruction on Italian noun-adjective agreement can transfer their improved processing strategies to future tense morphology. The results of this experiment involving 25 undergraduate students (L1 English) studying Italian suggest that this is indeed the case. Participants in the PI group improve in their interpretation and production of the Italian future tense from pre-test to post-test. Benati & Lee (2008: 86) conclude: "These small-to-modest gains suggest to us that these learners are developing a different intuition for processing Italian". Benati, Lee, & Houghton (2008) add to this their findings on secondary transfer-of-training effects from English Past Tense marker -ed to English Third Person Singular -s, supporting the previous findings. The participants in this study are 26 native speakers of Korean learning English in middle school. From a MOGUL perspective, this study is the most clear-cut as Processing Instruction on the primary feature involves explicit information on the processing problem, i.e. the non-salient -ed marker, with the potential of increasing the phonological awareness, which can lead to better input for the developing system: "The result will be conceptual representations of final consonants in CS [conceptual structure(s); S.P.], coindexed with direct representations of the sounds in AS [auditory structure(s); S.P.]" (Sharwood Smith & Truscott, 2014: 305). Once this phonological awareness is improved, it can be applied to other critical final consonants as well.

Token frequency effects

White (in press) reviews studies on the secondary effects of Processing Instruction and seeks to account for recent findings which are not as straightforward as the initial data

supporting the hypothesis that the effects of Processing Instruction can be transferred to other features. The main observation is that the initial treatments consist of more Structured Input items (e.g. Laval, 2008 comprises more than 100 tokens for instruction on the primary target feature, compared to 38 in e.g. Leeser & DeMil, 2013). White points out that this might be the reason why in recent studies no group makes significant production gains for secondary target features (White, *in press*: 7-9). To investigate this more systematically, he conducts an experiment where participants receive Processing Instruction on Spanish word order and 3rd person accusative clitics with varying token frequencies (one group receives 40, one group receives 60, and one group receives 80 tokens of Structured Input featuring the target structure). He then compares how the different groups transfer their learning to 3rd person dative clitics.

226 university students (L1 English) sourced from a beginner-level Spanish course are randomly assigned to one of the three frequency groups (note that entire classes are assigned though) or a control group. After completing a pre-test one week prior to instruction, the participants receive one Processing Instruction session on the target feature lasting between 25 and 35 minutes, depending on the token frequency. They then complete an immediate post-test on the day of instruction and a delayed post-test three weeks after instruction. The materials are based on VanPatten & Cadierno (1993), but focus exclusively on 3rd person forms and are adapted for different token frequencies. Interpretation results show no differences between the groups on the immediate or delayed post-tests, indicating that for interpretation gains even lower token frequencies are sufficient. For the production results, differences do emerge: Only the 80-token group improves on the immediate post-test, but even that group cannot maintain the gains three weeks after instruction (White, *in press*: 20). For the secondary target feature, only the 80-token group gains in interpretation scores after instruction and only the 60-token group demonstrates improvement for the production of dative clitic pronouns (White, *in press*: 23). This points to a non-linear correlation of token frequency with transfer-of-training effects, even if the primary and the secondary feature are very similar. The results should be interpreted with caution though, considering that White (*in press*) assigns entire classes to the treatment types. It is therefore possible that different teachers have provided different input for their students and that the divergent processing histories have led to a non-linear distribution of token frequency effects in this study. Further research seems to be needed, but it seems that token frequency should be substantial in order to see lasting changes at the behavioural level.

Long-term effects

To answer the question as to whether the effects of Processing Instruction are long-term or just observable immediately or shortly after instruction, VanPatten & Fernández (2004) replicate the original study and administer a delayed post-test eight months after instruction. Due to the widely spaced testing and the fact that three entire classes have to be eliminated from the participant pool because they have received instruction on Spanish object pronouns in the meantime, the final participant pool comprises 45 individuals sourced from a US university. They receive two sessions of Processing Instruction on Spanish direct object pronouns and interpret SOV as well as OVS sentence structures in an exact replication of VanPatten & Cadierno (1993). Immediately after the second session, the participants complete a first post-test consisting of ten OVS sentences and five distractors for the interpretation test and five fill-in-the-blanks items with five distractors for the production test. Eight months later, the post-test is repeated.

The results show that the participants improve significantly from pre-test to post-test, but cannot maintain these gains eight months after instruction. They perform, however, still significantly better on the delayed post-test than they do on the pre-test, which VanPatten & Fernández take as confirmation that Processing Instruction has lasting effects on the learner's developing system. They have to concede though that there are serious limitations to their study: Due to the high attrition rates in previous longitudinal research projects, they do not include a control group to avoid ending up with too few participants for a statistical analysis. In addition to that, the participants do not improve as much from pre-test to immediate post-test as previous participant populations (VanPatten & Fernández, 2004: 285–6), which leads to a potential over-estimation of significance. The testing materials allow for only a small range in possible scores to start with, which adds to this concern.

Learning rate

Despite VanPatten & Oikkenon's (1996) study identifying Structured Input as the active ingredient in Processing Instruction, questions regarding the potentially facilitating role of explicit information have remained, especially in view of the prevalence of explicit modes in foreign language teaching. Fernández (2008) re-visits this issue and presents two experiments on computer-delivered Processing Instruction, one targeting Spanish OVS word order with clitic object pronouns and one targeting the Spanish subjunctive of doubt.

In each experiment, she compares a full PI group that receives explicit information and Structured Input to a Structured Input only group. The computer-based instruction allows her to measure reaction times and to explore which group is the first to interpret the input correctly: “The criterion for correct input processing was established as the point when learners began and continued to provide correct responses for at least four items—three target items and one distracter—in a row” (Fernández, 2008: 285). This criterion approach is a new addition to the Processing Instruction framework. While the setting of the criterion value is arbitrary, it is based on the patterns that the participants produce and has therefore internal validity.

For her study with 84 college students (L1 English) in their third year of learning Spanish, Fernández (2008) adapts the instructional materials of the original study for individual computer-based instruction. During the participant screening, students listen to 15 sentences, seven of which are in OVS word order, and indicate the matching picture. If they are not sure, they can choose a “not sure” option. The cut-off score for inclusion in the study is 43%. The participants are randomly assigned to either a full PI group or a Structured Input group. The instruction consists of 30 Structured Input items, including 20 target OVS structures. The results indicate that not all participants reach criterion during instruction and that fewer students reach the criterion in the Structured Input only group (21 compared to 26 out of 42), but this is not a significant between-group difference. No significant difference in response times or accuracy after reaching criterion can be found either. Fernández therefore concludes that explicit information has not played a role for the processing of Spanish OVS sentences in this experiment, potentially because it has no value for the resolution of the processing problem.

In the second experiment, the same 84 participants distributed across the same two groups receive computer-delivered Processing Instruction on the Spanish subjunctive of doubt. During the participant screening, the students hear a subordinate clause in either indicative or subjunctive mood and decide which of the main clauses (one triggering indicative mood in the subordinate clause, one triggering subjunctive mood) on the screen fits. There is also the option to choose “neither” or “both”. Seven main clauses are target clauses that trigger the subjunctive. All participants qualify for the study. The instructional materials are based on Farley (2000) and comprise 30 items, 20 of which are target items. Results indicate that significantly fewer participants in the Structured Input only group reach the criterion: only 21, compared to the 32 in the full PI group. Those participants who reach criterion in the full PI group also need significantly fewer trials until they reach the criterion, show significantly higher accuracy rates after reaching the criterion, and

respond significantly faster, compared to the participants in the Structured Input only group who reach criterion. Fernández concludes that for the Spanish subjunctive of doubt, then, explicit information (EI) has affected the outcomes and has helped learners focus their attention: “It seems that EI helped learners notice forms that might not have been noticed or that might have taken more trials to notice, but it did not obviate or reduce the need for minimal exposure time; that is, correct processing was not instantaneous” (Fernández, 2008: 298). Ultimately, explicit information seems to have differential effects for different processing problems and target features.

Culman et al. (2009) conceptually replicate Fernández’s (2008) study using as a target structure German accusative case marking and word order. In an online format, they provide aural input and evaluate the effects of Processing Instruction on the interpretation of OVS word orders on a picture-matching task using Fernández’s trials-to-criterion approach. One PI group receives explicit information prior to interpreting Structured Input Activities, while the other does not. Like Fernández (2008) in her second experiment, they find an effect for explicit information: Participants who have received explicit information begin to correctly interpret OVS sentence structures before participants who have only received Structured Input Activities. The authors conclude that the differential findings of their study as well as Fernández’s (2008) study are caused by the structure and/or processing problem. This is further explored by VanPatten & Borst (2012) and VanPatten et al. (2013). Both reports support the findings that different structures in different languages, involving different processing problems, benefit from explicit information to different extents and that German seems to be one of the target languages that does benefit from explicit information.

Computer-based Processing Instruction

Due to the fact that Processing Instruction is input-based, VanPatten (1996: 158) raises the question whether it can be delivered successfully via a computer. Lee et al. (2007) attempt to answer this question for Processing Instruction on the Spanish preterite/imperfect distinction and on negative informal commands by comparing three modes of delivery: textbook, computer-based, and hybrid. They source participants from nine intact university-level Spanish courses and end up with 25 participants (10 in the classroom group, 8 in the computer group, 7 in the hybrid group) in total due to very high attrition rates. Six of these participants have learnt another L2 before starting to learn Spanish, but they are included in the analyses as the authors do not find any significant main effects

for previous language study (and because they cannot afford to lose more participants). Based on the Processing Instruction textbook Vitzazos (VanPatten, Lee, & Ballman, 2006), the authors develop a computer-delivered version of the materials that is as close as possible to the textbook and that offers the same additional information that a classroom environment would provide, i.e. there is a dictionary to parallel the classroom option of asking the teacher for a word meaning. The hybrid group receives prints of the website created for the computer-based delivery and participants go through the packages guided by their instructor. As usual, the materials consist of explicit information on the target feature, the processing problem involved, and referential as well as affective Structured Input Activities. Fonts, textually enhanced information, pictures, and audio materials are the same for all groups.

After completing the pre-test two weeks prior to instruction, the participants receive one session of Processing Instruction on the preterite/imperfect distinction and complete an immediate post-test. The following week, they receive one session of Processing Instruction on negative informal commands and complete an immediate post-test. The delayed post-tests are administered one week after the relevant instructional session, meaning that the participants complete the delayed post-test for the preterite/ imperfect distinction before they start learning about negative informal commands. The tests for the first target feature comprise five unrelated sentences with two blanks each and a paragraph of eight related sentences with eight items to supply. The format is multiple choice and students select either the imperfect or the preterite form of the verb in parentheses, meaning that they can provide an answer based on guessing, potentially. For the negative informal commands, five scenarios are described in English and learners are prompted to select the right negative command for the situation from four multiple-choice options. As distractors, present indicative, imperfect, and preterite forms are used and three different versions of each test are generated by altering the order of the items.

Considering that the whole experiment is taking place over the course of five weeks only, with a delayed post-test one week after instruction, this means that the delayed post-test scores might be inflated due to frequency effects. ANOVA supports this suspicion as there are no significant differences between the immediate and the delayed post-test scores. This is a concern that applies to all treatment groups though, and it is not that grave because it is the mode of delivery rather than the instructional content that is of interest in this study. The tests then show that there is no significant difference in the outcomes of instruction, no matter whether Processing Instruction is delivered based on a textbook in a traditional face-to-face classroom setting, individually via a computer, or guided by

the lecturer using print materials. Findings from Lee & Benati (2007a) support this by providing evidence of the equally beneficial effects of face-to-face and computer-delivered Processing Instruction on two target features, the Italian and the French subjunctive of doubt. This is relevant for the present study as the Processing Instruction materials will be delivered via a computer.

But Lee et al. (2007) is interesting for a second reason: For the first time in Processing Instruction research, a look at individual differences is attempted. Analysing the individual score developments from pre-test to immediate post-test and from immediate post-test to delayed post-test, the authors note that the performance of only six of the 25 participants mirror the statistical pattern, leaving room for other mediating factors. They also point to the fact that different participants have problems with the acquisition of different target features. A look at the pre-test scores suggests that the participants who start with low or medium scores improve significantly, whereas the participants who start with high scores do not. Whether this is due to statistical reasons, or a genuine finding, the authors conclude that Processing Instruction levels the playing field for all participants. Needless to say, the data collected by Lee et al. (2007) allow no insights into the backgrounds of the individuals who benefitted the most or the least from Processing Instruction on either of the target features. But they lead to the establishment of a new focus of interest in Processing Instruction research for the following years: individual differences.

2.4 Processing Instruction and Individual Differences

Individual differences constitute a research area of interest because they seem to have a greater impact on second language acquisition than on first language acquisition (Sanz, 2005: 15). Variables such as language learning aptitude and motivation contribute significantly to the prediction of language learning success, with only age of onset being a more reliable predictor (Dörnyei & Skehan, 2003: 589). As VanPatten and Benati (2010: 42) point out, L2 learners can be a very diverse group with respect to their personal backgrounds, their individual learning trajectories, and ultimately, their levels of language learning success. From a psychometric point of view, there are a number of different dimensions (e.g. personality traits such as extraversion, but also motivation, intelligence, etc.) and each individual is at a certain point on the continuum of each of those dimensions. It is the pattern of all those points that makes the individual's approach to (language) learning unique. All those points come together to determine the emotional,

cognitive, behavioural styles and preferences of a person. Traditionally, the problems of individual differences research “concern how to handle multiple measures of related constructs and how to account for a phenomenon, language learning, which is multi-causal” (Skehan, 1991: 292). Even cluster analysis, the technique that Skehan proposes in order to explore more than just additive effects of learner variables, does not necessarily shed light on the interactions of learner variables with each other or with concrete tasks: Sparks, et al. (2012), for example, look into several L1 and L2 measures as well as individual differences and conclude, based on *k*-means cluster analyses, that participants who achieve above-average L2 outcomes perform above the average in most respects, participants who are average achievers show average scores on various measures, and below-average achievements go along with below-average L1 and L2 measures. This, obviously, is a very global and abstract view of the interactions between learner variables and L2 achievements and does nothing to suggest ways of improving the learning experience for students.

Further evidence for the effect of individual difference variables can be found in the earlier Processing Instruction literature: Some post-hoc analyses (as in Lee & Benati with Aguilar-Sánchez and McNulty, 2007 or VanPatten et al., 2009) have documented that not all learners benefitted equally from Processing Instruction, i.e. different participants have improved to different extents on different target features and participants scoring low on the pre-test have improved significantly whereas high-scoring students have not. Identifying the participants who show this sort of pattern does, however, not say anything about the individual differences that caused it – unless selected learner variables are measured as well. To find out more about the learners in question, it is necessary to collect more data than a pre-/post-test quasi-experimental study traditionally does: A selection of learner variables which are likely to play a role in classroom second language acquisition would have to be explored in their relationships to each other as well as in their effect on instruction and its outcomes. Previous research within the area of Processing Instruction has been conducted with the focus on one variable at a time. In this section, research findings of studies on individual differences within the Processing Instruction framework will be presented and the implications for the present study will be highlighted afterwards. Where necessary, links to the wider context of individual differences Second Language Acquisition research will be pointed out, but a complete discussion of learner variables will not be possible.

Age

The role of age has been extensively debated in Second Language Acquisition research, especially with regard to the Critical Period Hypothesis (e.g. Lenneberg, 1967). A review of the relevant literature cannot be given in this context, but a very short summary could be that adult and child language acquisition seem to be qualitatively different, with certain sensitive periods for the acquisition of some aspects of a language, but not for others. In terms of grammar acquisition, for example, articles and plurals seem to be more age-sensitive and less robust than word order (e.g. Johnson & Newport, 1989 and DeKeyser, 2000). For Processing Instruction, this could mean that some default processing rules might be easier to optimise for younger learners than for adults. But:

despite the fact that the age effect is an established factor in second language acquisition, researchers have also found evidence that adults can achieve high levels of ultimate success in the acquisition of a second language. Adults are capable of higher-order language functions and learning strategies.

Lee and Benati (2013: 13)

On the one hand, those strategies might not be relevant in Processing Instruction as the learner has no choice of learning strategies, which could be a disadvantage for adult learners. On the other hand, keeping in mind that the age-sensitivity of a structure might be correlated with its saliency (DeKeyser, 2012), the very nature of Processing Instruction might hold the key to this problem and therefore be equally efficient for both child and adult second language acquisition. If the frequency and complexity-opacity of the target feature are also taken into account, the untypically inconsistent research findings on Processing Instruction and the Spanish subjunctive might make sense after all.

As argued by Birdsong (2005, 2006), age – in terms of second language acquisition – is not only a biological factor, but also impacts on the context in which learning takes place as well as the cognitive, motivational, and behavioural circumstances. It makes therefore little sense to look at those factors in isolation. While the initial studies on Processing Instruction focussed on university-level learners of a second language (with the exception of VanPatten & Oikennon's (1996) study on high-school students), school-aged learners eventually get onto the research agenda – with Benati (2005) researching Greek and Chinese children (12 – 13 years of age) learning English and Lee, Benati & Houghton (2008) looking into the effects of Processing Instruction on middle-school-aged learners of L2 English. Benati's (2013) study on 'Age and the Effects of Processing Instruction on the Acquisition of English Passive Constructions among School Children and Adult Native

Speakers of Turkish', however, is the first to directly compare groups of child and adult learners, namely children of 13–14 years of age and young adults of 19–21 years of age, all being Turkish native speakers.

None of the participants has had formal exposure to the target structure, the English passive construction. Both groups receive four hours of Processing Instruction with identical materials over the course of two days. The post-test takes place immediately after instruction and the delayed post-test is given three weeks later. Learners are assessed on a real-time aural interpretation task that requires them to match a picture to a sentence they have heard and on a written sentence-completion task that requires them to fill in the correct passive constructions of the verb in parentheses. As reported in the study, both the children and the adults start out with similar knowledge of the English passive construction and their gains from pre- to post-tests on the aural interpretation task and the written sentence-completion task are not significantly different between the two groups (Benati 2013, 94–5). Benati consequently concludes that Processing Instruction is equally beneficial for child and adult learners.

Another study investigating the benefits of Processing Instruction for school-aged learners is Laval's (2013) replication of a study on Transfer-of-Training effects reported in Benati, Lee & Laval (2008). While the earlier study researches university students, the replication investigates school-aged learners. Students in the Processing Instruction group receive explicit information on the French imperfective past tense and engage in referential and affective Structured Input Activities. After two hours of instruction, they complete the immediate post-test on the imperfective past tense (target feature), the subjunctive (to test for secondary transfer effects), and the causative with *faire* (to test for cumulative transfer effects). Assessments consist of a sentence-level interpretation task and a written sentence-level fill-in-the-blanks task for each feature. They take place three weeks prior to instruction (pre-test) and immediately after instruction (post-test).

Compared to the adult participants of the (2008) study, the children's scores in the replication study improve slightly more from pre-test to post-test (+64% as opposed to +58%) for the target feature interpretation and production (+159% as opposed to +140%). For the secondary feature that has never been taught to the participants, the children score significantly higher than the adults on the interpretation task: +55% as opposed to +20%. The production score gains for the secondary feature (French subjunctive mood), however, are reversed: While adults in the (2008) study improve by +10%, the children score zero points in both the pre- and post-test, effectively showing no improvement at

all. For the cumulative transfer-of-training effect, learners have to interpret and produce French causative structures with *faire*. Again, these have not been taught at any point. Compared to the adults in the (2008) study, children improve slightly more on the interpretation task (+50% as opposed to +34%) and quite significantly more on the production task (+30% as opposed to +10%). The findings lead Laval to conclude that, overall, Processing Instruction is as effective for younger learners as it is for older learners (Laval, 2013: 125).

In two parallel classroom experiments, Angelovska & Benati (2013) directly compare the performances of German adult university students (mean age: 26 years) and school-aged learners (mean age: 10.5 years). Both groups, over the course of two days, receive two hours of Processing Instruction on the English simple past tense, a feature of which they have no previous knowledge. One experiment adapts an existing interpretation task (see Benati, 2005), the other experiment utilises a modified interpretation task that is designed to focus learner's attention on the aspectual distinction between present perfect and simple past tense – a distinction which can be hard for German native speakers as the equivalent tense forms in their L1 can be used interchangeably.

Pre-tests are administered one week prior to instruction, one post-test immediately follows instruction, and a delayed post-test takes place two weeks after instruction. On the “traditional” interpretation task, both groups perform equally well. For the newly designed interpretation task, however, the authors find significant differences between the two group scores: Adult learners perform significantly better than school-aged learners. According to Angelovska & Benati, this is due to the task demands and individual processing capacities of the participants (Angelovska & Benati, 2013: 142). Seeing that one of the main objectives of Processing Instruction is to structure the input in a way that taxes the processing resources of the learners as little as possible, it is then questionable whether the modified interpretation task should be counted as a Structured Input Activity following in the tradition of Processing Instruction.

While the above-mentioned studies have concentrated on the comparison of adult learners with children, Mavrantoni & Benati (2013) investigate the acquisition of the English present tense third person singular by Greek school-aged learners of two different age ranges (8–10 and 15–17 years of age, respectively). This distinction is interesting because it translates as comparing a pre- and a post-puberty group, which can have implications for second language acquisition. Participants are enrolled in two different private schools and have no prior knowledge of the target feature. The pre-test is

administered one week before instruction, then participants are assigned to one of two treatment groups (Processing Instruction or Traditional Instruction). Instruction takes up three hours in total, split over two consecutive days. The post-test immediately follows instruction.

Assessment consists of a sentence-level comprehension task that requires participants to decide whether a verb refers to a single agent or to multiple agents and a written fill-in-the-gap task that requires participants to produce the correct form of the verb in parentheses. Results show that the pre-puberty group that has received Processing Instruction improves their interpretation scores by 55% from pre- to post-test, whereas the Traditional Instruction group does not improve. On the production task, both pre-puberty groups perform equally well in statistical terms. The post-puberty group that has received Processing Instruction improves by 59% on the interpretation task, but the Traditional Instruction group does not improve. Both post-puberty groups perform equally well on the production task in statistical terms. Comparing the scores of the pre- and post-puberty groups, there seems to be no statistically significant difference in terms of their performance. From the data collected, Processing Instruction seems to be equally beneficial for learners, pre and post puberty.

Unfortunately, the design of the studies described above makes it impossible to evaluate any potential qualitative differences in the learning process or the involvement of any mediating variables. In Benati's (2013) study, no significant differences show in a direct comparison of children and adults learning the English passive construction. This seems to suggest that both groups encounter and overcome similar challenges. Future research, however, should also take into account that different target features and structures might elicit different outcomes, especially in view of the fact that the First Noun Principle has been shown to affect both child and adult second language acquisition (see for example Bever, 1970; Nam, 1975; LoCoco, 1987; Lee, 1987). Other features and structures might be more age-sensitive.

Laval's (2013) replication of the Benati, Lee & Laval (2008) study on transfer-of-training effects shows that, in comparison, children's scores improve more from pre- to post-test on most tasks, raising questions as to whether the children have learnt at a higher rate than the adults. Without a more in-depth analysis of learner profiles and the learning process, and a bigger sample size, findings are hard to explain, however. Angelovska & Benati's (2013) study supports the evidence found in previous Processing Instruction research when it comes to the benefits of this pedagogic intervention for learners of

different first languages and age ranges and for a variety of target languages and features. It also shows that Structured Input Activities are designed in a particular way for a reason and that even slightly different task designs can lead to vastly different results. All this points to the necessity of analysing task conditions and perceived task difficulty in relation with the learner's profile and the characteristics of the target feature or structure. Mavrantoni & Benati's (2013) study does not show any statistical differences between pre- and post-puberty groups, but one cannot tell whether that is due to the fact that Processing Instruction is equally beneficial for learners of all ages or whether the lack of differences was due to the target feature. The simple present third-person –s is notorious for its late acquisition even in L1 contexts and might not be as age-sensitive as other structures. Further research on this hypothesised biological threshold, in combination with other potential background factors, seems therefore necessary.

Gender

Generally speaking, studies on gender within the area of Processing Instruction research are rare. Agiasophiti (2013) investigates the acquisition of German accusative case marking by secondary-school pupils in the UK and strives to find out whether textually enhanced or unenhanced Processing Instruction materials, respectively, affect the acquisition process differentially with regards to the gender of the learners. Participants are aged between 12 and 14 years and attend five different secondary schools. All classes are reported to be of mixed abilities, but most participants are placed at the A2 level of the Common European Framework of Reference for Languages. The author compares three randomly assigned groups: an enhanced Processing Instruction group, an unenhanced Processing Instruction group, and a control group. The pre-test takes place one week prior to instruction, then the participants in the treatment groups receive 90 minutes of Processing Instruction on German accusative case marking in object-verb-subject sentences. The first post-test immediately follows instruction; the delayed post-test takes place 12 to 14 weeks afterwards.

German word order is more flexible than English word order and most phrases (except for the finite verb) can move freely within a sentence. Object-verb-subject word order does not convey a different meaning than subject-verb-object word order, but it puts emphasis on the object. This is even more salient in spoken German as the object (phrase) in an OVS sentence structure is naturally stressed. In keeping with the First

Noun Principle, learners of L2 German might resort to taking the first noun (phrase) in a sentence to be the subject of the sentence. In the OVS word order, this leads to a wrong interpretation of the sentence. While a speaker conveys OVS word order by stressing the dislocated phrase, there often is no telling in written German. The only possibilities to discriminate between a direct accusative object and a subject in this context is to pay attention to the inflected definite article accompanying masculine singular nouns or to check person agreement of the verb. In Processing Instruction, however, instructors aim to design activities which force learners to attend to a specific feature in the input and make one-on-one form-meaning connections, so the focus in this study is on the inflection of the masculine definite article.

Due to the two experimental conditions (textually enhanced vs unenhanced Processing Instruction), all activities designed by Agiasophiti (2013) are in written mode. “For the coloured typographically enhanced group (+IE+PI) cases were enhanced with different colours in bold, i.e. for the Masculine Nominative light blue; for the Masculine Accusative dark blue; for the Feminine Nominative and Accusative pink and for the Neuter Nominative and Accusative, green” (Agiasophiti, 2013: 165). For the assessment, learners complete an interpretation task that requires them to match an English translation to a German sentence and a production task that requires them to fill in the correct form of the definite article. Instruction and assessment materials are all computerised. Results show that for the interpretation task, participants of both treatment groups (enhanced as well as unenhanced Processing Instruction) and both genders improve equally from pre- to post-test. For the production task, the author cannot find any significant statistical differences between genders, but the enhanced Processing Instruction group performs significantly better on the immediate post-test. Based on these findings, Agiasophiti concludes that gender has no effect on the outcomes of Processing Instruction, whether the materials are enhanced or unenhanced. It must be noted though that there is great variability within the groups, which can be due to the fact that participants are sourced from different schools – or it could be due to individual differences which have not been the object of investigation in this study.

More recently, the potential mediating effects of sex hormones on processing have been investigated within the Declarative/Procedural model (see Ullman, 2005 for a short introduction). Depending on the oestrogen level, learning success is hypothesised to vary. This is a path for potential future research:

Moreover, they [Ullman and colleagues; S.P.] hypothesized that due to their verbal memory advantage, females will tend to memorize previously encountered complex forms (e.g., regular past tense forms) in the declarative system, while males tend to rule-compute them in the procedural system in real time (Ullman, 2004; Ullman et al., 2002; Ullman, Maloof, Hartshorne, Estabrooke, Brovett, and Walenski, 2005). Both sexes must memorize irregular forms, whereas novel complex forms would be expected to be composed

Wood Bowden, Sanz, & Stafford (2005: 113)

First Language

In their (2013) study on Processing Instruction targeting the Spanish subjunctive/indicative contrast after the adverb *cuando*, Lee & McNulty look at the participants' language background as a variable. The participant pool consists of intermediate-level university students of Spanish and the different linguistic backgrounds are described respectively as 'English' (L1 English with Spanish as the only other language), 'English+' (L1 English with other L2's) and 'Other' (L1 other than English, L2 English and Spanish) (Lee & McNulty, 2013: 57–8). The students receive three hours of Processing Instruction (1 hour of explicit information and two hours of practice) over the course of two days. Post-testing is conducted seven to nine days after instruction due to curricular constraints. Participants complete two listening (comprehension and production), two reading (comprehension and production), and two composition tasks for both the pre- and the post-test.

It is noted that participants who do not have English as their first language generate more contexts for subjunctive use after *cuando* in the pre-test composition task, but they do not produce more subjunctive forms because they have had no prior formal exposure to the Spanish subjunctive (Lee & McNulty, 2013: 62). The difference in context generation scores disappears after instruction and there are no significant differences between groups on the post-test guided composition task. The initial divergence is attributed to the fact that the participants in the 'Other' language group have "broader language learning experiences" (Lee & McNulty, 2013: 65). Having a look at the participants' linguistic profiles presented in Benati & Lee with McNulty (2010, 129) though, this explanation is not satisfactory: The native English speakers in the 'English+' group report having experience with at least one other language – lasting between seven and twenty years. Assuming that the observed pre-test scores are not random, then the cause is likely to be more complex.

The experiment described above does not reveal any significant differences between the three groups described, prompting the authors to conclude: “Language background does not appear to offer any advantages or disadvantages in terms of the forms produced after instruction” (Lee & McNulty, 2013: 68). The authors, however, do suggest further research on the variable of language background, especially in relation with language aptitude and Working Memory. This fits with the general observation that Processing Instruction has been shown to be an effective pedagogic intervention for the improvement of processing strategies related to a multitude of target features. While the earlier studies have focused heavily on Romance languages, more recent studies have investigated target languages such as Japanese and Russian. Different native languages, such as Chinese, German, or Greek, have also been studied and are featuring in the current Processing Instruction research agenda (for a review see for example Lee & Benati, 2009).

Aptitude

In their (2012) study, VanPatten & Borst examine the effects of explicit information and grammatical sensitivity in an experiment with 46 English native speakers learning L2 German at college-level (third semester). All of the participants receive computer-delivered Processing Instruction on nominative/accusative case distinctions and have to interpret the subject and object roles within sentences by selecting the drawing that matches the aural input they hear. Only half of the participants receive explicit instruction (in English) on the difference between subjects and objects, German case marking for nominative and accusative definite articles, German word order patterns, and the dangers of applying the First Noun Principle to German OVS sentences. Both groups are exposed to a total of 50 sentences, with one SVO sentence following three OVS sentences.

During instruction, trials to criterion are monitored, i.e. it is measured how many trials it takes the participants to process three subsequent OVS sentences and one SVO distractor sentence correctly. According to the authors, participants score above chance level once they have reached the criterion, which is taken as an indication of their processing success. The authors find that the explicit information helps the participants in processing the case markers accurately as they reach the success criterion before the group which has not received explicit information. In fact, half of the explicit-instruction group immediately and consistently process OVS sentences correctly (VanPatten & Borst, 2012: 101), pointing to a significant effect of explicit information on processing of

German case marking in sentences which do not follow a canonical word order. These findings are in line with Henry et al. (2009), who also find an effect of explicit information in Processing Instruction targeting German case marking.

They also find that the grammatical sensitivity measure is weakly correlated with the performance of the explicit-information group, but not with the performance of the group that has not received any explicit information. Due to the small sample size and an unequal distribution of pre-test scores, the results have to be treated with caution though. Choosing the nominative-accusative distinction limits the scope of the instructional materials as only the accusative masculine singular forms differ from their nominative counterparts, potentially simplifying the teaching of German case marking more than is warranted. Why grammatical sensitivity should play a role in German case marking acquisition, but not in other target languages and structures that concern the same processing problem, namely the First Noun Principle, is not clear yet –especially when considering that similar experiments on Russian case marking and the First Noun Principle resemble the findings from Spanish studies involving clitic object pronouns rather than the German case marking study (VanPatten, et al., 2013: 518).

One possible explanation that the authors offer for the differential effects concerns the nature and applicability of the explicit information that is provided: the more straightforward the form-function mappings and the more easily applicable the information, the more beneficial explicit information seems to be (VanPatten, et al., 2013: 522). As to grammatical sensitivity: Processing Instruction might not be strongly correlated with this kind of aptitude because the goal of this pedagogical intervention is not the internalisation of rules, but the acquisition of reliable processing strategies. It is all the more surprising, then, that grammatical sensitivity is correlated with the outcomes of Processing Instruction on German nominative-accusative case markers, and on those specifically. VanPatten & Borst (2012: 104) conclude thus:

Perhaps there is a relationship between grammatical sensitivity and the ability to use EI during a processing task, a relationship that may involve a third variable such as working memory. Only future research that uses a multitude of measures of individual differences could address this issue.

Up until the present study, no Processing Instruction study has attempted to do this.

Working Memory

While the interest in individual differences has increased within the Processing Instruction framework recently (for an overview, see Lee & Benati, 2013), the impact of Working Memory Capacity on the outcomes of Processing Instruction has not received much attention yet. While some researchers hypothesise that Working Memory might have an impact under certain circumstances and for particular target features, for example with regards to the application of explicit knowledge (VanPatten & Borst, 2012), others have claimed that the nature of Processing Instruction makes it generally beneficial for learners with different backgrounds and abilities (Lee, 2004). However, there is one PhD dissertation that investigates the influence of Working Memory on learners' performance after receiving Traditional Instruction (TI), Processing Instruction (PI), or no instruction (Santamaría, 2007). The study targets French clitic objects and the author seeks to explore the effect of instruction type incorrect processing strategies as well as the relationship between Working Memory and type of instruction.

While previous research on Spanish has shown that Processing Instruction is a very effective approach to teaching clitic pronouns (VanPatten & Cadierno, 1993; VanPatten & Fernández, 2004; VanPatten & Oikkenon, 1996; VanPatten & Sanz, 1995), the results of Santamaría's (2007) study cannot support these findings for the interpretation and production of French clitics. Part of the problem might be that in French, clitics are not very salient. While they have the syntactic characteristics of a word, they often depend phonologically on another word or phrase. Their grammatical function is more pronounced than their lexical meaning – a condition which is addressed in the Primacy of Meaning Principle and its sub-principles (VanPatten, 2004). Additional processing problems are related to the often sentence-medial location of clitics and their semantic redundancy as placeholders, which means that they are unlikely to be processed, according to VanPatten's Input Processing Principles (e.g. *Preference for Nonredundancy Principle* and *Sentence Location Principle*). This is especially relevant for oral comprehension, as Erlam points out: "in spoken French the schwa vowel of the object pronouns me/te/le may be weakened to the extent that it is almost imperceptible" (Erlam, 2003a: 245).

In order to force learners to alter their inappropriate default processing strategies, Processing Instruction confronts them with Structured Input that makes use of alternative word orders. These non-canonical words orders are created with the help of a phenomenon that occurs in colloquial French: dislocation. Dislocation means that a noun

phrase or pronoun is put to the left or the right of a syntactically complete sentence. “By allowing different word orders, dislocation separates the topic from the comment. It introduces a potential psychological subject while maintaining the grammatical subject” (Santamaría, 2007: 7). Using dislocation, it is therefore possible to create sentences with flexible word orders and stress certain parts of the sentence without violating the global sentence intonation. However, dislocation makes it necessary to use a clitic that stands in for the dislocated element in its original position. This clitic is a co-referent and agrees with the dislocated element in case, gender, and number. Even when using correct processing strategies, these features require a lot of information to be kept in an activated state: The referent must be maintained in Working Memory and its connection with the clitic must be established, all while parsing the rest of the sentence.

Possible findings, according to (Santamaría, 2007: 41–2), are:

- 1) those with high working memory may take advantage of treatment (PI or TI) while those with low working memory will not benefit.
- 2) learners of all working memory capabilities may benefit from PI’s strategies and activities.
- 3) the high working memory group may perform better than the other group, but the subgroup that is high working memory and receives PI will score higher than all other subgroups.

Santamaría predicts that the third possibility will be supported in the experiment, i.e. that the combination of high Working Memory Capacity and Processing Instruction will be most effective when it comes to the acquisition of French direct object clitics.

The participants in this study are English beginner-level learners of French, they have received instruction on clitic pronouns before and are familiar with SOV sentences that contain direct object pronouns. All participants complete a pre-, post-, and delayed post-test, as well as a PowerPoint version of the Waters & Caplan (1996) reading span test and a Language Background Questionnaire. The instructional period and the immediate post-test take place one day after the pre-test and the delayed post-test is administered two weeks later. Included in the data pool are 65 participants for the interpretation task and 79 participants for the production task, reflecting the fact that the participants generally did better on the interpretation pre-test than on the production pre-test. Participants are randomly assigned to one of the instructional groups; the Working Memory Span groups are derived by a median split:

	Traditional Instruction	Processing Instruction	No Instruction
Interpretation	15 high-span individuals 10 low-span individuals	9 high-span individuals 15 low-span individuals	5 high-span individuals 5 low-span individuals
Production	14 high-span individuals 14 low-span individuals	14 high-span individuals 16 low-span individuals	11 high-span individuals 10 low-span individuals

Table 2.1: Participant distribution in Santamaría (2007)

The Processing Instruction group (PI group) receives explicit information on the grammatical phenomenon of dislocation and information on the diverging processing strategies applicable to the L1 and L2 and then engages in Structured Input Activities (two referential and four affective in nature). The Traditional Instruction group (TI group) receives the same explicit instruction on dislocation as the PI group, but is not taught about differing processing strategies. Instructional materials are taken from French textbooks and show SOV word order. Activities include oral and written exercises that focus on practicing dislocated sentences; they do not require the learners to process the direct object pronouns for meaning. In contrast to the PI group, participants in the TI group are required to produce the target item during the instructional treatment.

On the interpretation task, all learners improve their raw scores from pre-test to immediate post-test, but scores decrease from immediate to delayed post-test (Santamaría, 2007: 79). The Processing Instruction group does not outperform the Traditional Instruction group on the immediate post-test and the delayed post-test scores are even lower than the Traditional Instruction group's. An item-by-item analysis for the five different word orders occurring in the pre- and post-tests indicates that the learners of each group did best on those word orders to which they had been exposed the most, canonical word order for the Traditional Instruction group and non-canonical word orders for the Processing Instruction group. This might explain why the PI group does not outperform the TI group on the interpretation task: The SVO sentences in studies on clitics are usually not counted as test items, but as distractors (VanPatten & Cadierno, 1993), stacking the odds in favour of the PI groups. On the production task, the pre-test scores show that the PI group has a head start, but the TI group outperforms the PI group significantly, which is not the case in other studies comparing Processing Instruction and Traditional Instruction (e.g., VanPatten & Cadierno, 1993; VanPatten & Wong, 2004).

This leads to the second question, whether participants with certain cognitive resources benefit more from one type of instruction than another. Working Memory Capacity is measured in Santamaría's study using a PowerPoint version of the Waters & Caplan

(1996) reading span test. Span scores are calculated using the total number of points achieved and participants are classified as belonging to the low-span or high-span group following a median split through the results. It might seem problematic that the high-span and low-span participants are not distributed equally across instruction groups, but the main problem is the median split: "Splitting a distribution (e.g., at the median) inflates error variance by failing to use the discriminating power of the continuum" (Snow, 1991: 209). Moreover, participants with the same Reading Span score can (and did in Santamaría's case) end up in different groups.

Several participants in Santamaría's study have knowledge of a language related to French. Speakers of e.g. Spanish or Italian might very well be familiar with the role of object pronoun clitics in the context of dislocation. This familiarity might have influenced the outcomes, depending on the proficiency level and exposure to other Romance languages. These factors cannot be controlled, especially not when dealing with small participant samples, but they should have been treated more explicitly. Additionally, some of the sentences used in the materials are clearly favouring one interpretation over another, as predicted by VanPatten's Event Probabilities Principle (VanPatten, 2002: 758). In addition to that, only the B test included verbs that required different (direct/indirect) objects in L1 and L2 (Santamaría, 2007: 98). Unfortunately, it is not clear from the dissertation which group(s) received which pre- and post-tests, making it impossible to gauge the effects of the testing materials on the learners' performance.

All in all, there seem to be several factors that have influenced the outcomes of Santamaría's (2007) experiment: exposure to the target language before and during instruction, proficiency level and linguistic background, operationalisation of the pedagogical interventions, Working Memory Capacity measurement and subsequent participant classification, as well as the design of testing materials. Considering that it is unlikely to find a participant sample large enough for group comparisons, and that it would be impossible to control all the above-mentioned variables, it seems that the way forward is to embrace the diversity and conduct a small-scale study which focuses on Processing Instruction and individual differences in Working Memory Capacity, while at the same time keeping track of some potentially mediating variables.

2.5 Summarising the gaps: Implications for this study

After this presentation of previous findings from Processing Instruction research, it is time to take stock of the problems and questions that have been generated and to evaluate their implications for this project. The most persistent observation is probably that even the faithful PI studies vary a lot in terms of materials of instruction and assessment, criteria for participant inclusion, timeframes of data collection, and exposure to input. This makes them hard to compare. Most PI studies have a delayed post-test after one to four weeks, with eight months in VanPatten & Fernández (2004) and 12–14 weeks in Agiasophiti (2011) being rare exceptions. If one wants to find out whether there is any long-term effect of Processing Instruction, then a delayed post-test that takes place as far apart from the instructional period as logically possible is necessary. While learners can entertain several different solutions for one problem, in some cases never giving up inadequate structures, even if a native-like alternative has been acquired in the meantime, one can differentiate between a developing system that does not contain a certain feature of the language and one that does. The question is how the acquisition of a certain feature can be shown. “Whether, like Pienemann, you opt for the ‘emergence’ criterion—it is acquired when it shows up for the first few times in the performance of the learner—or whether you say it is acquired when it appears 80 or 90 per cent of the time is immaterial (Pienemann 1998: 137f)” (Sharwood Smith & Truscott, 2005: 222), there is a change in the learner’s behaviour. If this change persists in the delayed post-test, with no instruction on the target item provided in the meantime, it should be reasonable to say that change has taken place in the developing system. Therefore, the delayed post-test scores, maybe more than the immediate post-test scores, are of interest. Data collection should therefore aim to make optimal use of the whole term. Re-exposure to the target feature between immediate and delayed post-test should not be an issue as the undergraduate students hardly receive any explicit grammar instruction in their programme, as will be discussed later.

As White (in press) has pointed out, some of the PI studies have very short treatment periods: Smith & VanPatten (2013), for example, expose their *ab initio* learners of Japanese (L1 English) to 100 sentences within thirty minutes and claim afterwards that this PI treatment has helped the participants with their parameter setting. Culman, Henry & VanPatten (2009) use 30 input sentences and their participants finish the computer-based instructional session within 26 minutes. From a MOGUL perspective, the processing of features in the input leads to the acquisition of these features, so exposure

and accessible input are seen as crucial in this framework. While it seems relatively easy to statistically demonstrate an improvement from no knowledge of the L2 to a little knowledge of a particular feature (see VanPatten & Fernández, 2004), the *ab initio* condition means that vocabulary has to be taught before any relationships between the words can be addressed. The participant population for this study is going to include high-intermediate learners of German who have so far failed to acquire a particular target feature, so it seems necessary to expose them to as many instances of the target feature as possible in the attempt to achieve something that they have not achieved in years of learning German. These are clearly new circumstances in Processing Instruction research.

Another issue with the studies mentioned above is that the criteria for inclusion in the final data pool can be very arbitrary. Smith & VanPatten (2013), for instance, include only participants who demonstrate grammatical sensitivity (measured via a grammaticality judgement task, which is another concern) as well as at least a 65% accuracy rate on the interpretation task, to show that minimal exposure to the L2 can lead to parameter setting. This is circular reasoning. While it is understandable that the researchers are interested in addressing specific research questions, the discarding of participants' data without looking at why they differ from other data seems hasty and too convenient. For this case study, all participants that score 60% or lower on one of the pre-test measures will be included as it seems probable that different participants will perform more or less well on different measures. There might also be students who are capable of interpreting case inflection, but who cannot produce it accurately yet. In that case, it would be interesting to see whether intensive processing of the target feature can add some critical mass and enable learners to produce the target feature more accurately and/or more often. In addition to that, more and longer assessment tests will be used to allow for a wider range in possible scores, addressing the problem that most PI pre- and post-tests only include five to ten items per interpretation and production task, respectively. With the more advanced learners in this study, it seems crucial to look into tasks of different difficulty levels. The participants will also be prompted to rate the task difficulty as any discussion of the nature of assessment tasks in Processing Instruction research has only ever included the researcher's analysis of task demands at best, but has not given the participants an opportunity to indicate how they perceived the task difficulty.

Related to the above-mentioned problem is another assessment issue: In most Processing Instruction studies, interpretation is tested by using the same format as in the Structured Input Activities that the participants in the PI groups are familiar with, but on

which the Traditional Instruction groups have a disadvantage. Hikima (2010) and Agiasophiti (2011) are exceptions in as far as they look into discourse-level interpretation as well. However, none of the studies that are accepted as genuine Processing Instruction studies has explored non-PI comprehension materials to date. Again, with a more advanced participant sample, it seems that learners should be able to transfer the processing experience from the instructional sessions to more complex and therefore more authentic assessment tasks, such as a sentence-level translation task.

Some of the studies presented above (e.g. Fernández, 2008) have used a screening test, but then focused on the participants' on-task performance rather than the products of instruction. While the trials-to-criterion approach is an interesting addition to the Processing Instruction research methodology, it only has internal validity. The definition of the criterion, i.e. when a feature can be considered as acquired, is arbitrary. It may be checked against performance scores before and after criterion is reached, but it is not an objective measure of acquisition. As Sharwood Smith & Truscott (2014: 248–9) emphasise: There is no end state of language acquisition, if the target feature is not being processed, it will become less and less likely to be available for processing. And even if the activation levels are high, they do not always lead to appropriate processing outcomes as the target feature has competition. Initially, there can be substantial optionality, making accurate processing probabilistic rather than absolute. Based on the input that they receive in a classroom setting, even more proficient participants should therefore not be able (yet) to process German case inflections accurately 100% of the time. A more fine-grained form of performance monitoring will therefore be needed for this experiment.

Fortunately, the position on explicit information and learnt knowledge has shifted a little over the years: VanPatten & Cadierno (1993) attach no importance to learnt knowledge and claim that Processing Instruction alone can affect the learner's developing system (VanPatten & Cadierno, 1993: 238). VanPatten & Oikkenon (1996) operationalise explicit information (EI) as a variable quite harshly when they provide the EI group with information on the target feature without giving them any relevant L2 input, effectively telling them how to ride a bike and then watching them crash when put to the test. In later studies, Fernández (2008), Culman, Henry & VanPatten (2009), and VanPatten & Borst (2012) allow for a facilitative effect of explicit information on acquisition, at least for some target features. Considering that German case morphology seems to be a special case that benefits from explicit information, the participants in this study will be provided with explicit information on the target feature and their inappropriate default processing strategies at the beginning of each instructional session. The ability to apply this explicit

information to the tasks could also be connected to the learner's Working Memory Capacity, as some have argued (VanPatten & Borst, 2012). Interestingly, Working Memory Capacity limitations are a popular explanation when the initial hypotheses of Processing Instruction researchers are not supported by their data (see the discussion of results in Angelovska & Benati, 2013).

As the interest of this study lies in exploring the impact of individual differences in Working Memory Capacity, there is no need for a comparison with other instructional treatments at this stage. Considering that previous research within the field of Processing Instruction has come up with no explanations for differing degrees of success for different learners, it seems warranted to change the approach. The main difference that has been observed in terms of participants' task performance after receiving Processing Instruction is that participants that score low on the pre-test are more likely to gain significantly after instruction than those participants who already score high on the pre-test (Lee & Benati with Aguilar Sánchez and McNulty, 2007; VanPatten et al., 2009). Or, as Lee & McNulty (2013, 70) state it: "The low scorers closed the gap suggesting that they benefited perhaps not more from PI than the high scorers but differently." Assuming that the differences in gains are not just due to statistics, i.e. the size of the gap between bottom and ceiling scores on highly controlled tasks, this raises the question as to how we can learn more about the differences in the quality of the learning process for the respective participant groups.

Instead of comparing groups of different age, gender, aptitude, etc. respectively, it seems then more promising to look at several learner variables and how they interact. A small-scale case study should keep the data manageable even when measuring several learner variables at once. Other context variables are likely to play a role as well. One must be aware of the fact that the cognitive load of a task does not only depend on the status of the language (L1 or L2), but also on the proficiency level of the participant and their exposure to relevant L2 input. What is perceived as a complex sentence therefore depends on more factors than just Working Memory Capacity and might vary from participant to participant. Another issue is that while Processing Instruction materials are indeed maximally reduced in complexity and should not favour the participants with higher Working Memory Capacity scores, the testing materials used in the pre- and post-tests do not necessarily share these features. It is therefore conceivable that Working Memory Capacity differences do not impact the on-task performance during Processing Instruction, but have an effect on the pre- and post-test scores nevertheless.

The attitude of the participants towards German and German native speakers is another variable that has not been explored in Processing Instruction research, although it seems to be an established factor, together with motivation: In his longitudinal study, Schlak (2004) employs an ethnographic approach and collects data through observations, interviews with learners as well as teaching staff, and the analysis of relevant materials such as textbooks, exams, syllabi, homework assignments, curricula, and student surveys. Considering the role of explicit grammar instruction in two different German courses, his study with 14 learners of Business German and 9 learners of Intermediate German at the University of Hawai'i reveals that grammar instruction is perceived as important by both students and teaching staff, but for different reasons: The professors (all German native speakers educated in the German Higher Education system) rate the grammar knowledge of their students as very low in both German and their native language, English (Schlak, 2004: 72). They stress the point that grammar instruction in German will improve the students' awareness of their first language as well. The teaching assistants, however, see a marked difference between first and second language acquisition and see grammar as a shortcut to language acquisition rather than as the driving factor behind it (Schlak, 2004: 70). The students are aware of the importance of grammar knowledge because both the general language course and the Business German course penalise lack of grammatical accuracy. Learning German grammar and being able to produce it correctly in writing is therefore essential for good grades (Schlak, 2004: 72–3), even though it is not the main focus in the Business German curriculum.

A similar picture emerges in the United Kingdom: students generally mention a lack of L1 grammar instruction. The foreign language classroom is often the first place where they encounter explicit grammar instruction and concepts such as word classes and grammatical functions. The university curricula are suspiciously devoid of grammar courses unless they belong to *ab initio* courses. Looking at the German BA programme that the University College London (www.ucl.ac.uk/selcs/prospective-students/undergraduate/german-ba) offers, it is possible to study a BA in German with only 1 course unit on Modern German in year one and half a course unit in year two, out of 4 course units per year (=12.5%). Considering that the assignments are mostly written and that the alternative options involve discussions of German literature and society, this lack of systematic language instruction is rather surprising. For the reverse case of a German student enrolling in an English BA programme at a German university (<http://www.uni-goettingen.de/en/517119.html>), the 3-year programme would require them to spend 29 Credits on Modern English Linguistics throughout their studies, out of 66 Credits (=44%)

- despite starting out with a higher proficiency level in English, compared to the proficiency level of their English counterparts learning German.

Conclusion

Processing Instruction is a pedagogic intervention that focuses on form by encouraging learners to change their inadequate default processing strategies. This is achieved through Structured Input Activities that are designed to bring the target feature to the learner's attention. Based on the assumption that second language learners are limited-capacity processors of information (VanPatten 1993, 1996, 2007), the input that learners receive is structured according to the Input Processing Principles that VanPatten established in his earlier research on learner's attention to form and meaning in the input (VanPatten, 1990). If this assumption is true, Processing Instruction should be equally beneficial for learners with high and low Working Memory Capacity spans, meaning that they should overall be equally successful in interpreting Structured Input.

Yet, not all learners perform equally well during and after Processing Instruction (Fernández, 2008; VanPatten & Borst, 2012; Lee & McNulty, 2013), so the question as to what causes performance differences still remains. From the discussion in section 2.4 it has become clear that, while individual differences research has been popular within the area of Processing Instruction recently, one of the most interesting variables, Working Memory, has not received a lot of attention yet. Previous research investigating Working Memory Capacity effects on the outcomes of Processing Instruction (Santamaría, 2007) has not been able to answer this question, but research on other learner variables, such as age, gender, first language, etc. has yielded results that suggest that these factors do not predict performance differences on Processing Instruction or traditional output tasks on their own at the group level.

This study attempts to address some of the issues of previous Processing Instruction research by investigating individual differences in Working Memory Capacity together with potentially mediating learner variables, consequently working with a broad definition of processing resources and employing several different memory-related measures involving different modalities. This will be discussed in more detail in the following two chapters. From the review of Processing Instruction findings, however, we can already formulate the following research questions for this project:

Research questions:

1. Do learners benefit from Processing Instruction in equal measures, regardless of their Working Memory Capacity?
2. Are there other learner variables that potentially mediate Working Memory Capacity and thereby affect the learning outcomes?
3. Are there participant-treatment interactions?

Based on the findings from previous studies on Processing Instruction, which have been presented at some length in this chapter, a hypothesis for each of these research questions can be made:

Hypotheses:

- a) Processing Instruction is maximally reduced in complexity, yet there usually are individual differences in the post-test scores. Even if differences in Working Memory Capacity will not result in performance differences during instruction, they will result in differences in post-test scores.
- b) There might be other variables that compensate for lower Working Memory Capacity scores. If any of the measured constructs have an impact on post-test scores, then these should show up in a comparison of the individual learner profiles.
- c) It is conceivable that different individuals react to and interact with the instructional and experimental materials in different ways. This could influence the learning and/or assessment outcomes. Evidence for this might be found in the participants' approach to the less controlled tasks, in the way participants work with the materials, and in the observational data the researcher collects.

3 WORKING MEMORY: DEFINITION AND OPERATIONALISATION

One of the key assumptions in Processing Instruction research is that learners of a Second Language are limited-capacity processors of information (VanPatten, 2007: 116, 2005: 268) who process input according to particular default processing principles which might be suboptimal for the processing of the target language. Learners can, however, change their processing strategies if they experience the failure of their default strategies. Processing Instruction is a pedagogical intervention that provides learners with Structured Input that makes processing of grammatical forms in the input necessary if the learners are to satisfy their communicative needs. To this end, the input the learners receive contains no additional pragmatic (or other) information that learners could rely on. Materials do not offer any alternative cues as they are controlled for plausibility, animacy, event probabilities, and other context factors which might increase the likelihood for success of default or shallow processing. This means that the input is maximally reduced in complexity and the tasks are unambiguous, i.e. there is only one correct answer to any interpretation task. Proponents of the Input Processing Theory (VanPatten, 2007, 2004, 1996; Benati, 2013, Lee & Benati, 2007) have therefore claimed that Processing Instruction poses only a minimal strain on the individual's processing resources, making the creation or reinforcement of form-meaning connections more likely. If this is correct, then learners of different Working Memory Capacity spans should perform equally well on Structured Input interpretation tasks. The motivation for this study is to explore whether this claim is warranted.

Working Memory Capacity limitations are also the starting point for VanPatten's (2007, 2002, 1990) Input Processing Principles. Without these constraints, learners would not have to prioritise certain aspects of the input during processing, but would be able to attend to all features. What role, then, does Working Memory play in language processing? Working Memory can be seen as both the process and the product of keeping task-relevant information in a state of heightened activation. The construct of Working Memory has been conceptualised as the centre of attention (e.g. Baars, 1998; Kane et al., 2007), the accessible part of Long-Term Memory (e.g. Cowan, 1993, 2001, 2005), or a more sophisticated construct that is part of a complex system, oftentimes divided into multiple components (e.g. Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974; Baddeley, 1986, 2000). Different models of Working Memory account for a range of behavioural evidence, but there are still some rather elementary questions in relation to which no definite consensus exists, e.g. whether Working Memory is a single- or a multi-

component system, what the nature of Working Memory Capacity limitations is precisely, and how (individual, across-task, developmental) differences in Working Memory Capacity come to exist – to name just a few.

Within the field of Second Language Acquisition research, Working Memory has lately been seen as part of a multi-component aptitude complex (Skehan, 1991; Dörnyei & Skehan, 2003). These components not only interact with the learning environment, but also influence each other (Snow, 1991). From this new conceptualisation as well as from the description of the MOGUL framework, it has become clear that the research design for this study will have to diverge from what is customary in the field of Processing Instruction research because the primary variable of interest, Working Memory Capacity, cannot be conceptualised as a static and isolated construct. Instead, the learner's processing resources must be investigated as the outcome of a complex interaction of variables and with respect to the learning and assessment contexts. This has consequences for the measures used in the experiment, which will be described in this chapter and the next. First, the concept of Working Memory will be introduced and seminal Working Memory models as well as the most pertinent questions regarding individual differences in Working Memory Capacity will be presented. Then a suitable definition of Working Memory will be adopted and it will be pointed out how this construct is relevant for Processing Instruction research. The chapter will end with the justification of the operationalisation and measurement of Working Memory and Phonological Short-Term Memory Capacity in this study.

3.1 Working Memory

Working Memory is associated with a number of higher-level cognitive tasks, such as language processing, general learning, and reasoning. It is, broadly speaking, a place in which task-relevant information is kept in a state of heightened activation. This makes the information readily accessible and enables the individual to form links between existing and new knowledge, for example. Verbal Working Memory has been hypothesised to have evolved specifically for language acquisition (Szmalec, et al., 2013). It is therefore not surprising that this construct has received considerable attention within the fields of Second Language Acquisition research and psychology. Yet, there is no unified theory of Working Memory, just as there is no unified theory of cognition at this point. It is therefore impossible to give an exhaustive overview of existing models and measures of Working

Memory Capacity. Instead, the two most influential models of recent times will be presented and similarities as well as differences will be highlighted.

The most widely cited Working Memory model in the field of Second Language Acquisition research is that of Baddeley & Hitch (1974), later extended by Baddeley (1996, 1998, 2000, 2002). It consists of different modules that cater for different input modalities and has therefore a certain appeal for the exploration of language processing that focuses on input modalities and/or specific learning impairments (Melby-Lervåg & Lervåg, 2012). Cowan's Embedded-Processes model will be presented as an alternative (though not entirely contradictory) model of Working Memory – one which has received support from neuro-imaging findings (Chein, et al., 2003). This study has mostly adopted the perspective of the Embedded-Processes model as this interpretation of Working Memory is highly compatible with the understanding of input processing proposed by Sharwood Smith & Truscott (2014) for the MOGUL framework. It should be noted, however, that the modality-specific storage of representations that is built into the MOGUL architecture is more in line with Baddeley's multi-component model of Working Memory.

While Baddeley's modular model and Cowan's Embedded-Processes model focus on what is in the centre of attention, work in Randall Engle's Attention & Working Memory Lab has investigated the role that executive attention plays in the context of individual differences in Working Memory Capacity. Their understanding of capacity differences is less about the focus of attention than about the suppression of interfering stimuli while keeping goal-relevant information activated (Engle, 2002). This is one approach that could be integrated in the MOGUL framework, but cannot be discussed in detail here. The other model that has received a lot of attention in Second Language Acquisition research and must be mentioned is the Resource-Sharing model of Just & Carpenter (1992). This model is concerned with the processes involved in what is termed the executive module in Baddeley's model and states that there is a trade-off between storage and processing in terms of Working Memory resources and that the higher the processing demand, the smaller the storage capacity (and vice versa):

Supporting this dissociation between processing span and processing speed, Just and Carpenter found that subjects who were high vs. low in reading span did not differ in reading speed, although they did differ in the ability to keep in mind simultaneously more than one potential interpretation of a sentence structure.

Cowan (1995: 91)

This model has been very popular in Second Language Acquisition studies, especially in connection with ambiguity resolution as is necessary in the interpretation of garden path sentences. While the processing/ storage distinction is still a crucial aspect of Working Memory Capacity tasks, the model as such has been deemed simplistic (Lustig, May & Hasher 2001; MacDonald & Christiansen 2002; Baddeley 2000).

3.2 Working Memory Models

Even after decades of systematic empirical testing, researchers stress the lack of consensus when it comes to the exact nature of and processes involved in Working Memory (Baddeley & Hitch, 1974; Cowan, 1995, 1999, 2005; Altarriba & Isurin, 2013). Whereas previous models of Short-Term Memory were mostly conceptualised as a single system, Baddeley & Hitch argue in their seminal (1974) article that such a unitary system, made of uniform processes, would be more vulnerable to interferences than the behavioural data accounts for (Baddeley & Hitch, 1974: 50). Consequently, they propose multiple, modulated stores for visual/spatial and verbal (i.e. phonological) Working Memory, each working separately. They also suggest that there is a central storage system that is limited in capacity, based on their findings that a series of digits only interferes with a verbal task when the number of items to be recalled is challenging, but not when it comprises two items only (1974: 53). Moreover, the interference effects of the memory load created prior to a reasoning task depend on the instructions that the participants receive: If they are told to prioritise memory recall, their performance on the reasoning task deteriorates more significantly than when they are told to give priority to reasoning, which comes at the expense of their memory performance (1974: 53).

Under a concurrent-load condition, participants are told to rehearse items under four conditions while performing the verbal reasoning task: 1) silently, 2) while repeatedly articulating the word “the”, 3) while counting aloud from one to six, and 4) while repeating aloud a random six-digit sequence that changes after each reasoning trial. Compared to the control condition, the articulatory suppression and the counting just fail to reach significance, but the random sequence of digits significantly slows down the reasoning task performance, especially for the more difficult verbal reasoning problems. This is taken to mean that Working Memory limitations are not due to articulatory capacity limits alone, but that there is a component of Working Memory that is limited in capacity. Subsequent experiments involving free recall tasks and verbal comprehension suggest

that it is the phonemic similarity of items that impairs performance, which leads Baddeley and Hitch to the conceptualisation of the phonological loop, a “phonemic response buffer which is able to store a limited amount if speechlike material in the appropriate serial order” (Baddeley & Hitch, 1974: 77). While the capacity of this component is limited, the processing component can extend the functional capacity by applying rehearsal routines or by facilitating storage in Long-Term Memory. Based on findings from related research, the authors assume the existence of a similar modality-specific buffer system for visual Working Memory.

Baddeley has tested and further developed this well-known multi-component model of Working Memory and its slave systems, the phonological loop and the visuo-spatial sketchpad, extensively over the past decades. “He also removed the central store for the sake of parsimony, though he later added one back, called an episodic buffer” (Ricker, AuBuchon, & Cowan, 2010: 575). The processing in this model is attributed to the attention-demanding central executive, making the phonological and visuo-spatial stores attention-free:

(b) Baddeley (1986, 2000)

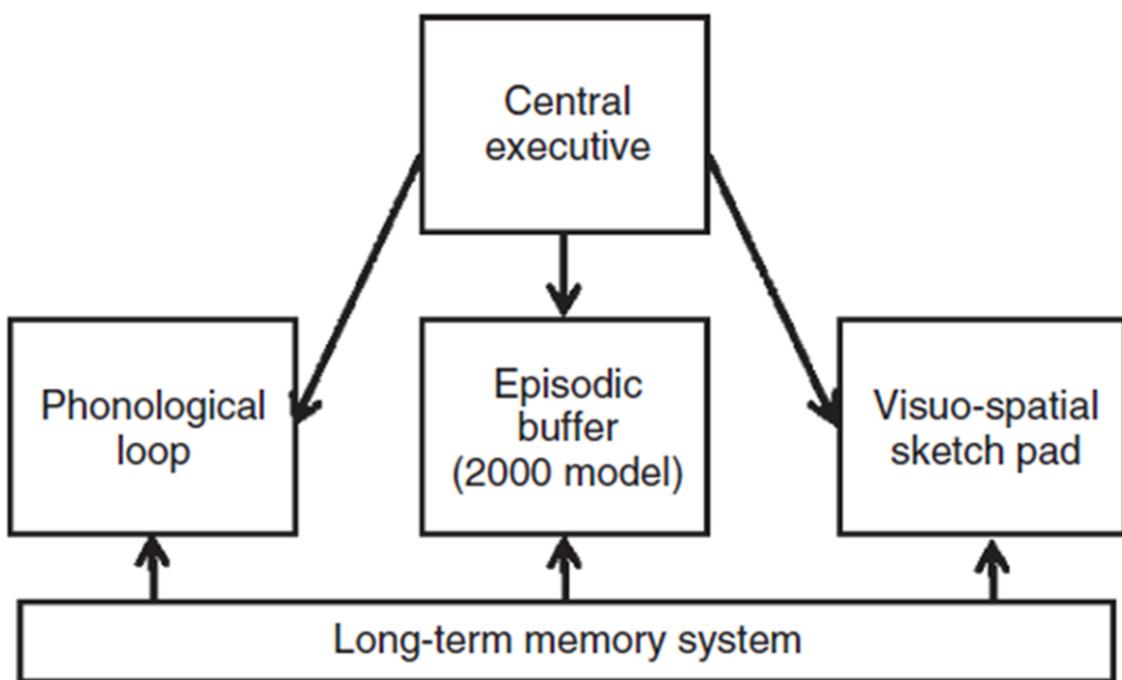


Figure 3.1: Baddeley's Working Memory model after Ricker, AuBuchon & Cowan (2010: 574)

Cowan, on the other hand, defines Working Memory as “cognitive processes that are maintained in an unusually accessible state” (Cowan, 1999: 62). He disagrees with the differentiation between phonological and visuo-spatial information on the grounds that there would have to be separate information stores for other sensory input as well, e.g. a specific memory for touch. Sharwood Smith & Truscott (2014: 60–1) show how this conflict can be resolved and offer a parsimonious solution in that the specialised sensory input processing modules each comprise a processor and a store, the activated part of which is the module’s Working Memory. Some of these stores’ contents never reach the level of awareness or only leave short ‘afterimages’, whereas other outputs from sensory modules can serve as input for the language modules and then have a longer shelf-life in this abstract form. Cowan’s understanding is domain-general, meaning that Working Memory is the activated and accessible part of a general Long-Term Memory that is virtually limitless. The state of heightened activation is subject to decay over time as well as interference from competing stimuli. To keep information activated and accessible, maintenance mechanisms must be employed, e.g. sub-vocal rehearsal.

But not all information that is available for manipulation within Working Memory is also conscious: Only a number of chunks can be held in the focus of attention at any one time. Miller (1956) reports that the number of items that can be recalled is about seven, subject to participant and type of item, with the possibility of increasing that number by chunking information or using rehearsal strategies. Cowan (2001) though states in his review of Working Memory Capacity research that, when chunking and rehearsal are prevented, the number of recalled items is usually about 3–5, which also happens to be the average span size that is elicited in Working Memory span tasks. While the amount of chunks seems to be relatively stable, the chunk size and complexity can vary significantly. Working Memory is thus restricted in terms of the amount of activated information, which is time-limited and subject to interference, and the number of items that can be held in the focus of attention, which is capacity-limited:

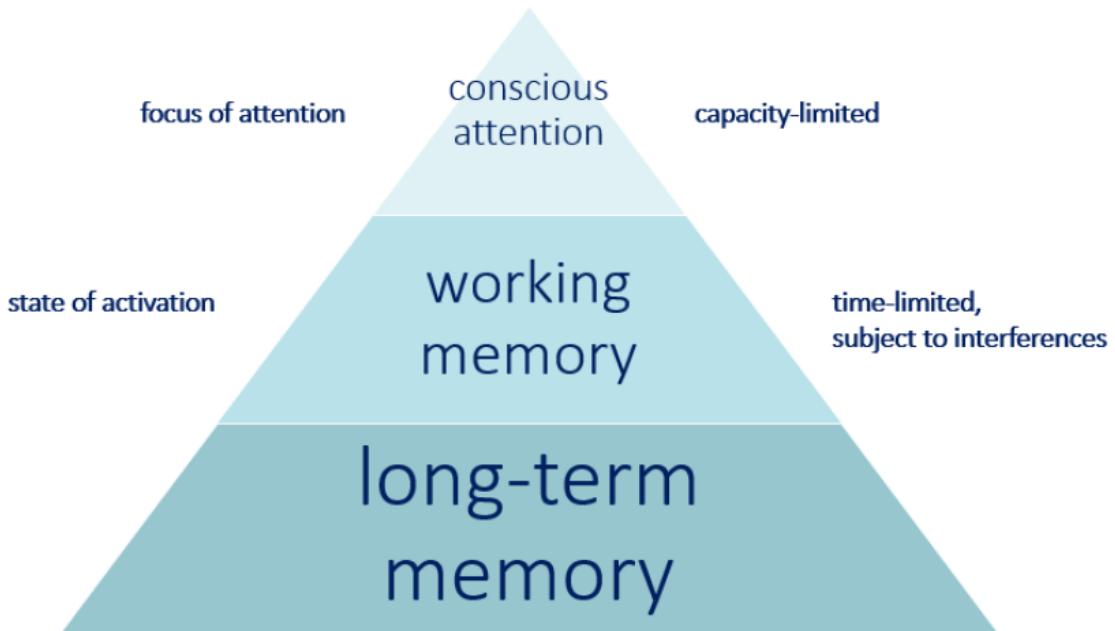


Figure 3.2: The Embedded Processes Model of Working Memory (after Cowan, 1999)

While studies on dual-task performance support the idea of separate memory stores for different modalities (Cocchini, et al., 2002), studies investigating latent variables have interpreted the evidence as resulting from a more domain-specific Short-Term Memory system that masks the domain-generality of Working Memory (Kane, et al., 2004). It is important to note though that the domain-generality that some researchers propose for the Working Memory store does not extend to the processors. The processed input still maintains its modality-specific attributes and strategies are modality-specific as well, e.g. sub-vocal rehearsal strategies only apply to auditory input. However, there seem to be such significant across-task and within-subject variations that a re-analysis of previous Working Memory studies with a more precise interpretation of task demands and contexts seems in order (for a meta-analysis, see e.g. Simmering & Perone, 2013). The improved temporal and spatial resolutions of neuro-imaging techniques might help in future research.

The Embedded-Processes model seems fundamentally different to Baddeley's multi-component model. Baddeley, however, argues that the Embedded-Processes Theory is dealing with one part of his model, i.e. the interaction between the central executive and the episodic buffer. It is true that both models are examples of a non-unitary approach to conceptualising Working Memory. While Baddeley separates between different domains and modalities, Cowan differentiates between states of activation and also introduces

focal attention. Both models assume a domain-general Working Memory store and both models account for behavioural evidence gathered in a great number of studies.

While the modality specificity of Baddeley's model has been attractive to researchers focusing on individual differences (Szmalec, et al., 2013), Cowan's model has sparked research on the domain-general processes involved in Working Memory. Both approaches have been fruitful. New evidence now comes from neuro-imaging studies that allow a more differentiated picture of the processes that are involved in cognitive tasks. With these new techniques, the potential for the empirical testing of Working Memory models has reached a new stage. Recent studies using functional magnetic resonance imagining have produced results that can supplement findings from behavioural studies with a new layer of data. They can also complement findings from research paradigms involving Event-Related Potentials, thanks to their higher spatial resolution. It seems that Working Memory tasks induce an activity of the same regions of the brain as are activated during perceptual processing, making a case for the notion that Working Memory is in fact the activated part of Long-Term Memory. These findings also provide support for the importance that Sharwood Smith & Truscott (2014) attach to the perceptual modules.

Baddeley & Hitch's (1974) model has enjoyed substantial popularity because it is modular and quantifiable, thus facilitating empirical testing. To that end, the authors themselves have proposed a collection of measures that are sophisticated, encouraging a systematic investigation of their model and providing excellent raw data for statistical analyses. The model's popularity, however, is also due to the theoretical paradigm of the time being one which favoured modularity. These days, researchers looking into Dynamic Systems Theory and emergentism embrace the complexity that has always been present as background noise, but previously needed to be filtered out. Spencer et al. (2012), for example, propose a 'soft assembly' approach. Processing resources are then the result of multiple interacting components that combine freely from moment to moment, based on the context and the nature of the task as well as the developmental history of the system. A resource pool like that should not display a static capacity limit, but one that depends on the circumstances. The MOGUL framework fits this notion in the sense that all processing is influenced by previous processing and learning, i.e. changes in the stores of the individual modules and the co-indexations across modules, is a consequence of processing. Depending on the available input, different routes can lead to similar processing outcomes, based on the relevant processors' outputs.

3.3 Individual Differences in Working Memory Capacity

As seen above, there is no doubt that Working Memory is limited in its capacity. There is only a “relatively small amount of information that one can hold in mind, attend to, or, technically speaking, maintain in a rapidly accessible state, at one time” (Cowan, 2005: 1). The question is rather what this limitation results from and why it exists. This relates to the question as to how the nature of Working Memory differs from that of Long-Term Memory, which is a lot more extensive. To a certain extent, Working Memory Capacity limitations seem pretty constant. Usually, 3–5 chunks (i.e. meaningful units of information) are recalled under the conditions of traditional Working Memory tasks. When rehearsal and chunking are controlled for, experiments show an even more limited capacity of more or less 3 chunks (Chen & Cowan, 2009). Opinions vary as to why this limit exists. It seems obvious that the human mind can only attend to so much information at any given time, giving preference to some information over other. Yet the number of items to be held in mind at once could be higher or lower. The functional answer to this question seems to be that this number of chunks is the optimal number in terms of information processing: Mathematical models have shown that searching for an item within a group of items is most efficient when the group size is similar to the average individual’s Working Memory Capacity (Cowan, 2005: 168). From an evolutionist perspective, a lower number might not have been sufficient for survival while a higher number might have been too costly to maintain.

In addition to the capacity limitation of the Working Memory store, it seems that there are time constraints as well. Especially for the modality-specific items in Working Memory, activation periods seem to be limited. Again, there are different theories regarding the nature of this limitation. The two main explanations are decay over time and interference. If the contents of Working Memory consist of activated items in Long-Term Memory, then there must be processes which determine what stays activated and what falls back below the activation threshold required for consciousness. If it’s due to the passage of time, strategies for rehearsal or refreshment can keep items accessible by maintaining the activation level. On the other hand, it is quite clear from extensive research on executive attention (e.g. Kane & Engle, 2002) that there is interference involved. Depending on how much processing and distractor tasks overlap and whether the concurrent processing tasks rely on the same input modalities, the extent of the interference effects varies.

From a MOGUL perspective, there would be competition within the respective modules, but also between the modules, leading to different interference effects at different levels

and stages of the processing (Sharwood Smith & Truscott, 2014: 66–7). Competition within a module, i.e. input modality, would affect the processing at an earlier stage and potentially have a heavier impact on the success or failure in establishing a consistent message than competition between the outputs of different modules that work in parallel. Working Memory research that focuses on individual differences is primarily concerned with the cause of these individual differences. The two main hypotheses are that differences in Working Memory Capacity either arise from a) different total capacities of the (storage) resources or b) differences in the efficiency with which individuals make use of their Working Memory resources (functional capacity) (Cowan, 2005).

Ever since Daneman & Carpenter (1980) demonstrated that verbal Working Memory span measures can predict comprehension, this question has been in the focus of Working Memory research. Linked to it is the question as to why a complex span measure, such as the Reading Span task, can predict performance on a wide range of cognitive tasks quite effectively. Despite these correlations, Daneman & Carpenter (1980) have received considerable critique, especially in terms of the reliability and validity of the measure. Yet, modified versions of the original Reading Span task remain a mainstream measure of Working Memory Capacity – especially in Second Language Acquisition research. Those who subscribe to the concept of Working Memory Capacity as an attentional resource focus on the nature of the tasks that are involved in span measures. Processes such as task-switching are widely believed to drain attentional resources, while gaps between the processing operations can be used to perform maintenance or rehearsal strategies. Engle and colleagues in the Human Working Memory Lab at Princeton University have done extensive research on the importance of inhibitory processes, arguing that individual differences are due to the individual's ability to inhibit interfering processes (Engle, 2002). Other factors concern the individual's ability to filter out irrelevant information or the capacity to store information. While there is a difference to be made between the activated items themselves and the processes that put them in an activated state as well as maintain them there, capacity is not a static product, but an on-going process.

There has been some debate on whether training can improve Working Memory Capacity. “The most dramatic example may be the case studies of improved memory span beginning with Ericsson et al. (1980). Over the course of a year, they trained an individual to increase his digit span from seven items to about 80” (Cowan, 2009: 14). The study mentioned above shows that it is possible to train a specific Working Memory task and improve performance on that task, e.g. by learning to chunk the items in a Digit

Span task. While it is then possible to increase the number of recalled items, this increased recall capacity is not transferable to other tasks that are supposed to tap into Working Memory resources, but affect different modalities, such as a Symmetry Span task (e.g. Heitz & Engle, 2007). For a meta-analysis of Working Memory training studies, see Melby-Lervåg & Hulme, 2013). It is conceivable though that, in terms of evolution, Working Memory Capacity has expanded over time, adapting to the increasingly complex stimuli of the environment (Cowan, 2009: 15). It has also been observed that bilinguals have higher Working Memory Capacities when measured with complex span tasks (Bartolotti & Marian, 2013). Whether this is due to higher processing efficiency or better attentional control, however, is not clear yet.

3.4 Working Memory and Second Language Acquisition

One problem that is often raised with regards to research in Second Language Acquisition and Working Memory is that models are interpreted rather narrowly, according to the task at hand. Depending on research trends and agendas, the concept of Working Memory has been operationalised rather inconsistently in the field of Second Language Acquisition research: From the rather narrow focus on Baddeley's phonological loop and its role in non-word repetition performance of dyslexic participants (Melby-Lervåg & Lervåg, 2012) to the much broader trade-off between storage and processing aspects of Working Memory in sentence interpretation (Juffs & Harrington, 1996). Martin & Ellis (2012), on the other hand, use a variety of measures to find out whether Phonological Short-Term Memory and Working Memory are correlated with vocabulary and grammar learning of an artificial language. They observe that Phonological Short-Term Memory and Working Memory seem to affect the outcomes of instruction (both in terms of vocabulary and grammar) as independent factors, both contributing significantly to the detection, generalisation, and application of rules in that language (Martin & Ellis, 2012: 396–97). Looking back at how Working Memory is defined within the MOGUL framework, the concentration on specific modalities is not problematic as each dedicated module consists of a processor and a store, the activated and accessible part of which is considered to be its Working Memory. It is therefore warranted to look at individual differences at the level of the input modalities involved in the processing within the context of a specific task.

More problematic is the fact that vastly different scoring procedures are in use, which makes a comparison of Working Memory studies almost impossible. Together with the fact that the Working Memory Capacity range is narrow to start with and that Second Language Acquisition studies often recruit university students as participants, a median split between a high-span and a low-span group does not generate much statistical power. More often than not, individual differences in language processing and learning trajectories are attributed to differences in Working Memory Capacity without any explicit reference to a Working Memory model and/or any measure of Working Memory Capacity (Malovrh, 2014).

On the other hand, efforts have been made to come to a systematic understanding of Working Memory effects on language learning. It has been argued (e.g. Wesche, 1981; Skehan, 1989) that there are two types of language learners: those relying on memory and those relying on analytical abilities. This distinction has recently received support from neurolinguistics, and not just for L2 learners, but also for monolinguals (Tanner, 2013; Tanner, et al., 2014; Tanner & Van Hell, 2014). Clearly, both types of processing can lead to advanced proficiency levels. Looking at this distinction from a MOGUL perspective, it is conceivable that in some situations and for some individuals (depending on the task, the available information, and the processing history), different processing routes can lead to similar results. If these processing contexts and conditions are repeated on a regular basis (i.e. in a classroom or an immersive context), then these processing routes will have advantages over the alternative routes. Considering the number of different processors which can generate parallel representations, the make-up (read: processing history) of the respective modules would then allow for a multitude of processing preferences or aptitude patterns, resulting in a continuum of foreign language processing aptitude. To include a measure that reflects these learning types, the Working Memory Capacity measures used in this study will be supplemented with measures of language analytical ability.

Following Dörnyei & Skehan's (2003) understanding of language learning aptitude, presented in Table 3.1 below, Working Memory and components of the aptitude complex are involved at different stages of the process. With regards to MOGUL, these stages are not necessarily serial, but have considerable temporal overlap as different processors work with their respective input modalities and features in parallel whenever possible (Sharwood Smith & Truscott, 2014: 78–82). And because every module consists of a dedicated processor and a store, it is clear that Working Memory, i.e. those representations which are activated and available for processing, affects all parts of input

processing, but potentially to different extents. For example, what Dörnyei & Skehan (2003) term *noticing* refers to the representation that is dominant in the perceptual output structures (Sharwood Smith & Truscott, 2014: 288–90). The phonemic coding ability is then the work of the auditory processor and the Working Memory component consists of the activated representations in the respective store – and ideally the co-indexed representations in the other modules. It should be noted, however, that within the MOGUL framework, notions such as attentional or pattern control are interpreted as emergences of the system rather than conscious processes directed by the learner (Sharwood Smith & Truscott, 2014: 287–300). Following the MOGUL approach, individual differences in Working Memory can be interpreted as affecting different task and input modalities in different ways.

SLA stage	Corresponding aptitude constructs
Input processing strategies, such as segmentation	<i>Attentional control</i> <i>Working Memory</i>
Noticing	Phonemic coding ability <i>Working Memory</i>
Pattern identification	Phonemic coding ability <i>Working Memory</i> Grammatical sensitivity Inductive language learning ability
Pattern restructuring and manipulation	Grammatical sensitivity Inductive language learning ability
Pattern control	Automatization Integrative memory
Pattern integration	<i>Chunking</i> <i>Retrieval memory</i>

Table 3.1: Second Language Acquisition stages according to Dörnyei & Skehan (2003: 597)

If we consider the involvement of Working Memory in the processing of Structured Input, as available in the context of Processing Instruction, we can think about the extent to which individual differences in Working Memory Capacity might impact the instructional outcomes of this intervention:

Attentional control and/or Working Memory are supposed to play a role in the learner's selection of *input processing strategies*. As Processing Instruction manipulates input in a way that forces learners to adopt optimal input processing strategies by letting them experience the failure of their sub-optimal default strategies, it could be argued that the effects of Working Memory Capacity differences should be minimal at this stage. Learners do not have a strategy choice and are trained to apply a pre-selected strategy to the input they receive.

Together with phonetic coding ability, Working Memory is also hypothesised to affect *noticing*. Processing Instruction does not address this aspect of the language learning process explicitly as the grammar acquisition supposedly occurs without conscious awareness of the learner (Benati & Lee, 2009: 89). Processing Instruction attempts to increase learners' intake by manipulating the input in a way that makes the target feature more salient and the formation of stable form-meaning connections more likely. The input is not ambiguous and the learners have to rely exclusively on the processing of grammatical form to derive meaning.

Another stage at which Working Memory (together with phonetic coding ability, grammatical sensitivity, and inductive learning ability) influences the process is *pattern identification*. Processing Instruction provides help for this as well: the input the learner receives is maximally reduced in complexity and the target feature is made meaningful, salient, and unambiguous. The increased salience in the auditory input should improve the phonetic coding and the nature of the Structured Input Activities should help in keeping the strain on Working Memory resources manageable. So potential sources of individual differences at this stage might be due to differences in grammatical sensitivity and/or inductive learning ability. Measures of these two constructs will be used in the experiment.

Grammatical sensitivity and inductive language learning ability are also mentioned in the context of *pattern restructuring and manipulation*, which is the stage at which patterns derived from the input are integrated in the learner's interlanguage. There is no way of testing the effects of instruction on an abstract construct such as interlanguage, but following Dörnyei & Skehan's (2003) reasoning and applying it to Processing Instruction, the manipulated input should help with the inference of (combinatorial) 'rules' from regularities in the input as more instances of the target feature are presented in a way that highlights form-meaning connections.

The next stage of the process would then be *pattern control*. At this stage, automatisation and integrative memory are supposed to be the main factors for success. Differences at this stage might be elicited through measures such as trials-to-criterion or the monitoring of on-task performance. If one counts how many times a learner produces an incorrect form before they consistently produce the correct form, then it should be possible to compare the differences in learning rate across learners. In the case of Processing Instruction, no output is being produced, so a process-oriented measure would have to count the number of wrongly interpreted items until the form is consistently interpreted correctly.

Lastly, *pattern integration* is argued to be linked to chunking and retrieval memory. According to VanPatten, the processing strain posed on the learner is minimal in Processing Instruction, so there should be no extensive chunking necessary. One could also argue that the retrieval is facilitated because Structured Input Activities are always meaningful and affective Structured Input Activities lead to affective as well as cognitive engagement of the learner, which could facilitate later retrieval as the co-indexation of novel representations with affective states supposedly adds a boost to the initial resting activation levels.

3.5 Operationalisation of Working Memory and Phonological Short-Term Memory Capacity

As previously seen, Working Memory Capacity limitations can be elicited via dual tasks that tap the same processing components at the same time while suppressing the rehearsal of information, commitment to Long-Term Memory, and strategies such as mnemonics or chunking (see Figure 3.3 below). There is a multitude of conscious strategies that individuals use to keep information activated, such as repeating a telephone number over and over when dialling, creating memory hooks, etc. Working Memory Capacity tasks ensure that there is no time and/or no opportunity to apply these conscious strategies. But even then Working Memory Capacity does not appear to be a fixed construct that holds true for the same individuals performing different tasks or the same task on different days (e.g. Simmering & Perone, 2013). Individual Differences and cognitive styles as well as the context of the task at hand and the interactions between abilities and context characteristics are hypothesised to play a role in determining the

extent to which domain-general or -specific processes are involved in Working Memory performance (Conway, et al., 2005: 770).

Measuring WMC

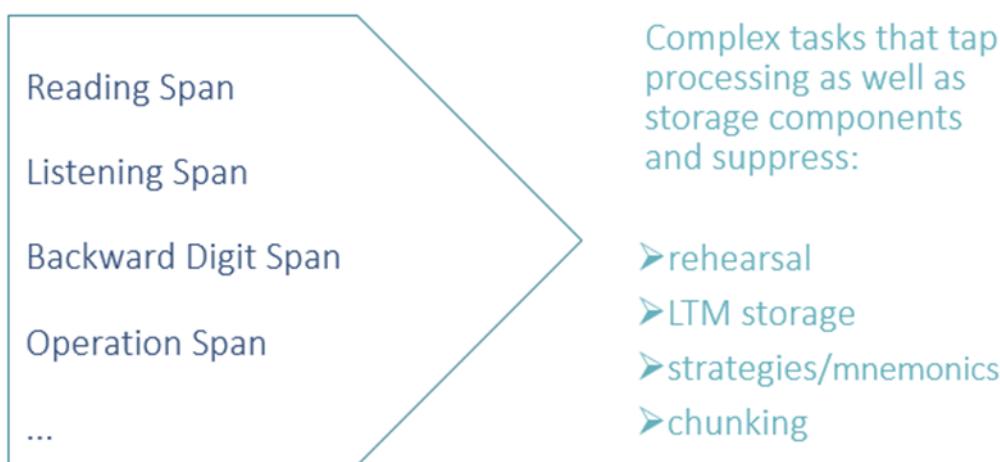


Figure 3.3: The shared characteristics of Working Memory Capacity measures

The task-specific view, as promoted by Daneman & Carpenter (1980), claims that the characteristics of the Working Memory Capacity task are important, but Turner & Engle (1989) have shown that Working Memory Capacity tasks involving no reading component can still predict reading ability. Recent neurocognition studies have indicated activation of the pre-frontal cortex during memory span tests, an area which is generally linked to the ability to control attention and conscious thought. It is unclear, however, whether Working Memory Capacity is the cause or the effect of these patterns and whether the limitations differ between individuals and are trait-like, or intra-individual and context-dependent, or a combination of both. The validity of the construct as a whole can be demonstrated, however, via the correlations with other complex span measures and by the lack of correlations with measures that do not require stimulus suppression or sustained attention (Conway, et al., 2005: 778). In order to mediate between the two separate stances, a Reading Span task is used in the present study. From a task-specific perspective, this should correlate well with the reading comprehension of the visually presented Processing Instruction materials and additionally reflect the verbal abilities of the participants. From a domain-general perspective, the nature of the complex Reading Span measure assures that competing tasks tax the participants' processing resources

either in terms of the allocation of executive attention, the suppression of interfering stimuli, or the switching of attention between tasks.

In addition to that, an Auditory Digit Span was chosen based on the following arguments:

- a) Phonological Short-Term Memory (PSTM) is another memory function whose individual differences are linked to variability in language learning success (Gathercole & Baddeley, 1993; Pimsleur, Sundland & McIntyre, 1966). It has been argued that it might be particularly important for the acquisition of function words and bound morphemes, as these cannot be readily associated with tangible imagery in the contextual stores (DeKeyser, 2005). While this study is not about the initial acquisition of morphemes, one could expect that those participants with a larger Phonological Short-Term Memory span should have acquired the case markers more easily and potentially formed more stable form-meaning connections than those whose span was more severely restricted.
- b) Phonological Short-Term Memory is supposed to measure the storage capacity of Working Memory under conditions where there is no added processing load or competing distractor task. This measure is therefore interesting as, theoretically, the aural Structured Input in Processing Instruction should tax the storage capacity more than the processing capacity – differences in Phonological Short-Term Memory might then be more influential for the performance on tasks with auditory input than differences in Working Memory Capacity.
- c) The phonological loop, as described by Baddeley & Hitch (1974) and elsewhere, seems to play a role in language learning, especially in terms of fluency and aural comprehension (Ellis & Sinclair, 1996). The modular Working Memory model has been criticised previously, but the observation that the input modality seems to play a role in eliciting memory effects remains. From a MOGUL perspective, the auditory processing module has its own store, in which the activated representations make up the module's Working Memory. It is conceivable that Phonological Short-Term Memory as a construct is a description of this module's characteristics and plays a role in the efficiency with which auditory input is processed. Many other tasks that appear as Working Memory tasks in the literature, such as the n-back task, have stronger correlations with Short-Term Memory than with Working Memory.

The common denominator of all traditional Working Memory tasks is that task-relevant information must be kept in an active and accessible state while a demanding secondary task interferes with rehearsal. To this end, Working Memory Capacity tests must be designed and administered in such a way that participants have to pay attention to both the storage and the processing aspects of the task and have no time to employ rehearsal strategies. Moreover, the challenge must be hard enough to elicit Individual Differences in Working Memory Capacity. Especially in college and university populations consisting of young, healthy adults, this is important in order to avoid ceiling effects (Conway, et al., 2005: 773). New and more dynamic Working Memory tasks are being devised all the time. However, they either lack the methodological validation of the older tasks, are not correlated strongly with other Working Memory Capacity measures, or do not demonstrate any added value (for a discussion, see Kane, et al., 2004). While traditional measures, such as the Reading Span task, do not exclusively measure Working Memory Capacity, but also other constructs (e.g. verbal ability), they can still demonstrate adequate reliability and validity scores.

As a measure of Working Memory Capacity, an automated Reading Span test is used in this study. The Reading Span test is a traditional complex span task, meaning that the presentation of a sequence of letters is interspersed with a distractor task, in this case a sentence judgement task where participants need to judge whether a sentence makes sense, or not. In this version, the test items are presented visually. In addition to remembering relevant information, it is also crucial to forget irrelevant information from previous trials that could interfere with the current trial. As a measure of Phonological Short-Term Memory, an automated Digit Span test is used in this study. The Digit Span test in its forward condition is a simple span measure that has been linked to Phonological Short-Term Memory, Short-Term Verbal Memory (Woods, et al., 2011), and in Baddeley & Hitch's (1974) model, to the capacity of the phonological loop. Participants listen to a sequence of digits and then recall them after a prompt. They are not required to manipulate the input and there is no distractor task. In the backward condition, participants listen to a sequence of digits and are prompted to recall them in the reverse order of presentation. While they do not have to suppress additional interfering stimuli, they do have to perform a mental transformation, i.e. a secondary processing task. Reversing the sequence of digits usually results in a lower span score as the test takers need to manipulate the input that they are holding in a state of activation before they can produce any output. In addition to that, as in the forward condition, one can expect interference effects of previously recalled numbers. The Digit Span test as a whole is

widely used in intelligence testing (Redick, et al., 2012: 164) and has good reliability and validity scores. It is not quite clear yet whether the backward condition taps the same construct as complex span measures. The task was chosen in order to include a simple span measure with an aural component, which “is thought to more directly assess the core systems of verbal working memory” (Woods, et al., 2011: 110). Employing a simple span task makes sense for the following reason: Working Memory span tasks are designed to suppress rehearsal/maintenance mechanisms and conscious strategies, thus eliciting the Working Memory Capacity limits described in length by Cowan (2005). But this is a very narrow definition of Working Memory Capacity which only covers a small part of the general processing resources. The simple span task is not designed to suppress the strategies and/or mechanisms mentioned above, meaning that the resulting score may well be the product of the interaction between different processing resources.

For both tests, automated versions were chosen in order to administer the tests to individual participants and to account for differences in reading times, making sure that quicker readers would not have time for rehearsal (Conway, et al., 2005: 771). In the case of the auditory Digit Span task, this also “reduced variability in the rate and clarity of digit presentation” (Woods, et al., 2011: 109) as the input is not dependent on the researcher’s presentation. Another forte of a standardised procedure for data collection and scoring is that it can be compared across samples and yields good validity and reliability scores. Using computer scripts for the testing also has the advantage that stimulus presentation can be branched, meaning that each span size to be tested is based on the participant’s previous performance. The Reading Span test also calculates an average response time for the practice trials and times out if participants take too long in the test trials. This is important because, according to Friedman & Miyake (2004), time-constraint complex span measures have a higher correlation with higher-order cognitive tasks than those measures which do not limit decision times.

The script for the automated Auditory Digit Span test is based on experiments reported by Woods et al. (2011) and was chosen because of its non-traditional scoring procedure “that showed reduced variance, improved test-retest reliability, and higher correlations with the results of other neuropsychological test results when compared to traditional DS measures” (Woods, et al., 2011: 101). The script for the automated Reading Span measure is based on materials copyrighted by Engle (2005). They have been used in previous research and are validated. The test-retest reliability of the automated Reading Span test is .76 for the absolute scoring technique and .82 for the partial-credit scoring (Redick, et al., 2012: 166). The internal consistency of Reading Span measures varies,

but Cronbach's α for the partial-credit scoring method is usually reported as being well above .70 (Engle, Tuholski, Laughlin, & Conway, 1999: $\alpha = .88$; Kane et al., 2004: $\alpha = .86$). This means that measurement errors, for example fluctuations in attention or fatigue, make up less than 30% of the scores. Complex span tasks such as the Reading Span task also correlate well (i.e. moderately or strongly) with measures that are indicative of higher-order cognitive abilities, such as fluid intelligence, and selective attention (Redick, et al., 2012: 167).

All tests were presented on an Acer Aspire V5 laptop with an 11.6" LED screen set to a 1366x768 pixel resolution. The tests were administered in a relatively quiet, not too bright, location. The researcher made sure that the test takers were seated comfortably and that there was good stimuli visibility and/or audibility. Participants were wearing headphones for the aural Digit Span measure and were encouraged to use them for the Reading Span measure to block out ambient noise. All participants had normal or corrected-to-normal vision and reported being comfortable with using computers. The test scripts were delivered via Inquisit 4.0 (Millisecond Software, LLC).

Auditory Digit Span Procedure

For this measure, participants hear a sequence of single digits (ranging from 1 to 9), each digit being presented for the duration of 1 second and at the frequency of 44.1 kHz (at a rate of 16 bits). Then they see a visual recall prompt and have to recall the digits in the order of presentation. The sequences are drawn from a pool of 200 semi-random digit sequences. They are selected so that digits are not repeated and no patterns can be determined (e.g. no equal distances between numbers, no "sequence rules"). If the participant's span surpasses 9, however, it is inevitable that a single digit occurs twice in the same sequence. The smallest presented set size is three digits for the Forward Span and two digits for the Backward Span. The participants have three practice trials before starting the forward trial block and two practice trials before starting the backward block. After each practice, the participants receive feedback on their performance. This procedure is supposed to make sure that the task has been learnt before the test trials start. For the test trials, the sequence length increases by one digit after each perfect recall. If the participant recalls an incorrect sequence, the same length is tested in the next trial. If the participant cannot recall that sequence correctly either, the next one will be shorter by one digit and the sequence length will build up again from there. The forward

span experiment comprises 14 trials. The biggest set size that is recalled perfectly twice is counted as the maximum set size, but the programme also counts the correctly recalled digits in partial sequences.

It should be noted that all participants go through all 14 trials, meaning that they have several chances of reaching their maximum span size. This is important as idiosyncratic interferences might arise from the randomly presented sequences (e.g. someone could hear a sequence that is either a duplicate or a permutation of their mobile phone number). But even more importantly, the performance variability is greatest around the maximal span, so several trials approaching that threshold are beneficial. In theory, the maximum size that could be reached in 14 trials and starting with a sequence of three digits is 16 digits (and 15 digits for the Backward Span). In the backward condition, the procedure of administration is the same as in the forward condition, but participants are prompted to recall the digits they have heard in the reverse order of presentation. The font style of the instructions and visual prompts is Arial, the exact wording of the instructions can be found in Appendix B.

Automated Reading Span Procedure

For this measure, participants start with a welcome screen and work their way through the initial instructions. They start the training with a letter set practice (sets of two or three letters to recall). Then they are presented with a selection of sentences and have to indicate whether the example sentences are TRUE or FALSE, based on whether they make sense, or not. After practising the letter recall and the sentence judgement separately and receiving immediate feedback, participants are then trained on the dual task. After five practice trials, the testing session starts. Up to then, participants have the chance to ask for clarification, if necessary. Letters and sentences are selected randomly from 81 sentences (see Appendix C) and the following letters: F, P, Q, J, H, K, T, S, N, R, Y, L. The font style is Verdana and the complete task instructions can be found in Appendix D. Each test trial consists of a number of letter/sentence combinations and ends with the prompt to recall all letters in the correct order of presentation. Item sizes vary from three letters to the maximum of letters that a participant can recall perfectly plus one. At the end of the test, the programme gives feedback on five measures: The Reading Span score, which is the sum of all perfectly recalled sets, the total of letters recalled in their correct position, the total number of sentence errors, the number of accuracy errors,

and the number of speed errors. A perfectly recalled set is defined as a complete set, with all its letters recalled in the sequence they were presented in.

To put the span measures into context, the following graphic illustrates which parts of the Working Memory system this study is investigating:

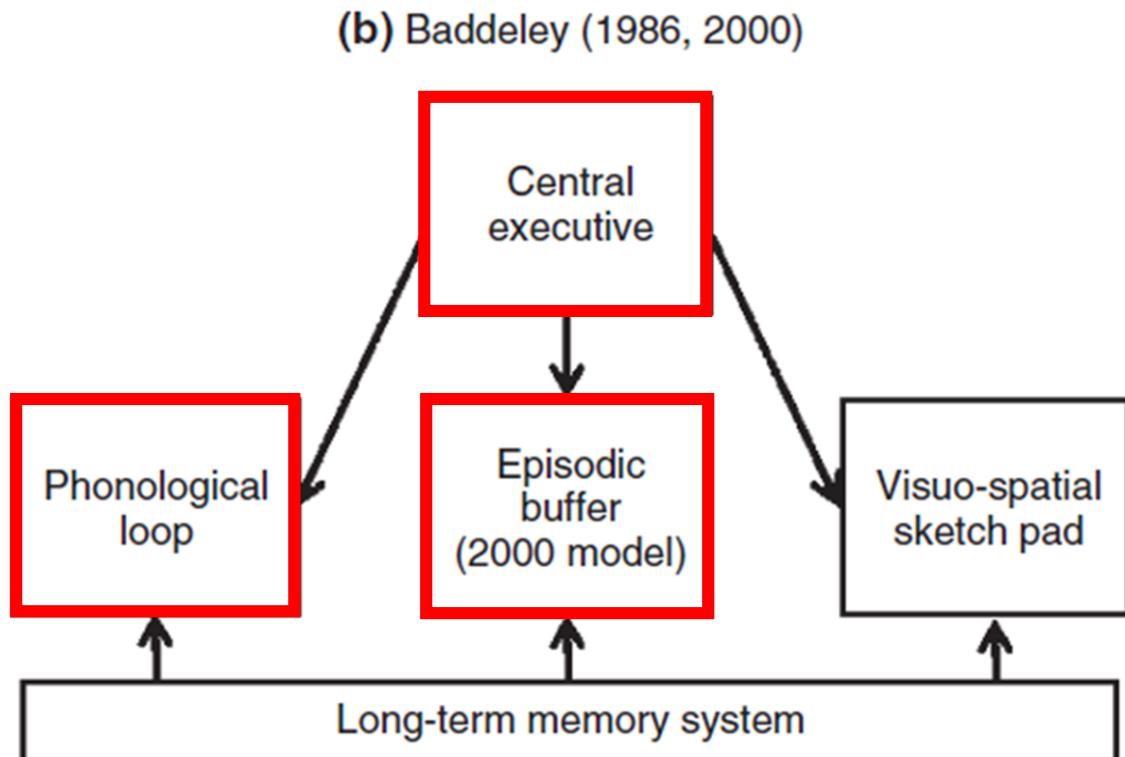


Figure 3.4: Baddeley's Working Memory model after Ricker, AuBuchon & Cowan (2010: 574), the highlights are mine

The Phonological Loop is important for the detection and processing of features in the aural input. Features must be kept activated long enough to be available for processing and acquisition. In MOGUL terms, the items in the Phonological Loop/Phonological Short-Term Memory are the activated representations in the phonological module's store. This part of the system is measured via the automated Aural Digit Span task. The Central executive is what directs attention to the items that are available for processing. In some researcher's opinion, executive attention is the main individual difference factor when it comes to Working Memory performance (Engle, 2002). From a MOGUL perspective, there is no need for a central executive, so the role of attention in keeping task-relevant information active is not an a priori determiner, but a result of processing. Cowan's Embedded-Processes Model is a more memory-specific conceptualisation of the central executive and compatible with this view. No matter whether the component is

conceptualised as a voluntary control mechanism or an emergence of the system, in this experiment, the attentional focus is measured in terms of the digits available for recall and manipulation despite decay over time (Digit Span) and the letters available for recall despite competing elements from the distractor task (Reading Span). The episodic buffer, according to Baddeley (2000), is the place where information from different sources is integrated and manipulated, i.e. where learning takes place. Within the MOGUL framework, this would be the total of activated and available representations that are being matched via the respective modules' interfaces. It can be hypothesised for this experiment that this construct will play a role for the manipulation of the digit sequences in the backward condition of the Digit Span task and for the establishment of form-meaning connections during Processing Instruction.

4 LEARNER VARIABLES: SELECTION AND OPERATIONALISATION

Following Snow (1991), the emphasis of this study is on *who* benefits from an instructional treatment *when* and *why*. Aptitude in this sense is not defined as a specific ability to perform an action, no matter what the circumstances are, but is “interpreted as personal readiness to profit from particular treatment situations” (Snow, 1991: 205). It is therefore necessary to collect as much data regarding the treatment context as possible: “Every research design involves multiple aptitudes and higher order interactions whether it includes them formally or not.” (Snow, 1991: 207). Previous research has shown that less controlled environments emphasise individual differences, especially in terms of anxiety and cognitive arousal. Students with very high or very low anxiety seem less able to self-control attention to the task and benefit from structured learning tasks, whereas a moderate degree of arousal/anxiety allows attentional control and leads to better performance on less structured tasks (Snow, 1991: 206).

This study aims to explore as many interactions as possible by measuring a selection of different learner variables which could all have a positive or negative effect on the processing of input before, during, and after a Processing Instruction treatment. The study strives to describe the processes and circumstances of data collection in as much depth as possible by presenting an account of the research process and by combining different formats and sources of data. To this end, the processing demands of the tasks are analysed (Robinson, 2001: 370), but also the participants’ subjective perceptions of task difficulty. The treatment context will be kept as stable as possible by using computer-delivered Processing Instruction materials, measuring Working Memory via automated span tasks, and generally making sure that the circumstances of the data collection are as similar as possible for all participants. Differences in performance before, during, and after instruction should then stem from the individual differences that the participants bring to the task. In the following section, the learner variables chosen for investigation in this study will be presented and their selection will be justified with regards to their potential mediating effects on the individual’s processing resources. Once their significance is established, the constructs will be presented one by one and their operationalisation in the context of this project will be described. Again, a process-perspective will be adopted and the theory will be presented together with the concrete instruments that have been adapted for the purpose of this study.

4.1 Variables of interest

This section introduces the potential mediating variables selected for investigation in this study. Working Memory is one part of this resource pool, but there are other factors which are likely to play a role in the processing of linguistic input and, ultimately, the learning rate and/or degree of proficiency displayed by the learner. It is clear that any Working Memory Capacity, however large, is obsolete if the learner does not pay attention to the input or does not engage with the instructional materials. Interest in the subject and a positive attitude towards the language can therefore help increase the learner's general willingness to allocate attention to the input (Gardner, 1985: 41). Attitudes towards the subject might also influence how open to new experiences the learners are (Cooper 2002: 211). Motivation is one of the main building blocks in individual differences research regarding Second Language Acquisition, next to cognition and emotion, but it is notoriously elusive to experimental research. While motivation might influence the willingness to allocate attention to a task or learning endeavour, it is a lot more dynamic than e.g. interest, meaning that it might be more relevant to individual tasks and learning situations whereas interest and attitudes are the variables that lead to the decision to study a certain language and help students to persist in the face of adversity.

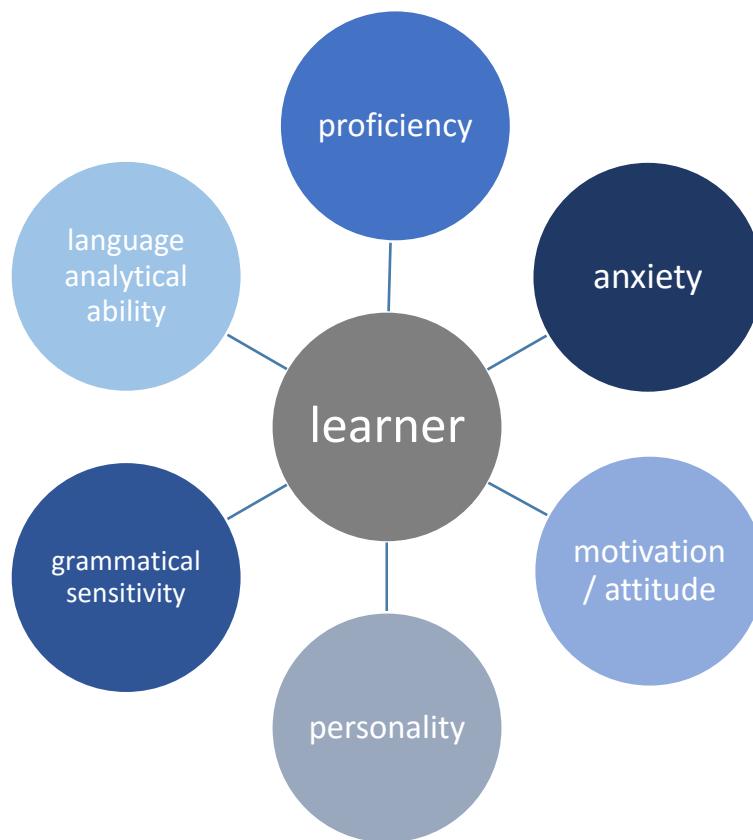


Figure 4.1: Overview of selected learner variables in context

The selected learner variables for this study are shown in Figure 4.1 above. All of them have been hypothesised to have a direct or indirect effect on second language learning and processing, but in the context of Processing Instruction, they have never been investigated as part of a functional whole that could be called processing resources. In line with a MOGUL interpretation of language acquisition, there is no isolated Working Memory construct. Working Memory is distributed over the whole network of processing resources because each processing module consists of a processor and a store. Working Memory comprises the activated representations in the respective stores, i.e. the representations that are available for processing. In accordance with that, language proficiency seems to be a factor which plays a role in language processing as well as language acquisition. Generally speaking: the more proficient the learner, the lesser the strain on their processing resources as representations get more complex over time. Co-activated representations are faster to process and have better chances of winning the competition, a phenomenon which is usually called chunking. The global proficiency level of participants in this study will be operationalised as the overall grade that the participants received in the previous term. Additionally, the pre- and post-test design allows to measure the more local proficiency level regarding morphological case marking in general and switch prepositions with variable case marking in particular.

Connected with the notion of high proficiency levels is the concept of language learning aptitude. This variable has been conceptualised in many different ways, but the key question has been: Which language-related abilities make the individual a successful language learner? Previous research has investigated concepts such as grammatical sensitivity and language analytical ability. These concepts are interesting because they concern the learner's ability to analyse language and make form-function connections in the absence of explicit meta-linguistic rules. But the role of language aptitude has been stressed for classroom-based language learning as well (Krashen, 1982: 31). It is, however, important to pay attention to the nature of the instructional and assessment tasks as different forms of learning and assessment might involve various sub-components of language learning aptitude to varying degrees, as previously described. While language analytical ability has not been investigated yet in the context of Processing Instruction research, some researchers have found an effect of grammatical sensitivity for the application of explicit information in Processing Instruction on German case morphology (Culman, et al., 2009; VanPatten & Borst, 2012; VanPatten et al., 2013). While language learning aptitude should not play a role in Structured Input Activities as the explicit information is provided and the input is manipulated to facilitate form-meaning

connections, there might be an effect for instruction on German case inflections. Moreover, individuals might rely on different processing paths for different tasks, relying sometimes on analytical processing and sometimes on memory-based processing (Tanner, 2013; Tanner & Van Hell, 2014; Tanner, et al., 2014).

Another factor that has been neglected in Processing Instruction research is personality. In combination with task modality and/or complexity, personality might play a role in the allocation of processing resources. This has consequences for the perceived difficulty of a task. An extreme introvert might find a spontaneous face-to-face dialogue a lot more demanding than a written composition task that requires extensive planning and sustained attention. The contrary might be true for an extreme extravert. In any case, the task itself needs to be considered when it comes to an estimation of processing demands on the learners' resources. Capacity limitations can be alleviated by sequencing course content, by using simplified or manipulated input (as in Processing Instruction), by helping learners recover from mistakes and by involving learners in the goal setting process. In the case of Processing Instruction materials, i.e. Structured Input Activities, the input is manipulated in a way that makes form-meaning connections more straightforward by discarding all additional information that could encourage the use of inappropriate processing strategies. The learners, on the other hand, do not have much choice in Processing Instruction: the input, the tasks, and the sequence of tasks is strictly planned, leaving little room for a personalised approach. It will be interesting to see how the participants rate the difficulty of the materials and how they engage with them, based on their personality and their general interest in the topic as well as the study. Interest, motivation, and attitude supposedly have a beneficial effect on language learning, whereas anxiety has been seen as the flipside of this. While a high degree of anxiety has negative consequences and can lead to the interruption of language learning, a moderate degree can help focus attention and raise the neural arousal levels, which is especially important for otherwise under-stimulated extreme extraverts. An interaction between anxiety and personality is therefore expected.

The approach taken in this study combines some characteristics of Robinson's aptitude research, i.e. the notion that learner variables should not be investigated in isolation as they do not exist in isolation, with the process-orientation of Skehan and Dörnyei's (2003) aptitude construct. It addresses the claim that language learning aptitude, and especially Working Memory as one part of it, should be investigated with regard to actual tasks and individual learners.

To recap, the rationale for this study can be summarised as follows:

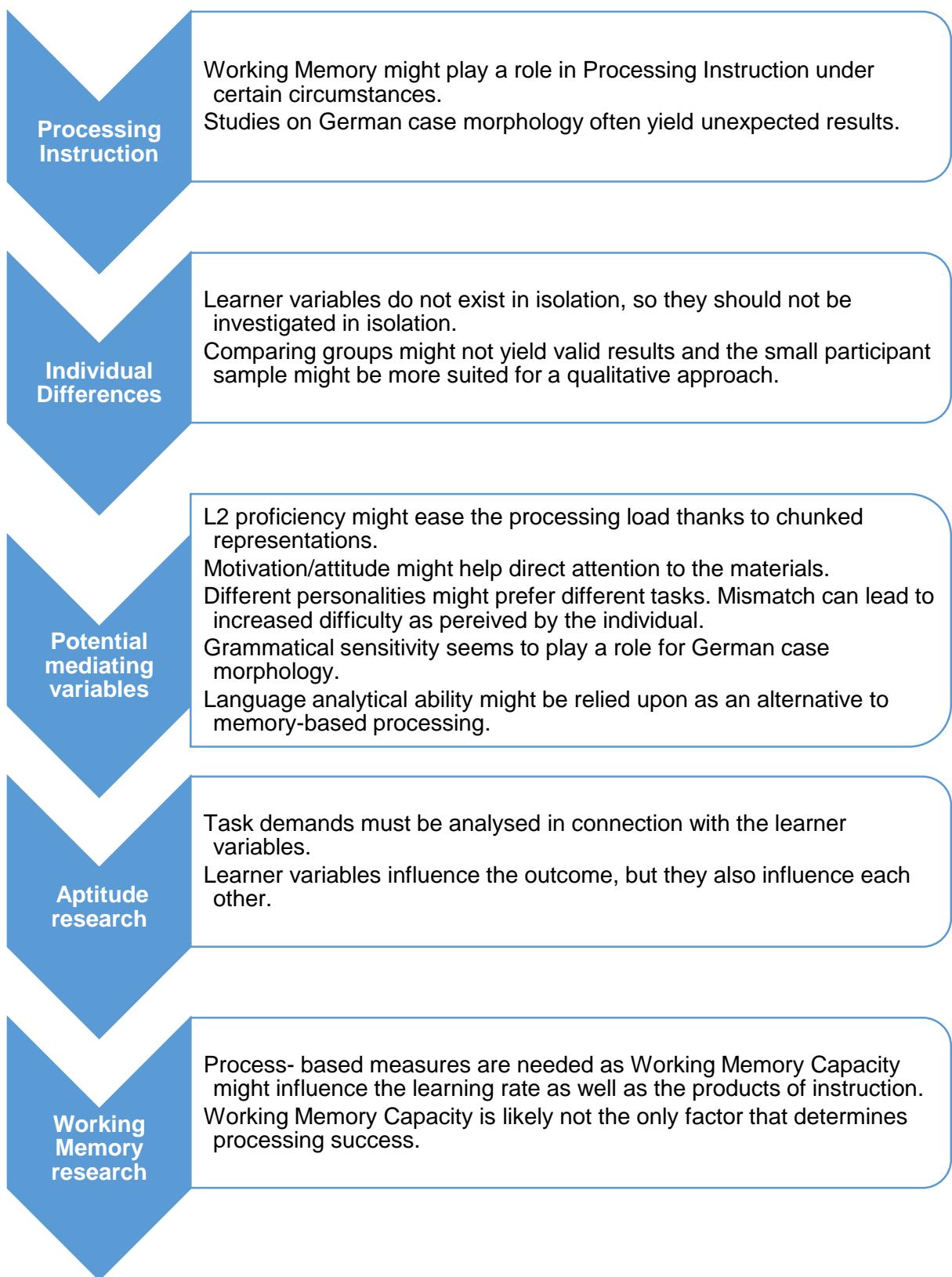


Figure 4.2: Gaps previously identified in the literature

4.2 Attitude, Interest, and Motivation

Motivation has emerged as the best predictor of successful language learning in many studies (for an overview of studies looking for the characteristics of a “good” language learner, see for example Dewaele & Furnham, 1999) and can be considered one of the monumental factors in second language acquisition, next to cognition and affect. It is important to consider, however, that different learning contexts might involve different factors to different extents, and potentially lead to different outcomes in terms of learner behaviour and/or performance assessment. As Skehan (1991: 281) points out, motivation is an accompanying factor due to the interaction of learner-external and -internal factors impacting on the learning process, but motivation is also a result of learning due to the fact that learning success comes with external as well as internal gratification. One could therefore say that motivation has a psychological and a social dimension.

The Attitude/Motivation Test Battery (Gardner, 1958, 1960; Gardner & Lambert, 1972) has been developed to measure some of the social and emotional factors that account for individual differences in second language acquisition, mainly along the two dimensions of integrative versus instrumental orientations of motivation. The original test battery has been designed for English native speakers learning French in Canada and comprised 12 factors: Interest in Foreign Languages, Parental Encouragement, Motivational Intensity, English Class Anxiety, English Teacher Evaluation, Attitudes toward Learning English, Attitudes toward English-speaking people, Integrative Orientation, Desire to Learn English, English Course Evaluation, English Use Anxiety, and Instrumental Orientation. Most of these original factors were measured by five positively and five negatively coded questions that participants rated on a six-point Likert scale.

For this study, all questions have been adapted to refer to German as a Foreign Language. 50% of the items have been positively worded and 50% have been negatively worded to counteract acquiescence and response bias effects. The questions are presented in a pseudo-randomised manner, counterbalancing the versions of each item. To abridge the questionnaire, only a handful of factors have been chosen and the questions for each factor have been reduced to four. The following constructs were chosen: Interest in Foreign Languages (as a baseline, as the participants were all language students), L2 use anxiety, motivational intensity, instrumental orientation, class anxiety, and attitude toward the language. Attitude, defined by Gardner (1985) as “an evaluative reaction to some referent or attitude object, inferred on the basis of the

individual's beliefs or opinions about the referent" (9), is hypothesised to have an effect on the learner's behaviour. If the learning of the target language is seen as desirable and worth the effort, then it is more likely that the learner will align their behaviour with the goal of learning that language. A favourable attitude is the foundation of motivation, which comprises not just the desire to learn the language, but also the effort that the learner puts into achieving that goal. The orientation indicates whether the learner's interest is directed towards the language itself, or whether they see the language as a means to an end, e.g. the improvement of their employability. The anxiety measures were included because anxiety, especially Foreign Language Classroom Anxiety (FLCA), is linked to self-related cognition, i.e. constant worrying about how one performs, how one is perceived by others, and what could possibly go wrong (Dewaele, 2013: 670). In a sense, excessive self-related cognition is a constantly interfering process that taxes the processing resources and decreases attentional capacity. Increased anxiety levels are also linked to the personality dimension extraversion-introversion, as will be discussed in the next section. The complete questionnaire can be found in Appendix E.

Factors not included in the questionnaire were discarded because they either had no great relevance for the participant sample (e.g. Parental Encouragement) or for the experiment (e.g. Teacher Evaluation). It should be noted that these sub-tests were used as independent indicators in this study whereas the original Test Battery also calculates four composite scores, namely integrativeness, motivation, attitudes towards the learning situation, and attitude/motivation. Gardner's research on attitude and motivation has contributed significantly to the theory and methodology related to the investigation of these constructs (Skehan, 1991), but the measurements are based on students' self-reports. While the reliability estimates for the Attitude/Motivation Test Battery are generally acceptable, it has to be kept in mind that the materials have been adapted for this particular study. While the researcher has taken every care to transpose the original questions as truthfully as possible, the estimates cannot be extended to the adapted materials. Furthermore, attitude tests generally are dependent on contextual influences. As Skehan (1991: 284) points out: the original dimensions proposed by Gardner are mediated by the context of the learning situation, the individual learner's background, and the characteristics of the language as well as the places in which the target language is spoken. The questionnaire data were collected at the start of the term, so scores might have been influenced by the relative distance of assignments, impact of new teaching staff, etc. Potential measurement errors could therefore stem from both the instruments and the dynamic constructs they measure.

4.3 Personality

As Dewaele & Furnham (1999) note in their article “Extraversion: The Unloved Variable in Applied Linguistic Research”, extraversion as an independent variable in its own right has not received a lot of attention in Second Language Acquisition research, apart from a short period of interest in the 1970s. Yet it is conceivable that this personality dimension plays a role not only in speech production and fluency, but also for the perception of a task’s difficulty by the learner. It is plausible that a complex tasks of longer duration and with substantial planning involved would be perceived differently by an introverted as opposed to an extroverted learner (Eysenck and Eysenck, 1985). As the extraversion-introversion dimension is linked to impulsivity, different styles of approaching language learning (materials) could be expected as well. Extraversion as a variable was chosen for this study as the ability to focus on task-relevant information and to keep the attention on-task might differ depending on the task and the personality of the learner. While no oral production measure is employed in this study, differential effects on different types of activity as well as on different pre- and post-test materials are expected.

In addition to the effect of extraversion-introversion on the products of instruction, there could also be an effect on the scoring consistency during instruction. The fact that the instruction is computer-delivered in this study might have differential effects on the participants as another factor that extraversion is correlated with is sociability. Individuals scoring high on extraversion generally find solitary tasks less enjoyable than more introverted persons (Eysenck and Eysenck, 1964: 8). The effects of extraversion-introversion are a lot more complex than that though: While there is a tendency for introverts to outperform extraverts on long-term memory tasks, the opposite is true for immediate-recall tasks (Eysenck, 1981: 203). There might be differential impacts of extraversion on the immediate post-test and the delayed post-test, respectively. It should also be noted that there appear to be links between extraversion and better stress resistance as well as lower anxiety levels (Dewaele and Furnham, 1999: 515) – which might affect how participants score on the Attitude/Motivation Test Battery. Higher anxiety levels would also be likely to result in a decrease of attentional capacity for extremely introverted participants, but not for extremely extraverted ones.

The Eysenck Personality Inventory was developed to assess personality traits based on two independent scales: extraversion-introversion (E-scale) and neuroticism-stability (N-scale). Those two scales are based on extensive research into factorial models of personality and their construct validity has been tested in terms of correlations with other

personality inventories, self-ratings, and third-person judgements (Eysenck & Eysenck, 1981: 16–19). The original version comprises 24 Yes/No questions for each of the scales plus nine Yes/No questions (L-scale) to account for response distortions, i.e. unfaithful answers based on acquiescence or desirability effects (Eysenck & Eysenck, 1981, p. 20). The EPI is one of the most widely and broadly used personality tests, in use since the 1960s and assessed as stable and reliable through factor analyses with data from 22 countries (Dewaele & Furnham, 1999: 511). The extraversion-introversion dimension has also the advantage of being based on Eysenck's (1967, 1990) Arousal Theory and can be linked to biological factors. While some of the hypotheses proffered by the Arousal Theory have been criticised in the light of more sophisticated insights into neural and chemical processes, it is an important attempt to link personality theory with behavioural as well as biological factors.

For the present study, the questionnaire has been shortened to contain only 16 questions for the E-scale and 5 questions for the L-scale. Any problematic questions (e.g. double negatives) were reworded, in accordance with recommendations for the construction and administration of questionnaires (Dörnyei, 2010; Dörnyei & Csizér, 2012), and the language has been slightly updated. The complete questionnaire can be found in Appendix F. Overall, the Eysenck Personality Inventory (EPI) is considered to be a reliable and valid measure of a well-established personality dimension, with the manual reporting test-retest and split-half reliability scores above 0.8 (Eysenck & Eysenck, 1981: 14–15). The EPI (Eysenck, 1964) was revised and extended to include a psychotism scale (Eysenck & Eysenck, 1975), but the new version, called the Eysenck Personality Questionnaire (EPQ), seems less relevant for this study as it almost exclusively measures sociability (Rocklin & Revelle, 1981). The older EPI, measuring both impulsivity and sociability, seems more appropriate in relation with computer-delivered instruction. Impulsivity as an underlying personality factor might play a role as a high degree of impulsivity can lead to decreased planning time on essay tasks and/or shorter response times on other tasks.

To a certain extent, personality influences the overall learning experience. It is a modifying variable and some personality traits might affect attentional resources, which might play a role in terms of task modality. For an extraverted learner, a spontaneous conversation in the target language might be less daunting than for an introverted learner. In the latter case, the felt pressure might contribute to a decrease in attentional resources. On a rather long, written essay task, the introverted student might feel more comfortable than the extraverted student who might get more easily bored or has trouble keeping their attention

on-task. More importantly though, Eysenck's (1990) Arousal Theory suggests that extreme introverts are constantly experiencing cognitive over-arousal, whereas extreme extraverts suffer from constant lack of cognitive arousal. This leads to the former avoiding situations that provide additional stimulation and the latter seeking out such situations. Together with anxiety as a state of heightened cognitive activation, personality could play a role in the quality and efficiency of processing as the heightened neural activation in introverts should spread and lead to processing noise and increased competition, hence interfering with the learning task. Extreme extraversion, on the other hand, would be accompanied by sub-optimal neural activation levels and could have a negative effect in as far as the learner would need to process representations more frequently to achieve the optimum resting activation levels.

4.4 Grammatical Sensitivity

The construct of grammatical sensitivity describes the learner's ability to detect the function of words in sentences (Skehan, 1991: 277). The most popular measure of this ability, aptly named Words in Sentences, is a subtest of the Modern Languages Aptitude Test and requires learners to indicate which word out of a selection of underlined words in a target sentence has a similar function as the word underlined in the example sentence. To make a correct decision, the test takers are not required to have metalinguistic knowledge of that function and they do not need to produce an explicit rule either:

- 1. Yesterday, Mary caught a fish at the lake.**
Cindy cut a cake with a knife.

Figure 4.3: Words in Sentences example

It should be noted, however, that some researchers have categorised grammatical sensitivity, together with rote rehearsal, as part of an ability factor called "metalinguistic rule rehearsal" (Robinson, 2001: 374), which is associated with an aptitude for explicit rule learning. While Processing Instruction is not about rule learning and while the role of explicit information for Structured Input Activities has been evaluated as facilitating rather than essential, Processing Instruction is still an explicit pedagogical intervention which

does provide explicit information on the target feature and the default processing strategies.

Robinson (2002: 131) suggested that Processing Instruction “may be a technique for inducing focus on form that is differentially affected by the fourth aptitude complex [...] – particularly the grammatical sensitivity component of what I have termed metalinguistic rule rehearsal”. Previous studies on Processing Instruction have found either no predictor effect for grammatical sensitivity or only for participant groups receiving explicit information on the target feature (e.g. VanPatten & Borst, 2012). Considering that the participants in this study will receive a moderate amount of explicit information on German case marking, switch prepositions, and the potential failure of their default processing strategies, it seems fair to include grammatical sensitivity as a variable. It should also be noted that most subtests of the Modern Language Aptitude Test (Carroll & Sapon, 1959) have been heavily criticised on the grounds that the authors tested a plethora of different measures and chose to include into the final package only those which correlated significantly with scholastic achievement scores (Robinson, 2001: 371). The applicability of those tests to different forms of instruction has therefore been questioned, as has the initial work of Carroll (Dörnyei & Skehan, 2003: 593). The Words in Sentences subtest, however, has yielded positive correlations with achievement in communication-oriented instructional settings (Robinson, 2001: 371–2). The test used in this study can be found in Appendix G.

4.5 Language Analytical Ability

As described previously, the key assumption in Processing Instruction research is that the Structured Input that learners receive is maximally reduced in complexity and minimally taxing for the learner’s processing resources. Because of this lack of ambiguity, redundancy, and pragmatic complexity, the form-meaning connections are supposedly relatively straightforward. Consequently, learners should not need exceptional language analytical skills to attach meaning to forms in the input. If this is correct, then it should not matter how well learners can abstract patterns and rules from the input as the skill gap between analytical and less analytical learners should be minimal when exposed to Structured Input Activities. This is one reason why a test of Language Analytical Aptitude has been included. Another reason is that, as proposed by Wesche (1981) and discussed by Skehan (1991, 1998), there might be two types of learner styles: one that relies on

memory and one that relies on analytical abilities. Those styles could be independent of one another, but recent research in neurolinguistics points to the opposite: brain potentials associated with a reliance on memory seem to be negatively correlated with brain potentials that are associated with analytical processing (Tanner & Van Hell, 2014; Tanner, et al., 2014). As this study is already exploring the memory component, it makes sense to see how learners score on a language-analytical task.

To this end, the artificial-language section of the University of Oxford Modern Languages Admissions Test (Admissions Testing Service, 2014) has been included. This test is a measure of what Carroll (1965, 1971) has termed inductive language learning ability or “the ability to examine language materials and from this to notice and identify patterns of correspondence and relationships involving either meaning or syntactic form” (Skehan, 1991: 277). In this task, the participants are shown sentences of an artificial language called Pip. For each sentence, an English translation is provided. By comparing the sentences’ common and diverging elements, the participants should be able to figure out which Pip word exactly stands for which English word. Part of the challenge is to notice that in Pip, the plural is indicated by a duplication of the vowel: *pit* = dog, *piit* = dogs. Another important observation is that Pip marks the object case by vowel gradation: *pit* = dog (subject), *put* = dog (object). The first section comprises sentences in Pip and their English translations. Based on these correspondences, the test-takers can infer which Pip word has which English meaning. The second section then requires the participants to translate two new Pip sentences into English, demonstrating that they have grasped the patterns of the artificial language and do not rely on word order. The third section is a sentence translation from English into Pip, requiring the productive application of these patterns.

While the first section builds on more general analytical skills that have to do with the systematic comparison of words which appeared in more than one sentence, the other sections build on the realisation that Pip marks subject and object cases as well as singular and plural. Without this observation, the likelihood of translating the sentences correctly is very slim. The materials used in the main study diverge slightly from those used in the pilot study inasmuch as the word order of the Pip sentences was changed to OVS. The notes of one participant in the pilot study indicated that the translation had been based on word order rather than case marking, so it was crucial to change this for the main study. The first part of the materials can be seen in Figure 4.4 below and the complete Pip materials for both studies can be found in Appendix H.

Now I'd like you to have a look at an invented language. It's called Pip. Read each of the examples carefully and you will be able to work out the meaning of each word. Don't rely on word order! (Note that there is no "the" or "a(n)" in Pip.)

pit sak run	The dog chased the cat.
rin lup kat	The cat watched the mouse.
mup taw kid	The horse saw the teacher.
liip puut kat	The mice watched the dogs.
kid taw muuk	The horse saw the squirrels.

Figure 4.4: The artificial language Pip

The presentation of aptitude as a multi-component system leads to the hypothesis that different components of this complex can be developed to different degrees in any individual, making it likely that there is not the one profile of a successful foreign language learner, but that there is a plethora of different patterns that benefit more or less from different forms of instruction and perform better or worse when assessed, depending on the nature of the assessment tasks. The MOGUL interpretation of Working Memory also points to a more diverse distribution of processing resources as well as modality-dependent capacity limitations that affect the individual's processing abilities to different degrees, depending on the nature and the consistency of the respective processor's output. A similarly complex picture emerges from research on anxiety, showing that, depending on the linguistic ability and cognitive arousal of the individual, moderate degrees of anxiety can be beneficial in certain learning environments (Robinson, 2001; Dewaele, 2013). The general predisposition for cognitive arousal can then be linked to Eysenck's (1967, 1990) Arousal Theory of personality, pointing to an impact of personality on processing. This all fits well within the MOGUL framework, where language acquisition is a consequence of processing and where (resting) activation levels decide which representation wins the competition and/or which processing route is the fastest.

As outlined previously, this study aims to collect data on several learner variables as well as pre- and post-test scores, thus generating a lot of data despite a small sample size. To reflect this and to keep the amount of information manageable for both the participants and the researcher, the data collection tools have been shortened where necessary. All of the instruments have in common that they are tried and tested materials with acceptable reliability scores. Some of them have been used for more than 50 years in psychometric research and still are widely used today. Their wording had to be updated in places and some of the questions have been reformulated based on Dörnyei's (2010)

recommendations, but overall, every effort has been made to preserve the original's characteristics. Table 4.1 below sums up the selected learner variables and their hypothesised effects on Working Memory Capacity limitations during input processing:

Variable	Potential effect on Working Memory Capacity (as seen from a MOGUL perspective)
Extraversion	If very extraverted individuals suffer from a lack of neural arousal, then they might have a disadvantage as representations need to be processed more often to reach significant resting activation levels. Extreme introverts, on the other hand, might suffer from too much competition between highly active representations.
Grammatical sensitivity and language analytical ability	If individuals are better at identifying functions of words within sentences, they should be able to establish form-meaning connections more quickly as they can use convergent information from inside and outside the core language modules to co-index representations. Whether this translates into a distinctive advantage for the processing of Structured Input in the context of Processing Instruction remains to be seen.
Proficiency/exposure to L2	The more often a representation is processed, the higher is its resting activation level and the more likely is the feature to win the processing competition. Proficiency is then a result of exposure to input and depends on internal processing frequency levels. Participants with more frequent/recent exposure to German should have an advantage, at least initially. According to previous Processing Instruction research, these individuals might benefit least from Processing Instruction though.
Background	Other languages, depending on their processing history, can facilitate or impede the processing of German case morphology. This could be one source of individual differences.
Motivation/attitude	Motivation, interest, and attitude play a more global role in Second Language Acquisition and might not be crucial factors in the immediate context of this experiment as the participants find themselves in special circumstances. However, these factors have played a role in the students' choice of programme and have led to the result that all participants have been studying German for years despite the language's reputation.
Anxiety	A moderate degree of anxiety supposedly helps under-stimulated individuals in attending to the (classroom) input. While the sample is too small to investigate the optimal degree of anxiety for each personality type, extremely high or low anxiety scores in combination with extremely low or high extraversion scores might give an insight into the effects of anxiety on performance before, during, and after instruction.

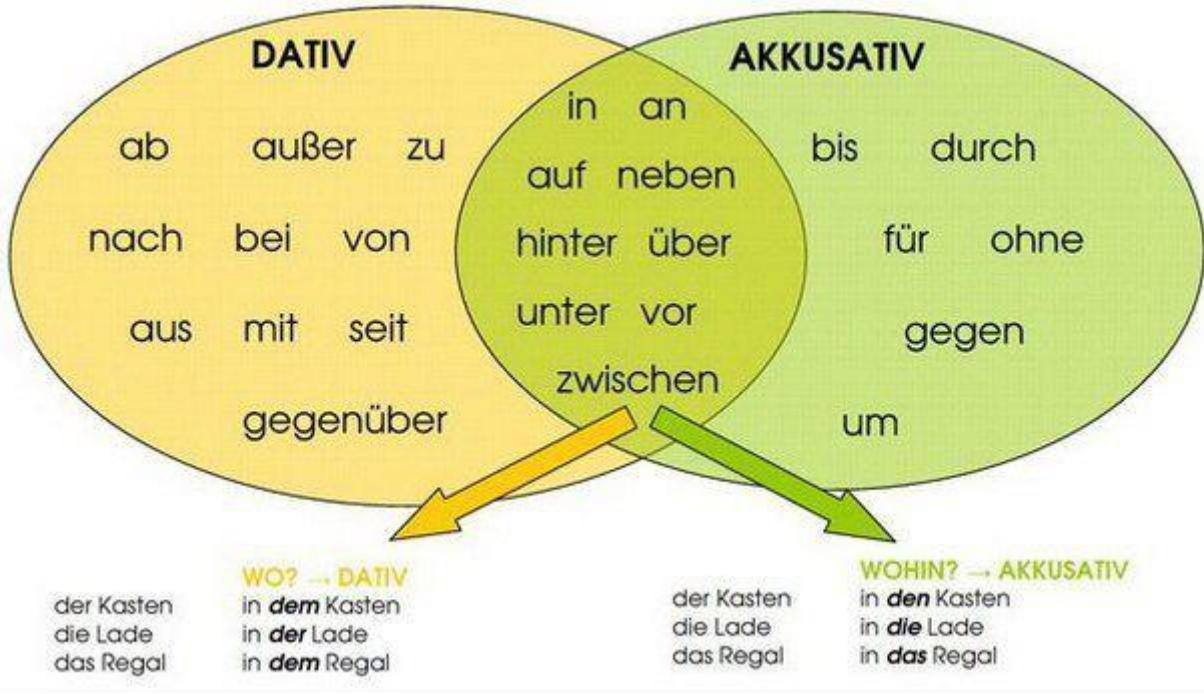
Table 4.1: Summary of variables of interest and their potential impact on Working Memory

5 EXPERIMENTAL TREATMENT

Previous research on Processing Instruction and German case inflection has shown that the interpretation and production of German case markers can be successfully taught using Structured Input Activities (VanPatten & Borst, 2012; White, DeMil, & Rice, 2015). Usually, the materials target the First Noun Principle as a processing strategy to be changed. Word order is not a reliable indicator of subject and object roles in German sentences, but morphological case marking is. Structured Input Activities targeting this principle therefore use OVS and SVO sentences in order to discourage the learner's reliance on word order for parsing and force them to process the case markers. This study chooses a different path. Based on the conviction that case markers are inherently meaningful, this experiment targets the semantic/pragmatic distinction between accusative and dative case following German switch prepositions. While the underlying cognitive schemata will be discussed in detail in section 8.4, this chapter will be describing the target feature and the processing problems that make it suitable for Processing Instruction research. Based on the identification of the problematic default processing strategies, the development of the Structured Input Activities will be explained and the instructional and assessment materials will be presented. Having introduced all variables and the target feature at that point, sections 5.4 and 5.5 will describe the data collection and the participant sample of the pilot study and the main study, respectively. The chapter will end with a summary of the steps that have been undertaken to maximise validity and reliability in this study.

5.1 Target feature: German switch prepositions

Switch prepositions, or dual prepositions, are a class of nine German prepositions that switch case depending on whether a location or a direction is encoded. Usually, prepositions require the following noun or pronoun to display one mandatory case, the preposition *ab* (English equivalents: from, ex) for example always requires the dative case, the preposition *durch* (English equivalents: by, through, per, via) always requires the accusative case. Switch prepositions can select either dative or accusative case, depending on how the action is framed. One can ask for the phrase following the preposition using *wo?* (where?) or *wohin?* (where to?). Figure 5.1 shows the most frequently used dative and accusative prepositions as well as the nine exceptions that can select either case.



<http://deutsch-mit-frau-dalo.blogspot.co.uk/p/grammatiktafeln.html>

Figure 5.1: German switch prepositions at a glance

Traditionally, the teaching of switch prepositions is based on insights derived from valence grammar: L2 learners of German are taught that some arguments are core arguments which depend on the verb, while others are optional and provide additional information. Prepositional phrases, then, either depend on the verb, or provide additional information. In the former case, students are encouraged to learn the preposition by heart, in combination with the verb. In the latter case, learners are prompted to work through an abundance of fill-in-the-blanks tasks or are required to produce discourse-level written output. Domínguez Vázquez (2005) provides an overview of the most common exercise types. Ritterbusch et al. (2006) show that most of the time, students can identify the gender of a noun if they are confronted with the form and explicit case information. They can also supply the case if they see the form and have gender information. But when presented with gender and case, they can produce the accurate form in only two out of three trials. This points to temporal differences in the development of interpretation versus production skills when it comes to German case marking. For native speakers of German, case marking is the most reliable cue and due to the relatively flexible word order and the potentially substantial gap between V2 position of the finite verb and a thematic verb at the end of the clause, it is preferred to relying on the semantics of the thematic verb: “studies investigating other types of verb-final constructions in German have shown that

German native speakers assign grammatical roles in a sentence prior to encountering the thematic verb" (Jackson, 2008: 885). It is quite obvious then that a learner's reliance on verb semantics can have negative consequences for the processing of German input.

According to Lee & Benati (2009: 42), the most important step in the development of Structured Input Activities is the identification of the processing problem that accompanies a target feature. In the case of German switch prepositions, several problems can be identified. The most fundamental problem, presumably, is that German marks most cases morphologically and therefore allows for relatively flexible word order. English native speakers, on the other hand, exhibit a tendency of relying on word order and their default processing strategy is to assign the subject role to the first (pro-) noun in the sentence (First Noun Principle). According to VanPatten & Borst (2012: 93), "[t]his strategy has negative consequences for object-first sentences, English-like passives in which the subject is not the agent, and case marking (learners may ignore it), among other grammatical structures." In the context of this study, it can therefore be expected that the participants will misinterpret OVS sentences if they do not pay attention to the case marking. It can also be hypothesised that they will not be sensitive to the accusative/dative distinction following switch prepositions. This does not only have consequences for the interpretation of the actual input sentence, but also for the further development of the learner's linguistic competence (Lee, 2004: 313).

Prepositions express additional information and therefore establish a semantic relationship with their context. They do have their own semantic content and are therefore less likely to be overlooked, according to the Primacy of Meaning Principle. Learners therefore typically struggle with the case morphology within the prepositional phrase rather than the preposition itself. While the meaning of different prepositions is quite clear, the differences in meaning expressed by the case morphology are less salient. The acquisition of German switch prepositions is especially difficult for learners because several decisions are involved in the correct interpretation and production of switch prepositions:

1. The learners must know the nine prepositions that can take either case.

Many learners of German learn each preposition together with the case it requires, i.e. *ab* + *Dat*. They learn about the existence of switch prepositions, but at first still express a preference for the first case they have learnt in combination with the preposition in question.

2. They must know when to use the dative and when to use the accusative case.

As mentioned before, the use of case depends on whether the preposition expresses direction or location. Learners often get confused because they distinguish direction and location based on the semantic content of the verb. This is not a valid argument though:

- a) *Die Katze läuft auf der Straße.* The cat walks on the road.
- b) *Die Katze läuft auf die Straße.* The cat walks onto the road.

Learners who pay attention to the verb only and associate a motion verb such as *laufen* (to walk) with the expression of directionality, are insensitive to the difference expressed by the two sentences above. Sentence a) describes a cat that walks on the road, whereas sentence b) describes a boundary-crossing event: the cat was walking elsewhere, maybe on the pavement, and now crosses a boundary when walking onto the road. The preposition in sentence a) requires the use of the dative, the preposition in sentence b) requires the use of the accusative. Note that in English, there is also a marked difference between the use of on and onto. This does not seem to help L1 English learners of L2 German though. They tend to express a reliance on content words, as described in the Primacy of Content Words Principle (VanPatten, 2007: 117).

3. They must know the right morphological markers for the case in question.

While the accusative case markers are identical with the nominative case markers except for the masculine singular and the personal pronouns, the dative case markers diverge from the nominative case markers and pronouns in all persons and genders. On the one hand, this makes the detection and interpretation of the accusative case difficult. On the other hand, it makes the dative case more salient for interpretation, but harder to accurately produce as there are more competing forms:

	Nominative	Accusative	Dative
masculine	der	den	dem
feminine	die	die	der
neuter	das	das	dem
plural	die	die	den

Table 5.1: Definite articles in different cases

4. The learners must know the correct gender and number of the following (pro-) noun in order to decline the article or use the right form of the pronoun.

Especially for English native speakers learning German, grammatical gender is hard to acquire. The categorisation seems arbitrary and the acquisition of gender can be frustrating at any stage of the process. The different (pro-) forms for different cases only add to this problem:

	Nominative	Accusative	Dative
1 st person singular	ich	mich	mir
2 nd person singular	du	dich	dir
3 rd person singular masculine	er	ihn	ihm
3 rd person singular feminine	sie	sie	ihr
3 rd person singular neuter	es	es	ihm
1 st person plural	wir	uns	uns
2 nd person plural	ihr	euch	euch
3 rd person plural/formal	sie/Sie	sie/Sie	ihnen/Ihnen

Table 5.2: Pronouns in nominative, accusative, and dative case, respectively

5. If there are weak nouns involved, then they receive a suffix to mark all cases but the nominative case.

The *n-Deklination* for weak nouns requires that not only the article word is marked for case, but also the noun itself:

- a) *Der Junge ist hungrig.* The boy is hungry.
b) *Ich gebe dem Jungen eine Banane.* I give the boy a banana.

Most of these weak nouns are also very “old” and very common nouns, so the learners are usually exposed to many instances in the input. According to the Input Processing principles, however, it is problematic that the suffix on the noun encodes the same information as the case marker on the article. Because the grammatical information is redundant, it is likely that learners do not become aware of the suffix and therefore do not acquire it easily, as laid out in the Preference for Nonredundancy Principle (VanPatten, 2007: 119).

6. In the absence of context, the question as to which case to use can depend on the verb (e.g. directive or stative verbs).

As seen in the cat example above, the semantic content of the verb does not always play a role. It does, however, provide crucial information in the case of directive vs stative verbs:

- | | |
|--------------------------------|---------------------------|
| a) <i>auf dem Tisch liegen</i> | to lie on the table |
| b) <i>auf den Tisch legen</i> | to lay s.th. on the table |

Note that there is only a minimal phonological difference between the verbs *legen* and *liegen*. This is a constant source of confusion and learners who do not perceive the difference and/or do not associate the correct meaning, cannot process the sentence accurately. Even in English, the distinction between lie/lay causes problems for native speakers. So the concept as such might not be absolutely clear to some L1 English speaking learners of L2 German or L1 German speakers.

7. Prepositional phrases can be rather complex, which means that the case marker usually ends up in a sentence-medial position.

The Sentence Location Principle (VanPatten, 2007: 125) sums up previous research on sentence processing by stating that learners tend to parse items in sentence-initial position and sentence-final position better than those in sentence-medial position. Case markers in prepositional phrases are therefore disadvantaged in several respects: They often appear in sentence-medial positions, are bound morphemes and therefore not very salient, and, in the case of the accusative markers, share forms with other cases – which makes them even harder to detect. Some of the endings, especially -em [əm] and -en [ən], are not easily distinguished in written and spoken German.

Considering the complexity of the target feature, it should thus not come as a surprise that this feature is not easily acquired. While the subject of switch prepositions is introduced as early as at the A2 level of the Common European Frame of Reference for Languages, even advanced learners of German struggle with its accurate interpretation and production. Considering this observation as well as the processing problems described above, German switch prepositions seem an ideal candidate for Processing Instruction. They also have the advantage that, thanks to relatively flexible word order in

German and the existence of more than 50 relatively common verbs which require the dative case for the direct object, the input can be structured in a way that forces learners to attend to the case markers in order to interpret the input correctly. With carefully manipulated input, determining whether the preposition in question is expressing direction or location can be rather straightforward, making the successful development of stable form-meaning connections more likely. For the computer-delivered Processing Instruction on dative/accusative case distinction following German switch prepositions, two 50-minute instructional sessions on Urban Life were designed. This overarching theme was chosen because it lends itself naturally to the description of locations and destinations, but also because it should provide a relatable topic for the London-based participants. As Farley (2005: 75) points out, it is important to create materials which do not consist of meaningless, mechanical drills whose only common denominator is the target structure: "Rather, all activity items should be interrelated and in some way refer back to the central theme of the activity. This will help instructors to keep meaning in focus and avoid designing mechanical, 'grammar-driven' exercises."

5.2 Instructional Materials

Before the materials are presented, it seems necessary to introduce one important definition. In the following sections, the instructional materials will be classified, described, and analysed as certain types of tasks. Underlying all that is the following definition of "task":

A task is a workplan that requires learners to process language pragmatically in order to achieve an outcome that can be evaluated in terms of whether the correct or appropriate propositional content has been conveyed. To this end, it requires them to give primary attention to meaning and to make use of their own linguistic abilities, although the design of the task may predispose them to choose particular forms. A task is intended to result in language use that bears a resemblance, direct or indirect, to the way language is used in the real world. Like other language activities, a task can engage productive or receptive, and oral or written skills, and also various cognitive processes.

Rod Ellis (2003: 16)

In addition to this understanding of what a task is, the literature on Processing Instruction details six guidelines which should be followed when designing Structured Input Activities (see e.g. Farley, 2005). Previous studies testing the merits of Processing Instruction have been rejected by VanPatten and colleagues on the grounds that the instructional

materials were not based on the Processing Instruction guidelines, so for the materials in this study, these guidelines have been kept in mind:

1. Present one thing at a time
2. Keep meaning in focus
3. Move from sentences to connected discourse
4. Use both written and oral input
5. Have the learner do something with the input
6. Keep the learner's processing strategies in mind

The guidelines are applied as faithfully as possible, while keeping the participant sample in mind, i.e. only one thing at a time is presented insofar as the distinction between direction and location in prepositional phrases is boiled down to the distinction between definite articles in either accusative or dative case. Only well-known, common nouns are used, so the participants' knowledge of gender should not be a decisive factor. This could have been simplified further by only using one gender per task, but the tasks generally contain definite articles in all genders and numbers as the students of the sample are familiar with the case markers as such and might have been bored by the monotony of the tasks if there had been even less variety.

Meaning is kept in focus as the dative-accusative distinction in prepositional phrases following German switch prepositions is a 1:1 mapping for the distinction between location and direction. In all of the Structured Input Activities, the participants have to accurately interpret the grammatical case markers in order to derive the correct meaning. There are no alternative cues and the sentences are not ambiguous. Form-function mappings should therefore be straightforward, allowing the application of explicit information as well.

The instructional materials start with sentence-level interpretations and then move on to connected discourse towards the end of the first session. The second session has a more intense training character and its sentence-level Structured Input Activities are designed to reinforce the form-meaning connections. In both sessions, the participants are exposed to both aural and visual input. While the textual input helps with the development of written form-meaning connections and affords the learners more processing time, the presentation of aural input is crucial for the development of real-time sound-meaning connections (Farley, 2005: 78).

Point 5, have the learners do something with the input, refers to the claim that “[l]earners need to be pushed to make decisions based on meaning and form instead of being mere recipients of input” (Farley, 2005: 15). As mentioned before, the participants in this study have to decide whether a switch preposition refers to a static location or to movement towards a destination. The decisions need to be made based on the morphological case marking as semantic content will not provide relevant information. Of course, the default processing strategies, as described in more detail in the previous section in combination with the target feature, are kept in mind at all times.

Processing Instruction, then, comprises three major components: explicit information on the target feature, explicit information on a default processing strategy which might impede the correct processing of the input, and Structured Input Activities in which learners need to correctly process the target feature in order to derive the meaning. The participants therefore receive information on the use of the accusative case, the use of the dative case, and the difference between the two cases when it comes to switch prepositions. They are also reminded that word order is not a reliable indicator of ‘who did what to whom’ in German and are prompted to rely on case marking instead. After each step of the explicit information, participants apply that information and work on the Structured Input Activities. Once they have completed the activity, a slide prompts them to indicate the difficulty level on a scale from 1 (too easy) to 5 (too difficult). Screenshots of the materials can be found throughout the next section and a full list of input items can be found in the Appendix I. The computer-delivered Processing Instruction in this study has been presented via Adobe Captivate 8.0, a programme that allows the creation and presentation of interactive slides through which the learners can work at their own pace. There is also the option to run the programme via a Learning Management System that records the answers of each individual learner and sends it to the lecturer.

Accusative and Dative Recap

After the welcome slide, participants receive a short refresher on the accusative case and then embark on a Structured Input Activity where they have to interpret a German sentence and drag the correct English translation onto it. The input sentences consist of 50% SVO and 50% OVS word orders and are controlled for animacy and plausibility effects, i.e. person A is as likely to do something to person B as person B is likely to do the thing to person A. The only way of determining who performs the action is to interpret

the case of the definite article. Once the participant has dragged the translation (in green) onto the turquoise target area and hits the “Submit” button (as demonstrated on an introduction slide), they receive immediate feedback (“well done” or “incorrect”). As in previous Processing Instruction experiments, no further explanation is given.

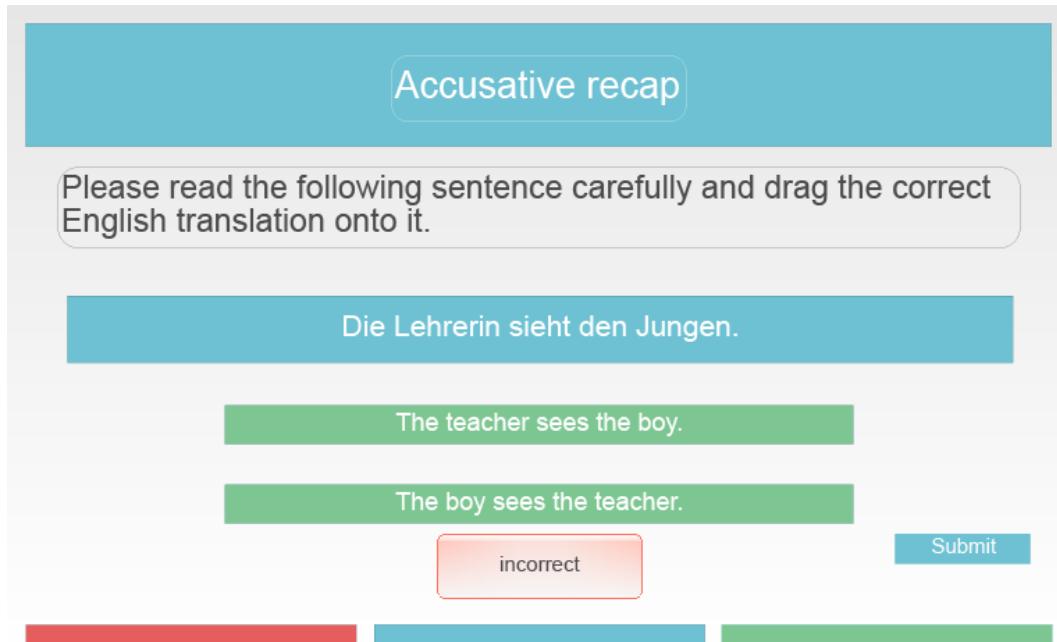


Figure 5.2: Example slide for the accusative Structured Input Activity

Picture-matching activity

After refreshing the accusative and dative cases via the translation tasks, participants are shown a table of switch prepositions and are given an explanation of how they work. They are cautioned not to rely on the semantic content, and especially not on the characteristics of the verb. The explicit information is followed by a picture-matching task in which the participants are asked to click on the picture that corresponds to the meaning of the sentence. The input is structured in a way that puts the prepositional phrase first. Note that in the canonical word order in sentences a) and b), the case of the prepositional phrase is less salient than in sentence c). Learners are more likely to skip the definite article in a) or b) and construct meaning based on the content words *Mann*, *stürzt*, and *Wohnung*.

- | | |
|---|------------------------------|
| a) <i>Der Mann stürzt in der Wohnung.</i> | The man falls in the flat. |
| b) <i>Der Mann stürzt in die Wohnung.</i> | The man falls into the flat. |
| c) <i>In die Wohnung stürzt der Mann.</i> | The man falls into the flat. |

Note also that the picture clearly depicts the boundary, i.e. one man is on the floor within a flat, whereas the alternative depicts a man falling through the door, literally crossing the boundary between corridor and flat. The participants then click on the picture they interpret as corresponding with the sentence and receive immediate feedback, as before. The task comprises 6 sentences with a 50/50 distribution of the two cases.

Picture-matching task

Have a look at the following sentence and decide which of the two pictures matches the meaning of the sentence.

In die Wohnung stürzt der Mann.



well done!

Figure 5.3: Example slide for the picture-matching task

Sentence-level comprehension

This task is also a sentence-level written input task. The participants are asked to read a sentence snippet which could be used as the answer to one of the two following questions:

- | | |
|---------------------------------------|-----------------------------|
| a) <i>Wohin gehen Peter und Anna?</i> | Where do Peter and Anna go? |
| b) <i>Wo sind Peter und Anna?</i> | Where are Peter and Anna? |

They have to recognise the difference between Peter and Anna being in a location and Peter and Anna going somewhere. Again, no additional information is provided and the prepositional phrase is isolated in a way that is supposed to make the case marking on the definite article more salient. As Martin & Ellis (2012: 404) note: “the phrase is the basic level of language representation at which form and meaning meet with greatest reliability”. But isolating the phrases is not only salient, it is also authentic: Someone who is asked either of the two questions in a native context will reply with the prepositional

phrase rather than the complete sentence. The only slightly artificial part of this task is that the preposition and the definite article for *in + dem* would most likely be contracted to “*im*”. This is only applicable to very few instances in written German though (*im* and *am*) and if the definite article is stressed (in THAT park, not a different one), then it is perfectly normal to not contract *in + dem*. The participants receive immediate feedback after dragging the sentence snippet onto the correct target area. After completing the task (9 slides), they are prompted to rate the difficulty.

Sentence-level comprehension

Read the following sentence snippet and decide which of the questions below it answers:

Wohin gehen Peter und Anna?

Wo sind Peter und Anna?

... in dem Park.

incorrect

Submit

Figure 5.4: Example slide for the sentence-level comprehension task

Sentence-level listening comprehension

This task is a sentence-level listening comprehension task, in keeping with the guidelines that recommend the use of written as well as aural input. The participants click on the ‘Play’ button to listen to a sentence snippet. Then they click on the question that the snippet answers. If they hear “*in der Bibliothek*” (in the library), they should choose “*Wo bist du?*”, if they hear “*in die Bibliothek*” (to the library), they should choose “*Wohin gehst du?*”. The sentence snippets have been spoken with natural intonation by the researcher and recorded on an iPad with an app called AudioMemos. The participants hear 10 sentence snippets and choose the corresponding question. The locations/directions and questions change from slide to slide. 50% of the sentences indicate direction, 50%

indicate location. The participants can listen to the recording twice, if necessary, but the third click on the Play button results in the presentation moving on (=incorrect).

Listening comprehension

Listen to the recording and decide which question the phrase you hear answers:

Play

Wo bist du?

Wohin gehst du?

Figure 5.5: Example slide for the sentence-level listening comprehension

Discourse-level interpretation

The last task of the first session is a discourse-level interpretation task. The original interview with Edward Glaeser on cities making us more human (Mensel, 2011) has been shortened and simplified for this task. The participants read one paragraph per slide, three slides in total, for this task. After each paragraph, they are prompted to answer a multiple-choice question. Translations are provided for potentially problematic words. The participants just have to roll the mouse pointer over a word they do not understand and the English translation is displayed. It should be noted that this task is not a Structured Input Activity per se as the input is not rigorously manipulated to draw attention to the target feature by placing it at the beginning of each sentence. From a Processing Instruction point of view, this might be considered a deficit, but it increases the authenticity of the text and avoids teaching the participants pragmatically incorrect choices. Using connected sentences, it would make no sense to start each sentence with the emphasis on the prepositional phrase, stressing that people drive their car to work, not elsewhere, then stressing that they drive to the supermarket, not elsewhere, etc. The focus of this activity is on the distinction between things happening within the city (*in der Stadt*) versus

on the way into the city (*in die Stadt*). All instances are balanced for probability so that the verb cannot be used as a cue to determine the location or destination.

Discourse-level interpretation

In 2011, Harvard professor and economist Edward Glaeser gave an interview titled “Städte machen uns menschlicher”. Below you can find a summary of his main arguments. Read the text and answer the questions that follow.

„Städte machen uns menschlicher“

- 1) Städte schützen die Umwelt, gerade weil sie Menschen von ihr fernhalten.

Menschen, die auf dem Land leben, fahren mit dem Auto in die Stadt, zur Arbeit, zum Einkaufen, und verbrauchen dadurch mehr Benzin als Leute, die bereits in der Stadt leben und alles zu Fuß oder mit öffentlichen Verkehrsmitteln erreichen können. Städter, die in Wohnungen leben, verbrauchen weniger Elektrizität oder Energie zum Heizen als Leute, die auf das Land ziehen, um in einem eigenen Haus zu wohnen. In Städten wird Raum oft geteilt, zum Beispiel in Restaurants oder Cafés.

Next

Figure 5.6: Example slide for the discourse-level interpretation task

After the first paragraph, as shown above, the questions are as follows:

- a) Where do people drive cars? b) Where do people save energy?

O in the city O to the city O in the city O in the country

Questions like a) are designed to elicit an interpretation of the case. The text states that people living in the country use cars to drive to the city. Both answers make sense as there are cars inside and outside of cities, but only “to the city” is correct in this context. Questions like b) are content questions to check general understanding. All in all, there are 12 questions, five target questions and seven content questions. The task ends with two Likert-type ratings prompting the participants to indicate whether they think this description of urban life is accurate and whether they consider themselves to be city dwellers. This is an attempt to make the learners relate to the materials on a more affective level. The participants do not receive feedback for this task. At the end of the session, the last slide displays the overall accuracy on the attempted tasks and the participants have the chance to review all answers to see where they have been right or wrong.

The second session, following one week after the first, is more intense and has more of a training character. For the picture-matching task, both versions of each item are used. This means that the sentences for both pictures appear in random order, so the participants are presented with twice as many slides. In total, they work through 12 slides for the picture-matching task. For the sentence-level comprehension with written snippets, the 9 sentences are even further simplified (see Appendix I).

Location versus direction

One of the new tasks in the second session is a written sentence-level task with stative and directive verbs. The potentially difficult differentiation between *legen* (to lay) and *liegen* (to lie) is facilitated by the use of the interrogative pronouns *wo* (where) and *wohin* (where to). The participants drag the correct answer onto the question and receive immediate feedback. After completing all 15 slides of the task (7 of which required the accusative), they rate the difficulty.

The screenshot shows a digital slide with a light blue header bar containing the title 'Location versus direction'. Below the header, a white text area contains the instruction: 'Please read the following question and click on the correct answer.' A teal-colored rectangular box contains the question 'Wohin stelle ich meine Tasche?'. Below the question are two teal-colored rectangular buttons with the options 'neben das Bett' and 'neben dem Bett'. At the bottom of the slide, there is a decorative footer consisting of three horizontal bars in red, teal, and green.

Figure 5.7: Example slide for the sentence-level task with stative and directive verbs

Aural sentence puzzle

The last task is an aural sentence-level task where participants click the 'Play' button to listen to the beginning of a sentence, e.g. "Auf dem Sofa", and then choose the correct

sentence ending. This task further trains the distinction between stative verbs and directive verbs. The sound bites are spoken with natural stress by the researcher, using the iPad app AudioMemos. The participants work through 18 slides (50% dative), receiving immediate feedback on their choices, and then rate the difficulty of the task. As in all of the aural-input tasks, participants have the chance to listen to each recording twice. If they click on the 'Play' button a third time, the presentation moves on and the slide is counted as incorrect. The complete list of input items can be found in Appendix I.

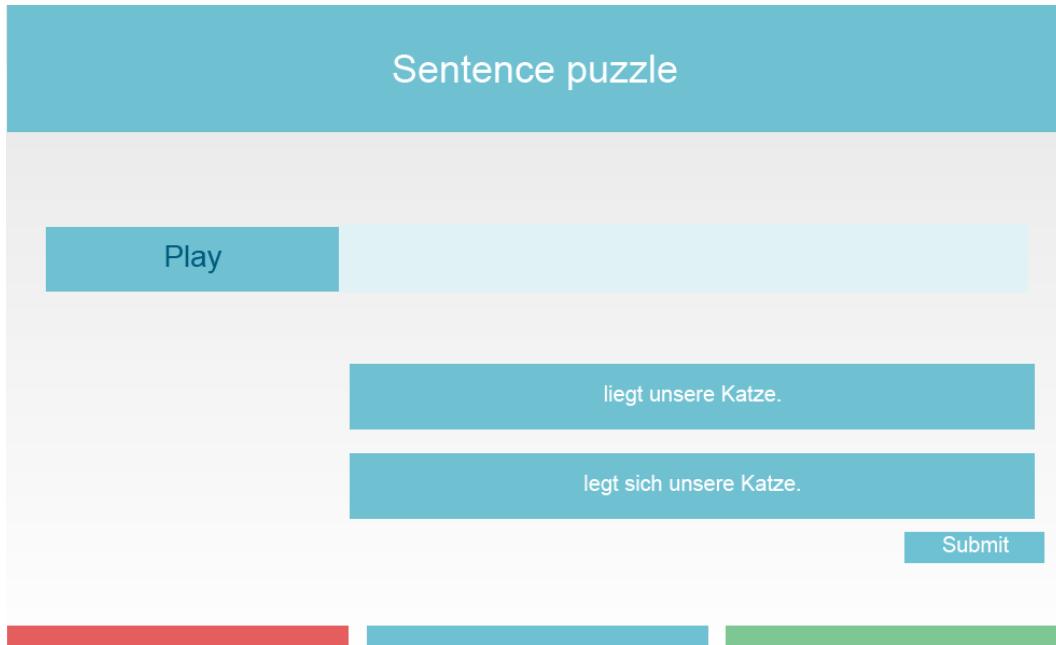


Figure 5.8: Example slide for the aural sentence-level task with stative and directive verbs

5.3 Assessment materials

Picture description

Half of the test materials have been used in both the pilot study and the main study. One of those tasks is a picture-description task in which participants see a picture (shown in Figure 5.9 below) and have to describe it in at least five sentences. The sentence beginnings are given (e.g. *Eine Frau...* - A woman...) in order to make the responses comparable and to ensure that the participants do not avoid grammatical constructions (e.g. dative forms) they are not comfortable with by choosing to describe only locations or only movements. The participants are also given some of the key vocabulary, including nouns and their gender, the prepositions, and some verbs to describe the events taking place in the picture. The vocabulary is provided to make sure that differences in

performance are not based on differences in vocabulary knowledge and to increase the likelihood that case markings are chosen based on the correct gender of the noun – which is crucial because some of the definite articles share forms across different cases, as detailed previously. All things considered, the task is pre-structured in a way that makes it a sentence-level task rather than a discourse-level task. Consequently, a genuine discourse-level writing task has been designed for the main study, as will be explained later. The picture-description task is the same in the pre- and post-tests and in both the pilot and the main study.

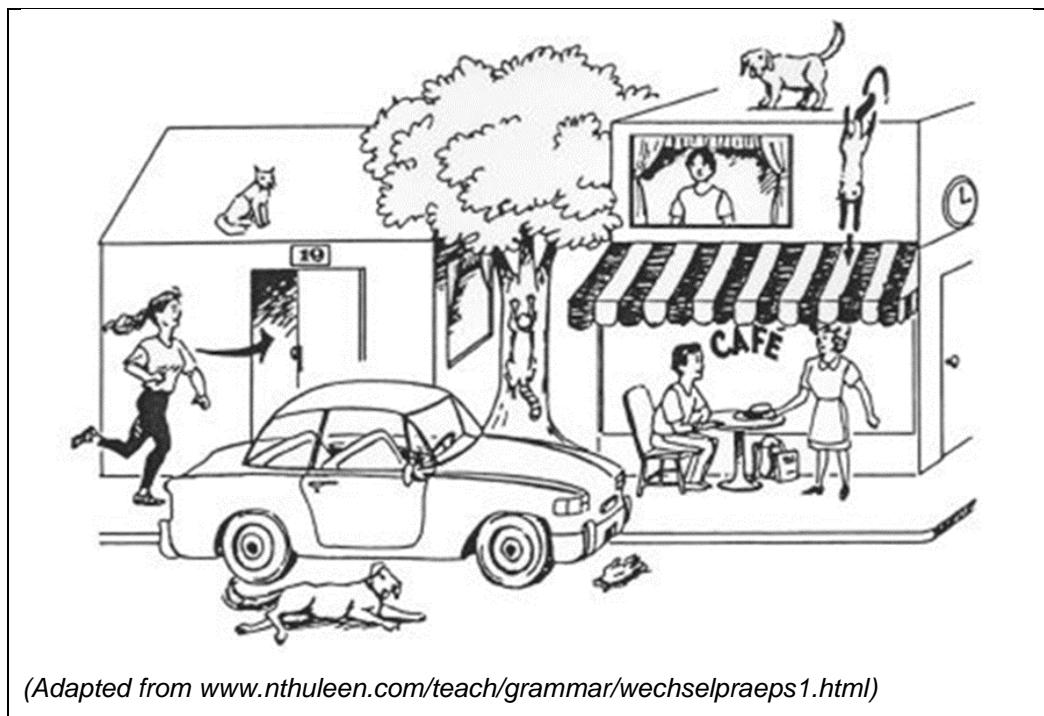


Figure 5.9: Picture prompt for the picture description task

Fill-in-the-blanks

Another task that has been used for both studies is a sentence-level written production task in which the participants are asked to supply the missing endings of the definite articles. As Domínguez Vázquez (2005) remarks in her overview of exercise types for the training of prepositional phrases, this task is very common in classroom-based instruction and is often used for grammar training. The task in this study focuses on the correct production of definite articles in either dative or accusative case:

- a) *Bei schlechtem Wetter verbringen wir viel Zeit in d__ Geschäften.*
[If the weather is bad, we spend a lot of time in the shops.]

b) *Bei gutem Wetter verbringen wir viel Zeit in d__ Parks.*
[If the weather is good, we spend a lot of time in the parks.]

Figure 5.10: Example sentences from different test versions

Sentence a) is an example of a pre-test item, whereas sentence b) is an example of a post-test item. Note that the overall sentence structure is the same and that the nouns chosen for this exercise are carefully controlled for gender and number. Out of the twenty blanks in total, ten require dative case and ten require accusative case. Across all blanks, there are five masculine, feminine, and neuter nouns, respectively, and five nouns in the plural. The participants can tell the gender and number from the colours used for the nouns, if they consult the legend at the top of the page. This information is provided to make sure that case marking is chosen based on the right gender and number. Any errors, in theory, should then be due to the use of the wrong case. The task is the same on the post-test and the delayed post-test as there are 15 weeks of term between the two assessments, which should be enough time to counter any potential memory effects.

Interpretation/ translation

For the pilot study with secondary-school learners, traditional Processing Instruction interpretation tasks are used, much in the style of the accusative and dative case activities described in the previous section. The participants have to decide which one of the two English translations is the equivalent to the German sentence. In order to control for guessing, a “not sure” option is offered:

- Die Tochter findet der Vater.*
- a) The daughter finds the father.
 - b) The father finds the daughter.
 - c) Not sure.

Figure 5.11: Example of Structured Input interpretation task

While this test elicits interesting data from secondary-school participants, it became clear during the participant screening for the main study that this kind of task would be too easy for a university population. The task has therefore been redesigned as a translation task.

The main idea, namely that participants should interpret the subject and object roles relying on case rather than context, has remained. The sentences are presented in a randomised mix of 6 SVO sentences and 7 OVS sentences. 5 sentences contain accusative objects and 7 sentences contain dative objects. One sentence contains no object. The canonical SVO sentences serve as distractor items and are not counted as test items as both default processing strategies and attention to case marking could have provided a correct interpretation. For the 7 OVS sentences, however, a reliance on the First Noun Principle, i.e. a reliance on word order rather than case marking, leads to a false interpretation of subject and object roles.

As mentioned before, the complexity of the sentences has been increased after a couple of pre-test administrations because some of the participants learnt from the Processing Instruction test materials and went back to revise their previous answers. All of the participants that are included in the study completed the complex version of the task. For this version, additional information and deliberately difficult vocabulary has been added to the sentences in order to make the task less drill-like and more varied:

Die Tochter findet der Vater nach dem Fußballspiel im Stadion nicht wieder.

[The father cannot find the daughter again after the football match in the stadium.]

Figure 5.12: Example sentence of main study translation task

This is not a Structured-Input-Activity type of task anymore because the design does not adhere to the Processing Instruction guidelines. If participants can demonstrate instructional gains from pre- to post-tests, however, this can count as support for the effectiveness of Processing Instruction. In previous research, the use of a Processing Instruction interpretation task was a good way of showing that those participants who had received Processing Instruction were able to do something that participants who had received traditional output-oriented instruction could not do: forego the default processing strategy. In this study, there is no comparison between instructional groups, so the exclusion of a genuine Processing Instruction interpretation task poses no problem. The complete pre- and post-tests can be found in the Appendix J.

Essay

Another change from pilot study to main study concerns the addition of a discourse-level production task (essay task). The participants in the main study are asked to write a short essay of at least 250 words of German on the topic of their favourite holiday destination (pre-test) and their ideal university experience (post-tests). Additional questions prompt the participants to describe the place/university and its location, how to get there, the characteristics, and potential activities. This is supposed to elicit plenty of prepositions, both expressing location and direction. During the material-testing phase, both essay questions elicited about the same amount of switch prepositions from two advanced learners of German and two native speakers. It would have been possible to answer all questions in fewer words, but the word minimum has been added to ensure that the participants experience a genuine discourse-level task of sufficient complexity. This kind of relatively free writing is another common exercise type that can be found in many foreign language classrooms (Domínguez Vázquez, 2005: 64) and therefore has ecological validity.

It is generally posited that “because writing involves a complex interaction between a wide range of different processes, it places extremely high demands on the limited capacity of working memory” (Galbraith, 2009: 7). What is written down, on the other hand, does not have to be kept in an activated state as the writer can go back to previous output by skimming the text. Additionally, the writing process seems to depend on the expertise of the writer: While novice writers tend to reproduce content stored in Long-Term Memory, expert writers tend to adapt the content and present it with their communicative goals and the target audience in mind (Galbraith, 2009: 9). It is therefore possible that the participants’ writing will not be affected equally by Working Memory Capacity limitations, especially if differences in the participants’ proficiency level additionally affect the availability of processing resources.

5.4 Pilot Study data collection

The aim of the remainder of this chapter is to make the data collection procedures as transparent as possible. In this section, the pilot study data collection will be described, then the findings of the pilot study will be presented and summarised with regards to the changes in the materials that followed from the pilot study insights. In the following

section, the main study will be described in terms of the methodology, the data collection procedure, and the participant sample. Section 5.6 then summarises the steps that have been taken to maximise the validity and reliability of the data collection instruments and procedures. Most of the materials chosen for this project have undergone a testing process that spanned several years. After the initial Material Testing Session (see Appendix K) within the Department of Applied Linguistics' Postgraduate Reading Group in March 2014, the materials were amended based on the suggestions of fellow PhD students and staff. Most of the questionnaires were abridged and problematic wording of questionnaire items was either improved or the items in question were discarded. Considering the number of questionnaires to be administered per participant, it seemed best to have a good variety of questions targeting the construct components for the individual variables (e.g. impulsivity and sociability as components of extraversion) rather than have several repetitive questions that address the same construct and artificially conflate internal validity. The next step then was to test the materials in a pilot study.

To this end, the researcher visited a mixed sixth form secondary school in South-East London several times over the course of a month to establish links with teachers and students of German and to observe German lessons. Based on the proficiency level of the students and the school's curriculum for German, an appropriate target feature was identified: German switch prepositions. The researcher then designed the Processing Instruction materials presented above and prepared them to be computer-delivered via Adobe Captivate. All the necessary licences were acquired and additional Ethical Approval was sought to include underage participants. In the end, two female students in year 12 volunteered for the pilot study, which was conducted in early 2015. Both students were preparing for the International Baccalaureate at the time of data collection. They very generously offered to work through the various parts of the experiment during their free periods and without any compensation. The data collection procedure was scheduled as in Table 5.3 below, all treatments were administered individually, but the succession and timing of the sessions were kept constant. The Working Memory Capacity and Phonological Short-Term Memory tests were administered at the beginning of the pre-test sessions because they were supposed to be the most exhausting tasks. The learner variable questionnaires were used as distractor items for the pre-tests. The two computer-delivered Processing Instruction sessions lasted roughly 50 minutes each and the pre- and post-tests could comfortably be completed within ten minutes as they were sentence-level structured tasks.

The materials were such as described in Chapters 3, 4, and 5 – with the following exceptions:

- The Language Analytical Aptitude test included translation sentences in SVO word order instead of OVS word order.
- There was no discourse-level production test.
- There was no translation task, the interpretation test was a classical Processing Instruction task as used in the instructional materials, see accusative and dative recap in section 5.2.
- Due to time constraints, the pre-test sessions had to be accommodated within 60 minutes, which did not allow for breaks. Instead, the questionnaires served as distractors for the pre-test tasks.
- Due to the fact that both participants were preparing for their IB exams, it was not possible to administer a delayed post-test.

Day 1	Day 3	Day 10	Day 17
Learner variables and pre-test	Learner variables and pre-test	Treatment session 1	Treatment session 2
Digit Span Task Interpretation task	Reading Span task Fill-in-the-blanks task	<i>Processing Instruction</i>	<i>Processing Instruction</i> Post-test
Demographic info	Language analytical ability		
Attitude/Motivation questionnaire	Personality questionnaire		
Picture-description task			
Grammatical sensitivity			
60 minutes	60 minutes	50 minutes	60 minutes

Table 5.3: Schedule of the data collection for the pilot study

Findings

Both of the participants find it important to speak a foreign language, but their motivation is not instrumental, i.e. none of them thinks that German is particularly prestigious as far as foreign languages go. A could have scored higher on the interest items, but she finds studying languages “annoying”. B, on the other hand, finds learning German “great”, but she does not invest as much time or effort as A. This illustrates well that interest in language learning does not translate into language learning success or even heightened motivation – unless students consistently put in goal-related effort. The class anxiety scores might be related to this: The more academically-minded, hardworking A reports lower class anxiety than B, but higher use anxiety in unstructured environments. That A scores slightly higher overall in terms of interest and motivation reflects the teacher’s observations.

	A	B
Interest in Foreign Languages	59%	83%
Motivational Intensity	83%	42%
Instrumental Orientation	59%	50%
Class Anxiety	9%	33%
German Use Anxiety	33%	9%
Attitudes towards learning German	67%	75%

Table 5.4: AMTB questionnaire results in percentages

Looking at the Working Memory Capacity scores, the composite scores of all correctly recalled letters from all sets are very similar and the Backward Digit Span maximum scores are the same. The difference between the two participants is that B’s Forward Digit Span, presumably a measure of Phonological Short-Term Memory, is higher and that she scores either close to her maximum Reading Span set size, or very low. A, on the other hand, does not reach her maximum span often, but she scores more consistently. She also takes longer to respond and uses the blank space when she cannot remember a letter, which shows that she is very aware of how much or little she recalls. Both students struggle a little with the judgement of the sentences in the training trials as they are not sure about what differentiates a sentence that makes sense from one that does not. This confusion does not last though and both are working swiftly through the test trials.

	A	B
Forward Digit Span (max)	6	9
Backward Digit Span (max)	7	7
Reading Span (composite)	58	55
Reading Span (range)	3 – 6	2 – 6

Table 5.5: Working Memory and Phonological Short-Term Memory Capacity scores

A scores higher on the extraversion scale (56%), which reflects both the descriptions of the teachers and the observations of the researcher. She is more willing to speak German outside the classroom and initiates more conversations, while B is quiet, but friendly when approached. B keeps a distance to other people and to the materials, never fully engaging or interacting with the environment. Her introversion (6% extraversion) does not express itself in slower response times, however.

A, who has more language learning experience and has learnt English as a second language, scores slightly higher on the grammatical sensitivity test, but only by one item. She does, however, solve all the artificial language problems and provides correct translations. Based on her notes, one can tell that she clearly noticed the singular-plural distinction. Unfortunately, she interpreted the subject-object marking as indications of positions within the sentence as the subject always came in first place in the translation examples. Which means that she scores 100%, but she achieves this based on false assumptions. The materials had to be changed accordingly to make this impossible for the main study. B scores a rather high 67%, but she does not give any indication of being aware that Pip marks subject and object roles and she does not translate the sentences correctly from English into Pip.

	A	B
Grammatical sensitivity	88%	63%
Language analytical ability	100%	67%

Table 5.6: Grammatical sensitivity and language analytical ability scores

On the assessment tasks, both participants start with relatively high interpretation scores, but provide very few correct forms in the picture description task. For the more controlled fill-in-the-blanks task, they provide correct forms roughly half of the time. While A scores slightly lower on the interpretation post-test, this is only due to the fact that she chooses the “not sure” option more often than in the pre-test. B also chooses the “not sure” option

more often in the post-test, but also shows improved scores on the sentences she is sure about. Apparently, the Processing Instruction treatment has made them question their default processing strategies. A produces the same amount of correct switch prepositions on both the pre- and the post-test of the picture description task, but B improves on this task and produces as many correct switch prepositions as A on the post-test. Both improve on the fill-in-the-blanks task. Unfortunately, there has been no opportunity to administer a delayed post-test, so it is hard to tell whether these gains could have been maintained in the longer run. For the main study, a delayed post-test after a significantly longer period of time seems desirable.

	A	B
Processing Instruction interpretation	-13%	+44%
Picture Description	+/-	+200%
Fill-in-the-blanks task	+71%	+100%

Table 5.7: Instructional gains from pre-test to immediate post-test

As can be seen in Table 5.8 below, Working Memory Capacity is the only individual difference measure on which B scores higher than A, but she also improves more from pre-test to post-test, which is in line with previous findings in Processing Instruction studies describing that the weaker students gain more from PI than the initially stronger performers (Lee & McNulty, 2013). Overall, both participants find the second session more difficult than the first session and both participants tend to rate the individual tasks as more difficult when the feedback they receive is mixed. A either produces the correct interpretations immediately or needs four to five trials until she consistently interprets the input correctly, which shows a clear learning effect. B, on the other hand, either interprets the input correctly from the start or shows an overall more mixed pattern. This is also reflected in her Reading Span scores. The on-task performance data can be found in Appendix L.

While B has the higher Forward Digit Span scores, which are supposedly measuring the Phonological Short-Term Memory component of the memory system, she struggles more with the aural input during instruction. A scores consistently and accurately on the same tasks and later admits that this might be due to watching German soap operas in the evenings. This points to the fact that exposure to and processing of German input is important for acquisition and future processing, as predicted by the MOGUL framework. Results might have shown a different picture for a delayed post-test.

	A	B
Age	17	18
Languages	Albanian English German	English German
Exposure to German	6 years (school) TV + literature	6 years (school)
Interest/Motivation	+++	++
Personality	Extra	Intro
Grammatical sensitivity	++	+
Language analytical ability	++	+
Working Memory Capacity (Max)	++	+++
Working Memory Capacity (Composite)	++	++
Instructional Gains	+	++

Table 5.8: Overview of variables and instructional gains for the pilot study

Implications for the main study

For the main study, some of the materials will have to be amended, e.g. the word order in the language analytical ability test should reflect the use of subject-object marking and display a more flexible word order that actually tests whether the participants have detected this feature and are able to apply the rules that they have abstracted. It is possible that the Processing Instruction treatment de-automatises default processing strategies to the point where students are hesitant about their choices, at least initially. A “not sure” option then is the easy way out, but does not shed light on potential instructional gains. It is also possible that some of the changes would have taken longer to show in production measures, which means that a delayed post-test or repeated exposure to Processing Instruction might have yielded different results. An additional discourse-level production test might bring out the effects of individual differences in Working Memory Capacity more clearly. What has become clear from the pilot study is that looking at the scores, both participants seem to be fairly similar. Looking at the scoring, however,

reveals significant differences in terms of consistency and trajectories. This adds weight to the argument that traditional group comparison studies investigating the effects of individual differences on the outcomes of Processing Instruction are unable to capture the interplay of the learner variables that each participant brings to the task and that a lot of potentially interesting information is lost or discarded in traditional quasi-experimental studies.

5.5 Main Study data collection

Processing Instruction research has initially focused on quasi-experimental studies in order to compare this pedagogical intervention to other forms of instruction. Even in the studies on individual differences, group-comparison designs have been employed, even though the authors routinely acknowledge the limitations of this design (Lee & McNulty, 2013; VanPatten et al., 2013). The so far only study on Processing Instruction and Working Memory effects is no exception. To restate the problem: Sample sizes in Second Language Acquisition research (as opposed to psychometric or medical research programmes) are severely limited, making the use of intact classes or a median split between high-level and low-level Working Memory span scores necessary. This has at least two negative consequences: The small size leads to very low statistical power and the between-group differences are not necessarily greater than the within-group differences. The present study therefore appreciates the diversity in any given participant sample and takes one step back. Instead of comparing a low-span and a high-span group and their respective learning gains, it explores a selection of Memory Capacity measures as well as potential mediating factors in order to look at individual differences at the level of the individual learner. In a mixed-methods approach, qualitative questionnaire data on learner variables are combined with a more traditional pre- and post-test experimental design. The advantage of this case study is that several variables can be explored qualitatively and in relationship with each other.

As described previously, the main aim of this study is to explore whether Processing Instruction is equally beneficial for learners at different points of the Working Memory Capacity spectrum. To this end, the participants' Working Memory Capacity is measured with two different instruments. Different scoring techniques are employed to find out which one is most sensitive to individual differences and/or makes optimal use of the available information. The learning outcomes are measured before, during, and after instruction.

Looking at both product- and process-oriented data yields insights not only in terms of the level of achievement, but also regarding the learning trajectories. Data on other, potentially moderating, variables are collected via questionnaires and supplement the Working Memory and Phonological Short-Term Memory Capacity scores. The task conditions have been analysed and the participants additionally rate the perceived difficulty level of each task after completion. The automated span tasks allow for the collection of complementary observational data. All in all, several layers and sources of data are collected with the aim of abstracting the individual learner profile for each participant and the impact it has on instructional as well as assessment tasks.

The main questions to be addressed in this study can be restated as follows:

Research questions:

1. Do learners benefit from Processing Instruction in equal measures, regardless of their Working Memory Capacity?
2. Are there other learner variables that potentially mediate Working Memory Capacity limitations and thereby affect the learning outcomes?
3. Are there participant-treatment interactions?

Based on findings from previous Processing Instruction and individual differences research, the following hypotheses are put forward:

Hypotheses:

- a)** Processing Instruction is maximally reduced in complexity, yet there usually are individual differences in the post-test scores. Even if differences in Working Memory Capacity will not result in performance differences during instruction, they will result in differences in post-test scores.
- b)** There might be other variables that compensate for lower Working Memory Capacity scores. If any of the measured constructs have an impact on post-test scores, then these should show up in a comparison of the individual learner profiles.
- c)** It is conceivable that different individuals react to and interact with the instructional and experimental materials in different ways. This could influence the learning outcomes. Evidence for this might be found in the participants' approach to the less controlled tasks, in the way participants work with the materials, and in the observational data the researcher collects.

For the reader's convenience, the table below summarises which research question is addressed using the instruments described with the aim of measuring the respective key variable. A more in-depth description and justification of measures and procedures has been provided in Chapters 3, 4, and 5.

Research question	Data collection tools	Variables
1) Do learners benefit from Processing Instruction in equal measures, regardless of their Working Memory Capacity?	<i>Automated Digit Span</i>	Working Memory Capacity
	<i>Automated Reading Span</i>	Phonological Short-Term Memory Capacity
	Online scoring during instruction	Learning rate (process)
2) Are there other learner variables that affect the learning outcomes and potentially mediate Working Memory Capacity limitations?	Pre- and Post-tests: Fill-in-the-blanks task Picture description task Translation task Essay writing task	Instructional gains on tasks of different complexity levels (products)
	<i>AMTB</i> (abridged and modified for L2 German)	Interest L2 use anxiety & class anxiety Motivational intensity Instrumental orientation Attitude towards the language
	<i>Eysenck Personality Inventory</i> (abridged and re-worded)	Extraversion
	<i>Words in Sentences</i> section of the MLAT	Grammatical sensitivity
3) Are there participant-treatment interactions?	Artificial Language section from the Oxford University Modern Language Test	Language analytical ability
	Difficulty ratings Interactions Behavioural data	Potentially all of the above

Table 5.9: Research questions and operationalisation of variables for main study data collection

Experimental Design

To explore whether Processing Instruction is equally beneficial for all participants, regardless of their Working Memory Capacity scores, a pre-/post-test design has been chosen. The first post-test took place immediately after instruction and the second post-test was administered 15 weeks after instruction. The pre-test also had a screening function: Only participants who scored below 65% on at least one of the four pre-test measures were eligible for participation. Once the final participant pool was established, Working Memory Capacity, and one additional memory component, Phonological Short-Term Memory, were measured prior to instruction. Hypothesising that Processing Instruction would not elicit Working Memory effects, but the more challenging assessment tasks would, the participants' on-task performance was monitored during the two instructional sessions and the participants were prompted to rate the difficulty of each task once they had completed it. This made an analysis of both the product- and the process-perspective possible.

To answer the second research question, i.e. whether the effects of Working Memory Capacity limitations could potentially be mediated by other variables, participants also completed the questionnaires and tests presented in Chapter 4. Table 5.10 below summarises the experimental design, with learner variables in yellow, pre- and post-tests involving the target feature in blue, and on-task performance during the instructional treatment in green boxes. The exact timeline of the experiment will be presented in the next section.

The third question will be addressed in the results section, when it is clearer from the analyses which learner variables (shown in yellow boxes in Table 5.10 below) seem to have influenced the behaviour and ultimately the learning outcomes of the participants (for the proficiency measures, see the blue and green boxes). Individual learner profiles will be analysed to find out whether a particularly successful pattern of variables emerges, or whether different predispositions affect different tasks to different extents.

PRE-TEST	MEMORY TESTS	INSTRUCTION	POST-TEST 1	POST-TEST 2
Day 1	Day 8	Days 10 and 17	Day 17	Day 122
L2 Proficiency: target item <i>Translation task</i> <i>Fill-in-the-blanks task</i> <i>Picture description task</i> <i>Essay task</i>	Working Memory Capacity <i>Automated Reading Span task</i>	On-task performance <i>Recorded via Adobe Captivate</i>	L2 Proficiency: target item <i>Translation task</i> <i>Fill-in-the-blanks task</i> <i>Picture description task</i> <i>Essay task</i>	L2 Proficiency: target item <i>Translation task</i> <i>Fill-in-the-blanks task</i> <i>Picture description task</i> <i>Essay task</i>
Task difficulty <i>Ratings (Likert Scale)</i>	Phonological Short-Term Memory Capacity <i>Automated Digit Span task</i>	Task difficulty <i>Ratings (Likert Scale)</i>	Task difficulty <i>Ratings (Likert Scale)</i>	Task difficulty <i>Ratings (Likert Scale)</i>
Learner background <i>Questionnaire</i>	Behaviour <i>Observational data</i>	Task novelty <i>Improvement from first to second session</i>		Debriefing <i>Exit questionnaire</i>
Learner variables <i>Attitude/Motivation Test Battery</i> <i>Eysenck Personality Inventory</i> <i>Grammatical sensitivity</i> <i>Language analytical ability</i>				

Table 5.10: Experimental design of the main study

Timeline of the data collection

The schedule of the data collection can be found below. While the data were collected individually on different days, the researcher made every effort to keep the distance between the respective data collection sessions as stable as possible across participants. Weekly re-occurring free periods in the students' timetables made this possible for almost all sessions. There were minor delays in the administration of two delayed post-tests as the participants were busy in their final week of university. In these cases, the delayed post-tests were taken two to five days later than the other delayed post-tests, respectively. Considering that the delayed post-tests took place 15 weeks after instruction, this slight delay did not seem problematic.

On day 1, potential participants received the Participant Information Form, gave their informed consent, and had the opportunity to ask the researcher for further information. Then they completed the demographic background form, the personality questionnaire, and the essay task. After a break, they worked through the second block, consisting of several sub-sections of the Attitude and Motivation Test Battery (abridged and adapted for German), the picture description, and the translation task. After another break, the last block comprised the fill-in-the-blanks task, the grammatical sensitivity test, and the language analytical ability test.

The questionnaires were used as distractor items for the pre-tests as all the preliminary data had to be collected in one session and the participants were not supposed to guess the target feature from the materials. The succession of tests was the same for all participants: First, they wrote the essay, then they described the picture, followed by the translation task, and finally they provided forms for the fill-in-the-blanks task. The progression was from the least controlled task to the most controlled task in order to keep priming effects as small as possible. The order of administration was the same for all participants and across all testing sessions.

For the post-tests, on Days 17 and 122+ respectively, breaks were enforced to counteract priming effects in lieu of distractor tasks. After completing the last task on the day of the delayed post-test, participants received an exit questionnaire and were asked to report any formal instruction received or independent revision undertaken relating to the target feature. For the full questionnaire, see Appendix M. Table 5.11 below summarises the data collection procedure:

Day 1	Day 8	Day 10	Day 17	Day 122+
Background data	Working Memory	Processing Instruction session 1	Processing Instruction session 2	Delayed Post-test
Demographic info Personality questionnaire Essay task	Digit span	50 minutes of computer-delivered Processing Instruction	50 minutes of computer-delivered Processing Instruction	40-minute post-test Exit questionnaire
Break (15 minutes)	Break (15 minutes)		Break (15 minutes)	
AMTB questionnaire Picture-description task Translation task Break (15 minutes)	Reading span		40-minute post-test	
Grammatical sensitivity Language analytical ability Fill-in-the-blanks task				
Approx. 120 minutes	60 minutes	50 minutes	105 minutes	45 minutes

Table 5.11: Timeline of data collection (main study)

Participant sample

Participants for this study were recruited by emailing a flyer (see Appendix N) to lecturers of German based at universities in the UK. The flyer was distributed via internal mailing lists and interested students were asked to contact the researcher who, in turn, provided more information regarding the time and effort required for participation as well as on the participant screening process. The researcher then conducted the pre-tests with 14 participants over the course of three weeks. All of the participants in the initial pool were briefed on the voluntary nature of their participation, the characteristics and purposes of the study, as well as the researcher's efforts to ensure the anonymity, confidentiality, and safekeeping of their data. All participants gave informed consent. None of them reported any psychological or physical impairments that would have made them particularly vulnerable participants. Based on a cut-off score of 60% on at least one of the pre-test measures, suitable participants were then contacted via email by the researcher and individual appointments were scheduled. The final sample comprised four participants, two male and two female, sourced from a high-ranking London-based public research university. Their age ranged from 18 to 29 years (mean age: 21.5 years). Two participants were in their first year, two in their third year. Further demographic information can be found in Table 5.12 below.

The students were all enrolled in undergraduate-level German studies and therefore no naïve participants as such. This has been taken into consideration in the choice and administration of the test materials and is detailed in the relevant section. All of the participants indicated that they were comfortable with the use of computers and/or tablets and that they have been using them for more than ten years. This was important as a substantial portion of the experiment was computer-based. None of the students had been taught by the researcher previously and it was stressed that any potential participation and/or withdrawal would have no consequences for their academic progression. Participants who took part in the full experiment received a financial compensation of £130 at the end of the experiment.

	MgrYyVvU	YKX5Bwvj	74YuGCRD	nLQ6aVus
Sex	M	F	M	F
Age	29	18	21	18
Country of birth	Australia	UK	UK	Libya
Study programme	German & History of Art	Russian & German	French & German	German
Year of study	3rd year	1st year	3rd year	1st year
L1	English	English, German	English	Hindi
Home language	English	English, German	English	Hindi, English
Can speak	English, German	English, German, Russian, Italian	English, French, German	Hindi, English, German
Can read	English, German, French	English, German, Russian, Italian	English, French, German	English, German
Can write	English, German	English, German, Russian, Italian	English, French, German	English, German
Years of exposure	5	18	11	7
Previous grade	55%	A	60%	A

Table 5.12: Demographic data of the participant sample (main study)

The data were collected close to the university, but off-campus, for the participants' convenience. Apart from the potential ethical issues that apply to any research that collects personal data from individuals in an educational setting (see Ethical Research Form in Appendix O), no additional ethical implications could be identified. One participant had been diagnosed with dyslexia, but the researcher decided to include him for the following reasons: Firstly, it is estimated that at least 10% of the general population are affected by this condition, so there is a fair chance that at least one dyslexic person ends up in the average German classroom. Including a dyslexic participant therefore reflects an authentic sample. Secondly, the scores of this participant, especially the Working Memory Capacity scores, did not seem to be affected by this predisposition, contrary to expectations. This will be elaborated on further in the discussion of the results.

One reservation that must be addressed is that none of the participants was exposed to German case markings for the first time in this experiment. All of the participants had

received instruction on the German case system before, and most of the participants who took the screening test demonstrated mastery of switch prepositions, but the four participants who were included in the final sample did not produce the correct case markings consistently enough to speak of successful acquisition of the target feature prior to the experiment. Some teachers estimate that it takes at least two years of instruction and/or six months of immersion (Arnett & Jernigan, 2014: 69) for students to grasp most aspects of the German case system, but in the case of these participants, even several more years of instruction and/or a longer immersion experience did not lead to mastery of the target feature. It will therefore be interesting to see whether Processing Instruction can make a difference despite the odds.

The participants were briefed on the general reason for data collection, the experimental design, as well as the processing and safekeeping of collected data. A code was assigned to each individual respectively and no given names were used on any of the materials. Participation was voluntary and could be withdrawn at any point before the submission of the dissertation. The collected data were protected against unauthorised access: They were kept in a secure location, any hard copies were stored in locked cabinets, and electronic copies were password-protected. Data were transferred in a secure manner and movable storage media were protected against loss, damage, and destruction. Participants were made aware of the fact that the data will be retained for five years before being disposed of securely, in keeping with the regulations of the University of Greenwich and its Data Protection Act.

5.6 Steps undertaken to maximise validity and reliability

One measure to maximise reliability was to use established instruments of data collection that have been demonstrated to possess decent validity and reliability scores in the past. Every care has been taken to administer the questionnaires and to collect the Working Memory Capacity data in the most controlled way possible, for example by using automated span tasks and by making sure that the conditions of the experiment were as similar as possible for each individual participant. The Working Memory Span tests were administered in the dominant language of the participants to avoid effects resulting from lack of second language proficiency (Gass & Mackey, 2007: 9). The use of a backward Digit Span measure that tapped to a large extent the same processing resources as the Reading Span task enhanced the construct validity of Working Memory Capacity and at

the same time the forward Digit Span task extended the construct to include a measure of Phonological Short-Term Memory. Where possible, the study employed more than one measure and/or perspective for the elicitation of each factor as well as more than just one assessor. For the marking of the essay task, for example, an experienced teacher of German was asked to evaluate the essays and provide a second opinion. Great care was taken to balance questionnaire items so that they were exhaustive enough to measure the construct without risking fatigue of the participants, and to administer them in a succession that kept priming effects at a minimum. As another validity measure, the materials prompted participants to rate the task difficulty in order to complement the task complexity analyses by the researcher.

The overall design of this study has taken shape over several years and has been refined with each Reading Group discussion and after each conference presentation that generated valuable feedback. An overview of the research process can be found in Appendix P. The materials were tested in a first instance with staff members of the Department of Applied Linguistics and fellow PhD students. A checklist with questions for the material testing session can be found in the Appendix K. The main aim was to see whether the reworded questionnaires were comprehensible and to see how long it would take to work through all the materials. Based on the feedback that the session elicited, the questionnaires were amended and problematic items were discarded. The attitude/motivation questionnaire, for example, was halved in length to make it more manageable. The materials were then tested with two participants at an all-girls sixth form comprehensive school in East London. A pilot study is an “important means of assessing the feasibility and usefulness of the data sampling and collection methods and revising them before they are used with the research participants” (Gass & Mackey, 2007: 3).

In this case, the results from the pilot study inspired the following changes:

- The translation items in the artificial-language test were converted to OVS word order.
- A discourse-level writing task was developed.
- A more challenging interpretation task was designed.
- A delayed post-test was included.

6 RESULTS

In this chapter, the results of the main study on Processing Instruction and individual differences in Working Memory Capacity will be presented. Different scoring techniques for the Working Memory and Phonological Short-Term Memory Capacity measures will be explored with the aim of finding the scoring technique that is most sensitive to individual differences. Then the participants' individual performances before, during, and after Processing Instruction will be presented in a fine-grained analysis that allows both a process-oriented perspective on the interpretation of Structured Input Activities and a product-oriented perspective on the differences between pre- and post-test scores, both within and across participants. In the next step, the data on selected learner variables will be introduced and described in the context of their respective measure. All this will be supplemented with the observational data that the researcher collected while the participants were working through the automated span tasks, in the hope of providing a richer and more complete picture of the assessment situation. Some of the findings will be explored in further detail to evaluate the risk of measurement errors that might have been caused by the nature and/or the administration of the Working Memory tests. Bringing the different variables together, a tentative explanation of the results within the context of this study will be offered before the next chapter discusses and interprets the findings in order to situate them within a wider research context.

6.1 Memory Span Tasks

Scoring-related decisions are important because they affect the statistical power of data analyses. Even with such a small sample size as in this study, scoring can directly affect the rank order of participants, as will be seen later. The Digit Span scores are therefore calculated using composite measures as well as the maximum span scores for both conditions. For the Reading Span task, which is a lot more complex, several different scoring techniques are employed in order to find the technique that is most sensitive to individual differences. Depending on whether one counts individual letters, complete sets of letters, or accurately judged sentences in combination with letters and sets, the performance results differ more or less between participants. As described previously, some scores were automatically provided by the computer programme, Inquisit. But for a complete comparison, additional scoring techniques are applied. Previous research

indicates that a partial-credit approach to scoring has better internal consistency and makes the best use of participant information (Conway et al., 2005; Waters & Caplan, 2003). Especially for research on individual differences, an absolute span score is problematic as the scores can only take few discrete values. Partial-credit approaches, on the other hand, take into account that task performance can depend on several different factors, such as sentence length, display duration, or semantic similarity.

Digit Span Scores

For the Digit Span score, the maximum score is defined as the number of digits in the longest sequence that a participant has perfectly recalled. For the participants in this study, the maximum forward span ranges between 7 and 10 digits. For the composite forward Digit Span score, all those digits which were recalled in their correct position within a sequence are added up. Even if a sequence was not recalled accurately in its entirety, the digits which were recalled in their correct position still count for this score. This scoring technique therefore makes use of all the available data by not discarding inaccurate or incomplete digit strings. Note that the maximum span scores do not differentiate performance levels as much as the composite scores do, e.g. two participants have a maximum span of 9 digits and appear to be performing equally well. If the composite scores are consulted, however, it is clear that one of the two participants performs better (i.e. scores more consistently) than the other. The scores for the forward Digit Span measure can be found in Table 6.1 below:

Participant	Forward Digit Span maximum	Forward Digit Span composite
MgrYyVvU	9	50
YKX5Bwvj	9	42
74YuGCRD	10	62
nLQ6aVus	7	37

Table 6.1: Maximum Forward Digit Span scores and composite Forward Digit Span scores

For the backward condition of the Digit Span (see Table 6.2 below), the same two scoring techniques are employed. The maximum span is the longest sequence of digits that a participant could recall in the reverse order of presentation. The maximum span for the

participants in this study ranges between 6 and 9 digits, which is remarkably close to the span range observed in the forward condition. For the composite backward Digit Span score, all the perfectly recalled digits are added up. Even if a recalled sequence was not entirely accurate, the digits which were recalled in their correct position still count for this score. As in the case of the forward Digit Span scores, the composite scores for the backward condition are more sensitive to performance differences than the maximum span scores:

Participant	Backward Digit Span maximum	Backward Digit Span composite
MgrYyVvU	8	42
YKX5Bwvj	6	31
74YuGCRD	9	52
nLQ6aVus	7	39

Table 6.2: Maximum Backward Digit Span scores and composite Backward Digit Span scores

Reading Span Scores

First, the Reading Span scores are counted in terms of all letters that have been recalled in their correct position, irrespective of the accuracy of judgement or the completeness of the set. Then the number of complete recalled sets is counted. Both of these scoring techniques are composite scores and have been employed in previous Working Memory research. The results for the absolute (whole sets) and partial-credit (individual letters) scores can be found in Table 6.3 below. It should be noted that the scoring of letters from complete sets only gives an additional layer of information in terms of the “recall style”: YKX5Bwvj and 74YuGCRD both recalled the same total number of letters, but YKX5Bwvj recalled more individual letters, whereas 74YuGCRD recalled more complete sets.

Participant	score of all letters recalled	score of all complete sets
MgrYyVvU	65	45
YKX5Bwvj	56	26
74YuGCRD	56	43
nLQ6aVus	40	8

Table 6.3: Reading Span scores for all perfectly recalled letters vs all perfectly recalled sets

Another scoring decision concerns the question whether all correctly recalled letters count, or whether only those letters count which have been recalled after correct judgement of the sentences. The argument for the latter stance is that the Reading Span task is only a complex span measure if the distractor task receives appropriate attention. A lower accuracy rate on the judgement task could also point at a greater strain on Working Memory resources. Counting only correctly judged trials would therefore make differences between high-span and low-span participants more salient. Kane et al. (2004), on the other hand, argue that participants who recall the most letters usually have the highest accuracy scores on the processing component of the task as well. A comparison of scores based on all items versus scores based on accurately judged items only can be found in Table 6.4 below. It should be noted that the comparison of the scoring techniques highlights different aspects of the task:

Comparing YKX5Bwj and 74YuGCRD again, the scores show a difference between judgement and recall performances. Considering the total number of recalled letters, these two participants fare equally well. But while 74YuGCRD recalled either complete sets after accurate judgement, or judged inaccurately and recalled no letters in the correct place afterwards, YKX5Bwj recalled more individual letters and her sentence judgement did not affect the whole letter string, merely the subsequently displayed letter. This will be discussed in more detail in the next section.

Participant	all letters recalled	all letters recalled (accurate items only)	all sets	all sets (accurate items only)	maximum set size	maximum set size (accurate items only)
MgrYyVvU	65	54	45	39	6	6
YKX5Bwj	56	43	26	16	7	7
74YuGCRD	56	38	43	37	6	6
nLQ6aVus	40	20	8	3	5	3

Table 6.4: Measuring all perfectly recalled letters/sets vs accounting for accuracy of judgement

The descriptive statistics for a normative sample of US university and college students showed a very slight tendency for males to make more processing errors while females had slightly lower storage scores (Redick, et al., 2012: 168). In this study, the highest storage score goes indeed to a male participant and the lowest to a female participant, but the middle positions are equally distributed for storage capacity. The participant with the lowest storage score has also the lowest processing accuracy, whereas the

percentages for the other participants are almost identical. Table 6.5 below shows the storage and accuracy scores for the four participants in this study:

Participant	accuracy %	all letters recalled	sex
MgrYyVvU	96	65	M
YKX5Bwj	94.7	56	F
74YuGCRD	94.7	56	M
nLQ6aVus	86.7	40	F

Table 6.5: Storage and processing scores in combination with participants' sex

The overview of the absolute span scores across participants for all scoring techniques (see Figure 6.1 below) shows which scoring procedures offer the best differentiation between the participants' performance levels. Maximum span scores, whether they refer to sets of letters in the Reading Span test or digit sequences in the Digit Span test, produce the smallest ranges and do not show clear differences between the participants' performances. Composite scores indicate wider gaps as they cover a wider range of potential scores. They are therefore more sensitive to individual differences in Working Memory (or Phonological Short-Term Memory) Capacity. Note that the counting of letters from correctly judged items only does not necessarily provide better discrimination between participants, but can affect the participant ranking:

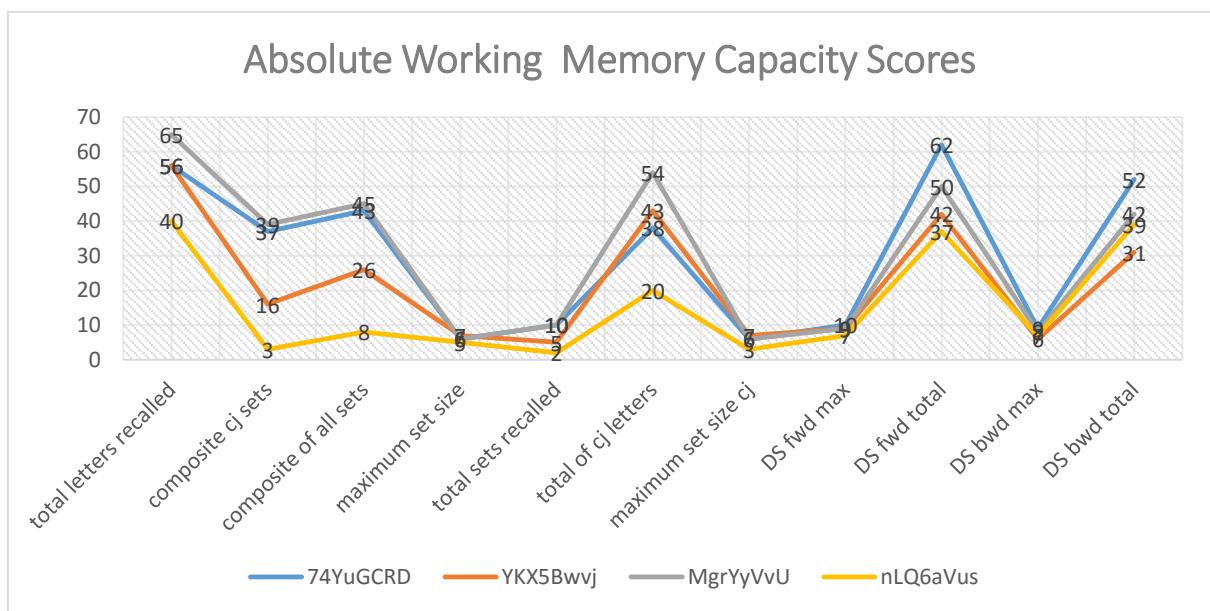


Figure 6.1: Overview of the absolute span scores by measure

Considering that the score range, while continuous and theoretically open-ended, can be very narrow or moderately spaced, depending on the scoring technique, a comparison of Working Memory Capacity scores can be difficult without descriptive statistics. In order to compare the performances of the participants in this study, relative span scores are calculated (see Table 6.2 below). To that end, the highest score for each measure is taken as the maximum value (=100%) and all scores below the maximum are converted into percentages. If the scores for all the letters that were perfectly recalled as complete sets and after accurate judgement of all relevant sentences are compared, then participant nLQ6aVus, for example, recalled 8% (or 3 letters) of the number of letters that participant MgrYyVvUu recalled (namely 39 letters).

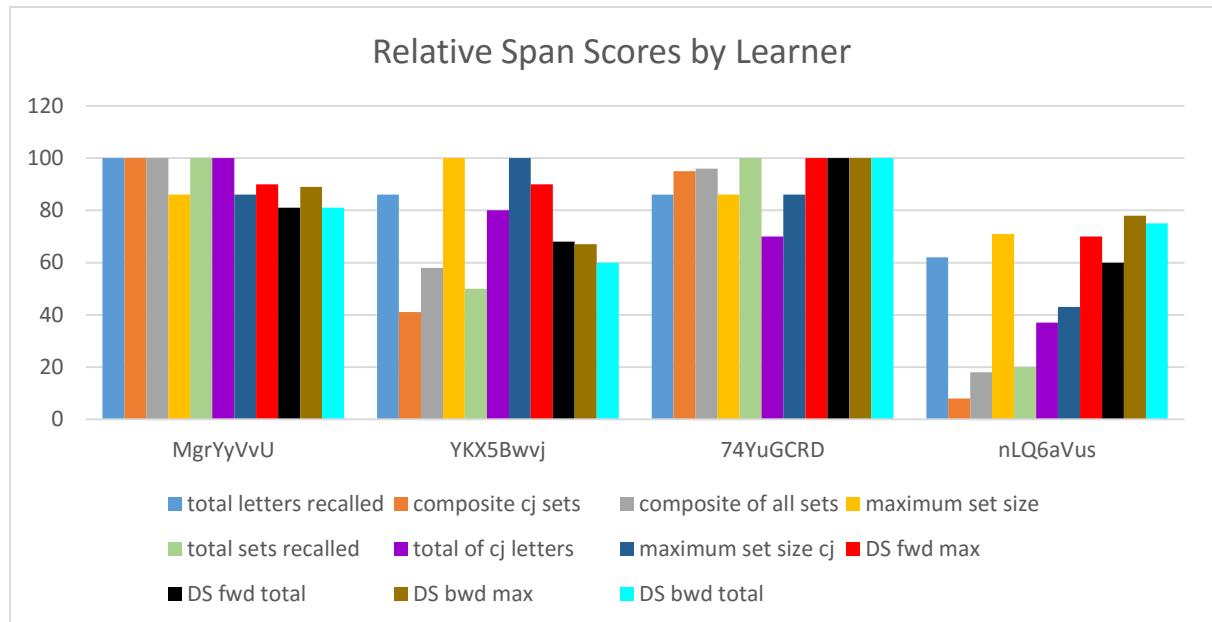


Figure 6.2: Relative Span Scores by Learner (in percentages)

As the relative Working Memory scores in Figure 6.2 show, the comparison of maximum set sizes (shown in yellow) is the least sensitive measure and hardly distinguishes the participants' performance on memory span tasks at all. The maximum set size with all sentences judged correctly (shown in dark blue) is already more discerning. Both the composite scores of all letters recalled in the correct position (shown in light blue) and of all letters recalled in the correct position after correct judgement of the sentence (shown in violet) are increasing sensitivity. Scores of the number of complete sets recalled (shown in light green) and the composite of all those sets (shown in grey) highlight performance differences even more, but the most sensitive measure is the composite score of all complete sets that were also completely correct in terms of sentence judgement (shown in orange). In a sense, the latter scores describe the most challenging task, i.e. the recall

of all letters in a set, without fail, combined with 100% accuracy in judging the distractor task. It makes therefore sense that this should be the measure that elicits the differences in span scores best. This scoring technique goes against the recommendations of psychometric guidelines for Working Memory measurement as, from a statistical point of view, it does not make use of all the available information because it discards correctly recalled letters from incomplete or misjudged sets. It does, however, seem to bring out the most pronounced differences in performance on Reading Span tasks in this experiment.

If one wanted to rank the participants by Working Memory Capacity scores, for most of the scoring techniques, the ranking would look as follows:

MgrYyVvU > 74YuGCRD > YKX5Bwj > nLQ6aVus.

Note that the maximum set size, whether based on all sentences or on correctly judged sentences only, leads to a different participant ranking than the other scores. Participant nLQ6aVus, however, seems to score lowest, no matter which score is used. We will get back to this ranking in the discussion of the results section.

Looking at the relative Digit Span scores in Figure 6.2, the differences are not as obvious. The maximum forward Digit Span (shown in red) varies between nine and ten digits for the first three participants. The participant with the lowest Working Memory Capacity scores also fares worst on this measure, recalling a maximum of 7 digits. The total of all digits recalled in the correct position (shown in black) highlights these differences without changing the ranking of the participants. Note, however, that the participant ranking for the forward Digit Span scores differs from the Working Memory span ranking:

74YuGCRD > MgrYyVvU > YKX5Bwj > nLQ6aVus (forward)

The scores for the backward Digit Span measure show another slightly different picture. The maximum backward scores (shown in brown) are within a range of six to nine digits, so slightly lower than in the forward condition. Considering that the participants had to manipulate the sequence of the digit string before recall, these are still rather high scores. The total of all digits recalled correctly in the reverse order of presentation (shown in turquoise) is again more sensitive to span differences than the maximum score, but the

participant ranking is the same for both scoring techniques. Note that the ranking differs from the forward condition though:

74YuGCRD > MgrYyVvU > nLQ6aVus > YKX5Bwvj (backward)

6.2 Test scores before, during, and after instruction

On-line performance during Structured Input Activities (session 1)

A complete list of items and individual scores can be found in the Appendix I. The Figures 6.3 and 6.4 below show the on-line performance of the participants on the case revision tasks. Note that the participants only commit one error during the accusative revision and the dative revision, respectively. Two out of four participants interpret the Structured Input correctly for all ten sentences and for both activities. Three out of four participants interpret the input correctly for all ten sentences in one activity and for nine sentences in the other activity. The errors occur on the third and fourth slides. Once the participants had completed the task, they rated the difficulty level on a five-point Likert scale from ‘too easy’ to ‘too difficult’. For the accusative and dative Structured Input Activities, all participants indicated that the tasks were a 2 on the Likert scale, meaning ‘easy enough’.

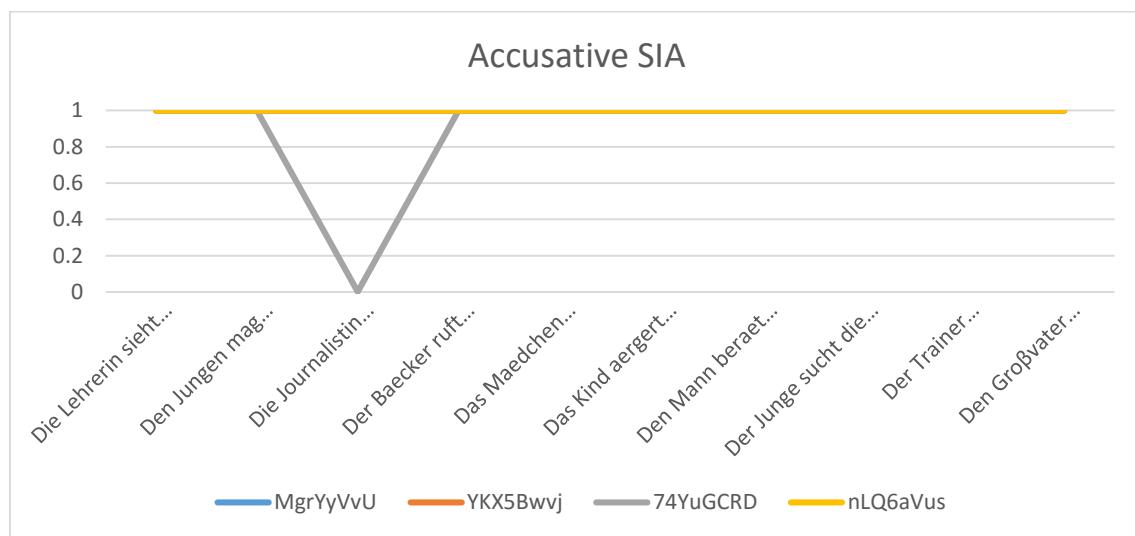


Figure 6.3: Performance on the accusative SIA

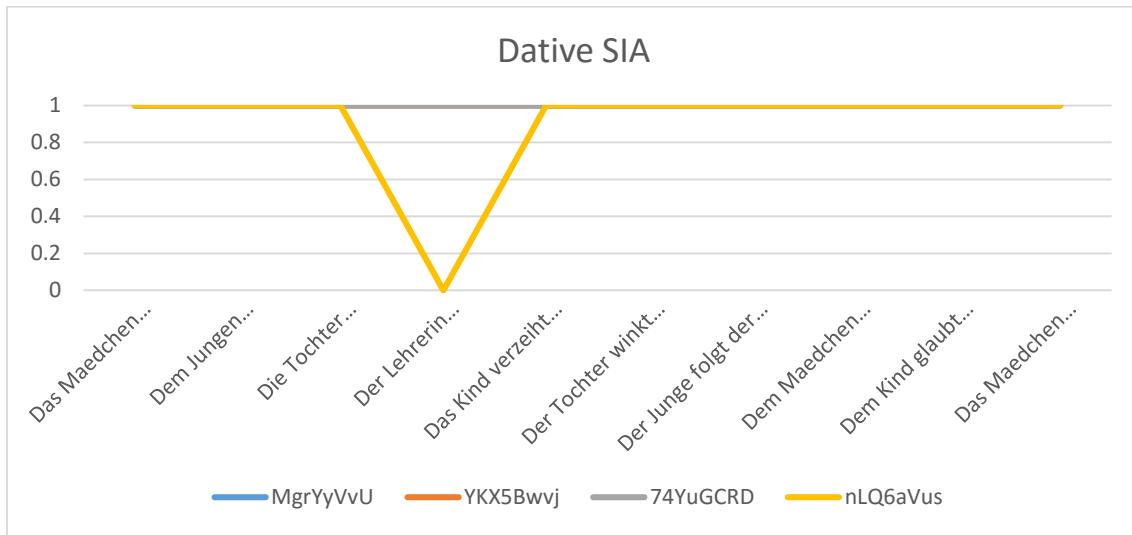


Figure 6.4: Performance on the dative SIA

Figure 6.5 below shows the performance scores for the first Structured Input Activity on switch prepositions, namely the picture-matching task. Compared to the first two activities, more errors were made and only one participant, the English-German bilingual, matched all the sentences with the correct picture. One participant, 74YuGCRD, matched five pictures correctly after getting the first match wrong. Another participant, the one with the lowest Working Memory scores, starts with three correct items and then matches the wrong pictures to the last three sentences. The fourth participant shows a more dynamic pattern, matching the first, third, and sixth sentences with the correct picture, but not the other three sentences in between. The two most successful participants judged the difficulty level as ‘easy enough’, whereas nLQ6Vus, who started with three right matches, rated the difficulty level as ‘moderate’ (3). The participant with the most inconsistent scoring, MgrYyVvU, rated the task as ‘challenging’ (4).

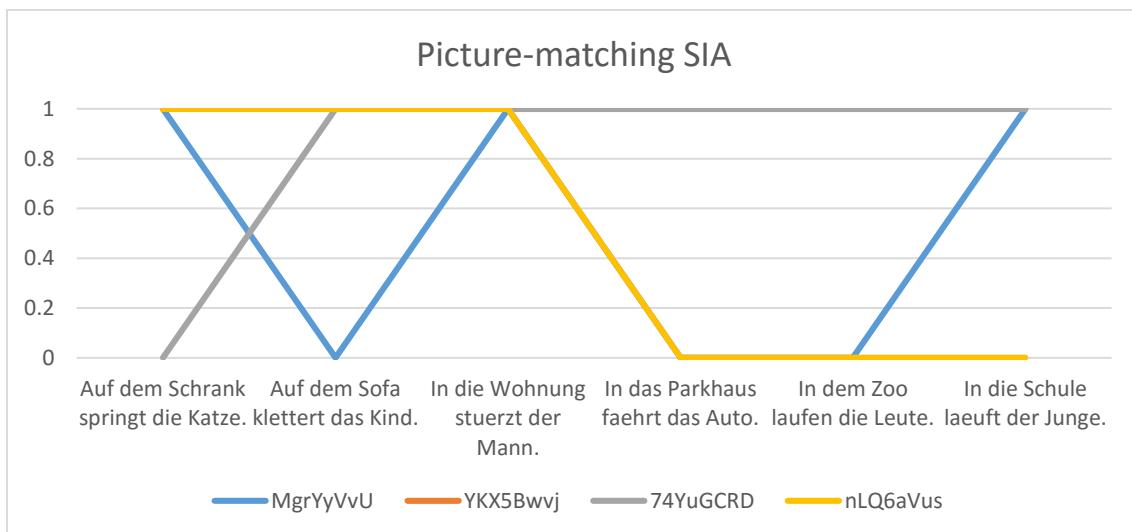


Figure 6.5: Performance on the picture-matching SIA

The sentence-level comprehension task, as shown in Figure 6.6 below, elicited a more homogeneous picture again: Three out of four participants achieved full scores for all nine sentences and one participant made a mistake on the very first slide, which could have been a blip in attention. All in all, it seems that this Structured Input Activity, true to the nature of Processing Instruction, did not challenge the participants. This is also reflected in the difficulty ratings: all of the participants perceived this task as ‘easy enough’ (2).

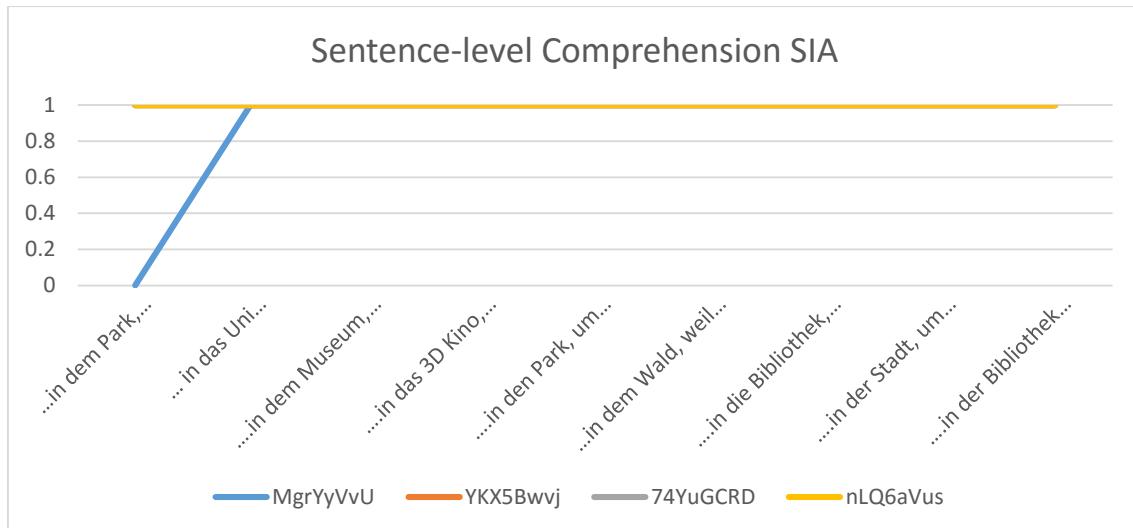


Figure 6.6: Performance on the sentence-level comprehension SIA

The sentence-level comprehension task, which used aural input, shows a lot more variability in scores (see Figure 6.7 below). Only two participants, MgrYyVvU (whose scores are hidden under the yellow and orange lines, unfortunately) and 74YuGCRD produce five correct answers in a row at any point. The latter is the most consistent scorer with only one error. Two other participants got three of the ten sentences wrong, and nLQ6aVus four. The difference between MgrYyVvU and YKX5Bwvj is that the first participant scored slightly more consistently towards the end. That 74YuGCRD scored nine out of ten in the listening comprehension Structured Input Activity is somewhat remarkable as this participant, having just returned from a year abroad in France, had received no German input directly before the study whereas MgrYyVvU and YKX5Bwvj had spent several weeks to one year in Germany prior to the experiment. While 74YuGCRD and YKX5Bwvj rate the task as ‘easy enough’ (2), the other two participants rate the difficulty as ‘moderate’ (3).

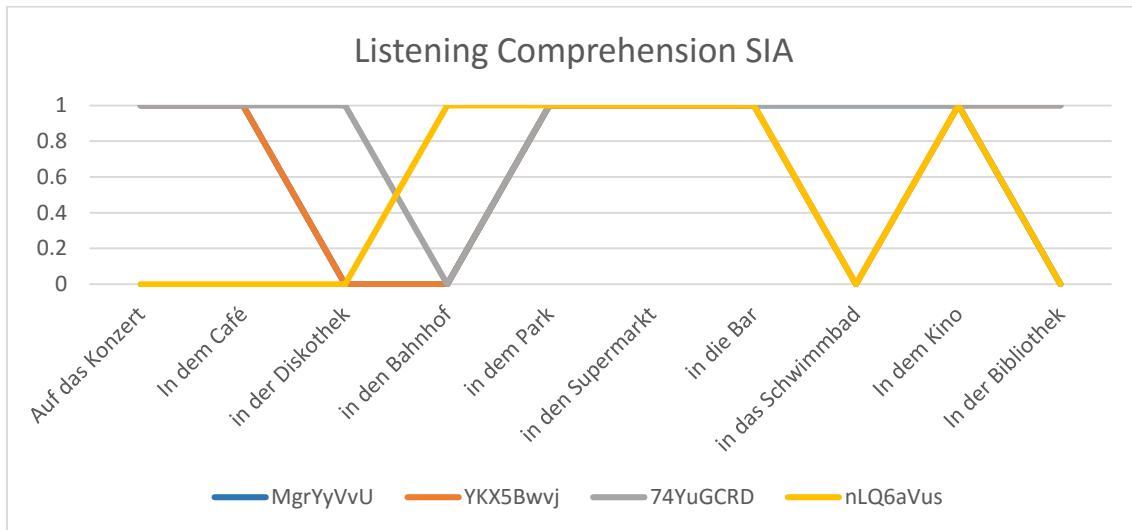


Figure 6.7: Performance on the sentence-level listening comprehension SIA

The discourse-level Structured Input Activity (see Figure 6.8 below) on how cities make people more human consisted of five target questions requiring the correct interpretation of the case marking, five distractor questions requiring the correct processing of the text content, and two opinion questions where the participants were asked to which extent they agreed with the text. nLQ6aVus was the only participant to answer all ten questions correctly. 74YuGCRD and MgrYyVvU got one target question wrong each. YKX5Bwj, the bilingual participant, chose a wrong answer for one target question and for one content question. Only 74YuGCRD rated the task difficulty as ‘moderate’ (3), the other participants found it ‘easy enough’ (2). It should be noted that participants did not receive feedback on this task.

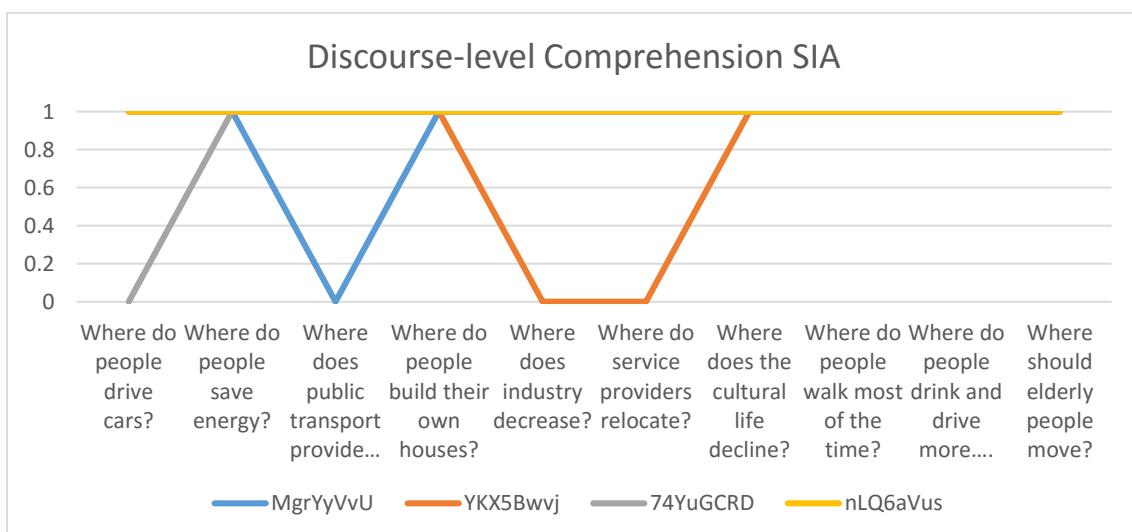


Figure 6.8: Performance on the discourse-level comprehension SIA

On-line performance during Structured Input Activities (session 2)

As described in more detail in the materials section, the picture-matching task in session two recycled the sentences used in the first session and also used the alternative versions of each sentence. The participants did better than in the first session (where 6 sentences lead to 7 errors in total): Two of the participants matched the correct picture to all 12 sentences and two participants chose one wrong picture, respectively (12 sentences, 2 errors in total). The improvement in performance is also reflected in the difficulty ratings: All four participants now rate the task as ‘easy enough’ (2), whereas before, one participant rated it as ‘moderate’ (3) and one as ‘challenging’ (4).

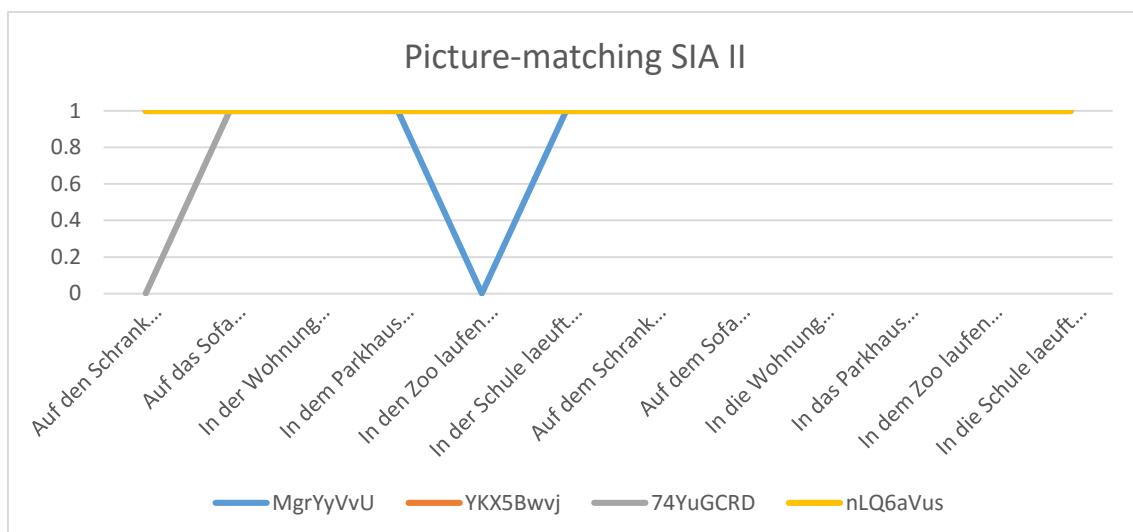


Figure 6.9: Performance on the picture-matching SIA in session 2

The sentence-level comprehension task was essentially the same in the second session and the nine sentences were interpreted correctly by three out of four students, with the fourth participant, YKX5Bwvj, interpreting one sentence incorrectly. Overall, the picture is the same as in the first session, indicating that the participants were performing at ceiling level on this task. This is also reflected in the difficulty ratings as all participants rated the task as ‘easy enough’ (2) once more. The performances can be seen in Figure 6.10 below.

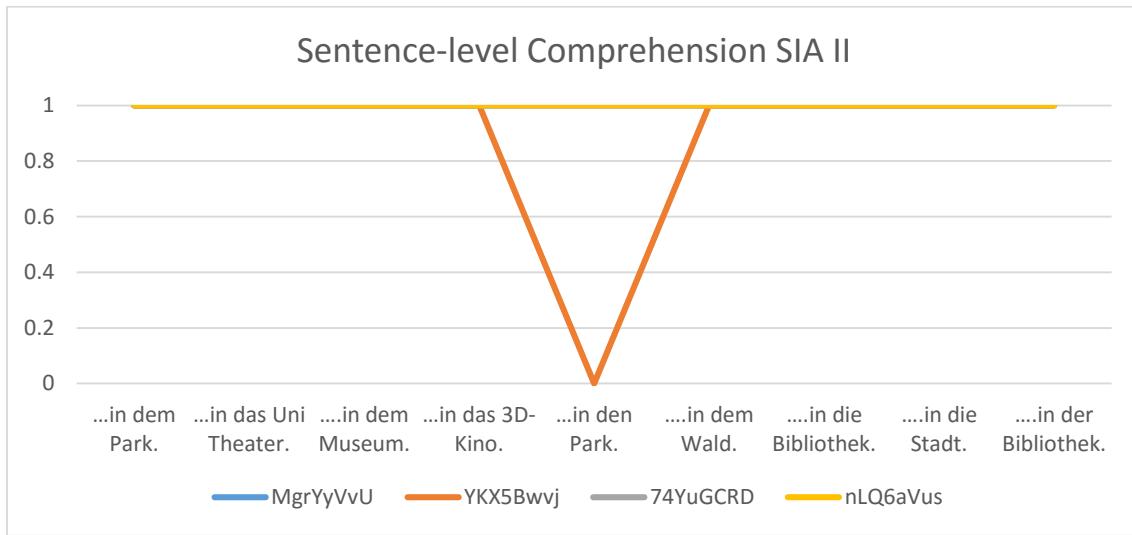


Figure 6.10: Performance on the sentence-level comprehension SIA in session 2

The Direction versus Location task was a new one. One participant, MgrYyVvU, answered all 15 questions correctly. YKX5Bwj answered one question incorrectly. Those two participants were also the participants who had received the most native-speaker input and had been immersed in a German-speaking context for the longest time directly prior to the experiment. nLQ6aVus answered three questions incorrectly. 74YuGCRD, the participants with the least exposure to German input, answered five questions incorrectly. He also rated the difficulty as ‘moderate’ (3), while all other participants rated the task as ‘easy enough’ (2).

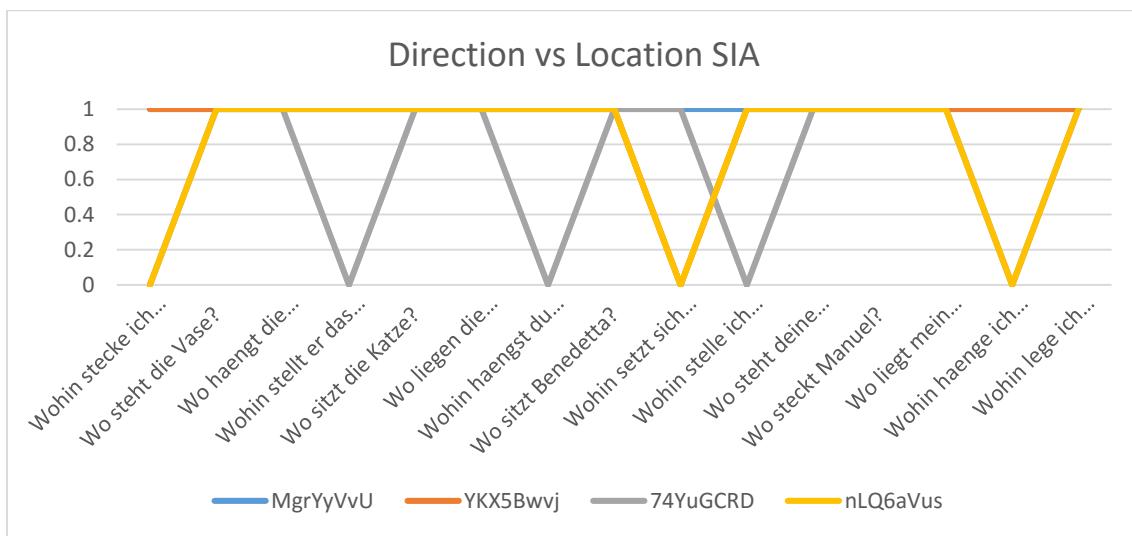


Figure 6.11: Performance on the direction vs location SIA in session 2

The Aural Sentence Puzzle was another new task that had not been presented in the first session. Two participants, YKX5Bwvj and nLQ6aVus, chose three wrong sentence endings each for the 18 aural sentence snippets. The other two participants chose four wrong sentence endings, respectively. The difficulty ratings reflect this again: MgrYyVvU and 74YuGCRD rated the task difficulty as ‘moderate’ (3), whereas the slightly more accurate YKX5Bwvj and nLQ6aVus rated the task as ‘easy enough’ (2).

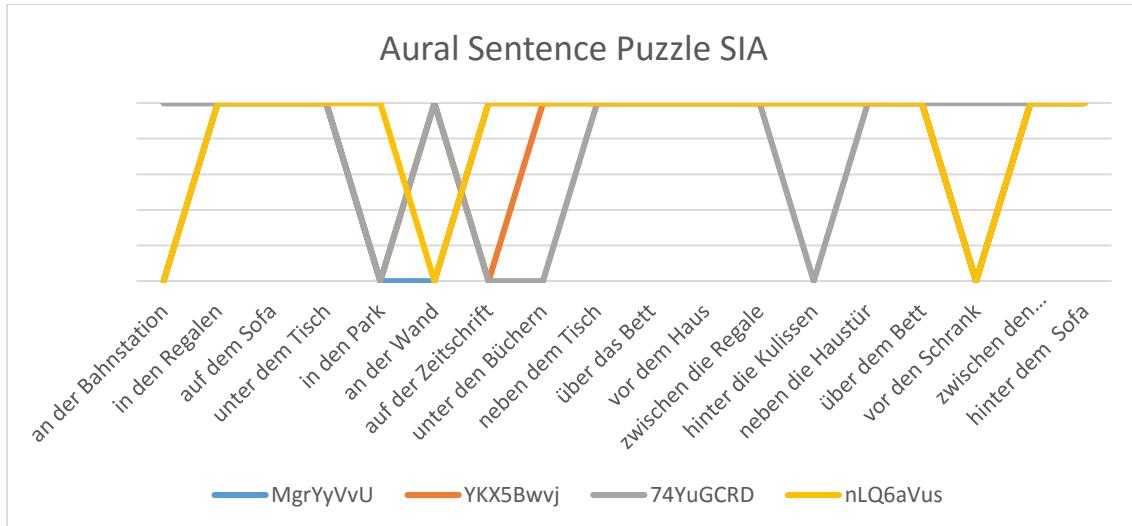


Figure 6.12: Performance on the aural sentence puzzle SIA in session 2

Pre- and Post-test scores

As described in Chapter 5, the administration of the pre-test took place in the week before instruction, the first post-test was administered immediately after the second instructional session, and the second post-test was administered circa 15 weeks after instruction. The pre- and post-tests can be found in the Appendix. A detailed list of the scores and responses is also available there.

Fill-in-the-blanks task

This task is one of the classic test formats that one can find in any language classroom. The participants were in all probability very familiar with it. For this task, which provided gender and number information, only completely correct forms were counted. The pre-test scores for this controlled, output-oriented task were already relatively high to start with and improved from pre- to immediate post-test. There was a rather sharp decline in scores from immediate to delayed post-test though, with three out of four participants

scoring lower than on the pre-tests. Only one participant, the bilingual YKX5Bwj, managed to improve consistently from pre- to immediate and ultimately even the delayed post-test. It is possible that this participant might have been exposed to a different quality and/or quantity of German input between tests than the other participants. It should also be noted that the two participants with the highest Working Memory Capacity scores, MgrYyVvU and 74YuGCRD, showed an almost identical trajectory.

The participant with the lowest Working Memory Capacity scores started as the strongest candidate on the pre-test, but then her scores decreased steadily from there to post-test 1 and even further to post-test 2. These results will be discussed in more detail in the next section. The participants' difficulty ratings for the task varied: Some ratings were more stable over time (YKX5Bwj rated this task as 'easy enough' (2) across all sessions and nLQ6aVus rated the task difficulty as 'moderate' (3) across all sessions), others fluctuated. MgrYyVvU rated the pre-test and the immediate post-test difficulty as 'moderate' (3), but the delayed post-test as 'challenging'. While this might be reflected in MgrYyVvU's scores, 74YuGCRD rated the first post-test, in which he scored highest, as the most difficult ('moderate') of the three test sessions.

Participant	Pre Fill-in-the-blanks	Post 1 Fill-in-the-blanks	Post 2 Fill-in-the-blanks
MgrYyVvU	60	95	55
YKX5Bwj	75	85	95
74YuGCRD	60	100	55
nLQ6aVus	80	70	50

Table 6.6: Pre- and post-test scores for the fill-in-the-blanks task

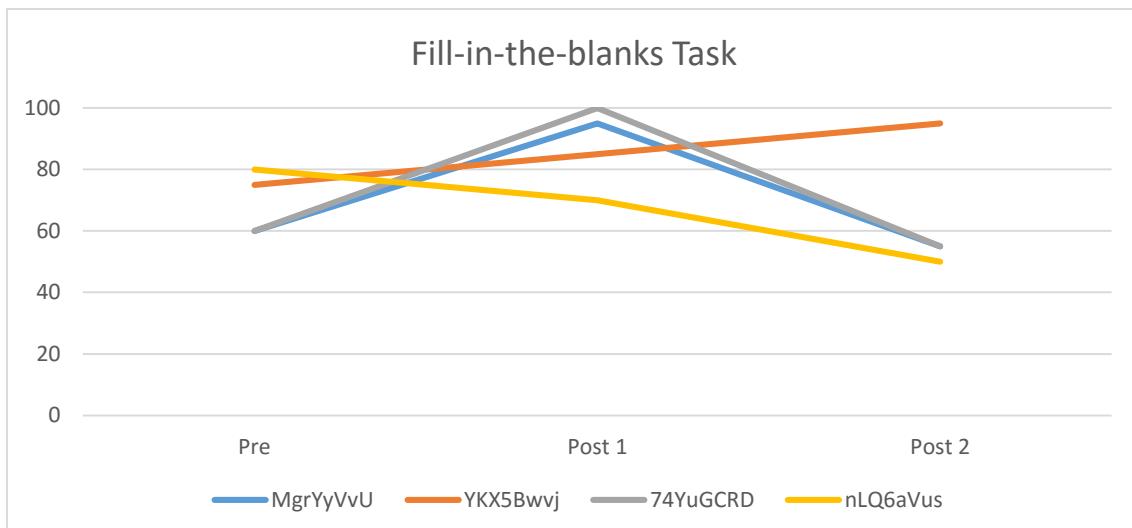


Figure 6.13: Pre- and post-test scores for the fill-in-the-blanks task

Picture description task

The picture description task is another task that can be found in any language classroom. It was a relatively basic and structured task that still left enough freedom for the participant to avoid structures and/or words they were not comfortable with. For a university sample, this task was almost too basic. Yet, the scores showed more variability than expected. While one participant, 74YuGCRD, consistently performed at ceiling level in all three instances of the test, the other three participants started out with lower scores. nLQ6aVus started with 86% on the pre-test and managed to be 100% accurate on both post-tests. MgrYyVvU started at a similar level, improved similarly, but then dropped below the pre-test score. It should be noted that the bilingual participant started and finished with the lowest scores of the group. Both 74YuGCRD and YKX5Bwvj rated the task difficulty as 'easy enough' (2) across all test sessions. nLQ6aVus rated the task difficulty as 'moderate' (3) across all sessions. MgrYyVvU rated the pre-test and the delayed post-test as 'moderate' (3), but the immediate post-test as 'easy enough' (2). The scores will be discussed in more detail in the next section.

Participant	Pre Picture Description	Post 1 Picture Description	Post 2 Picture Description
MgrYyVvU	83	100	80
YKX5Bwvj	50	100	20
74YuGCRD	100	100	100
nLQ6aVus	86	100	100

Table 6.7: Pre- and post-test scores for the picture description task

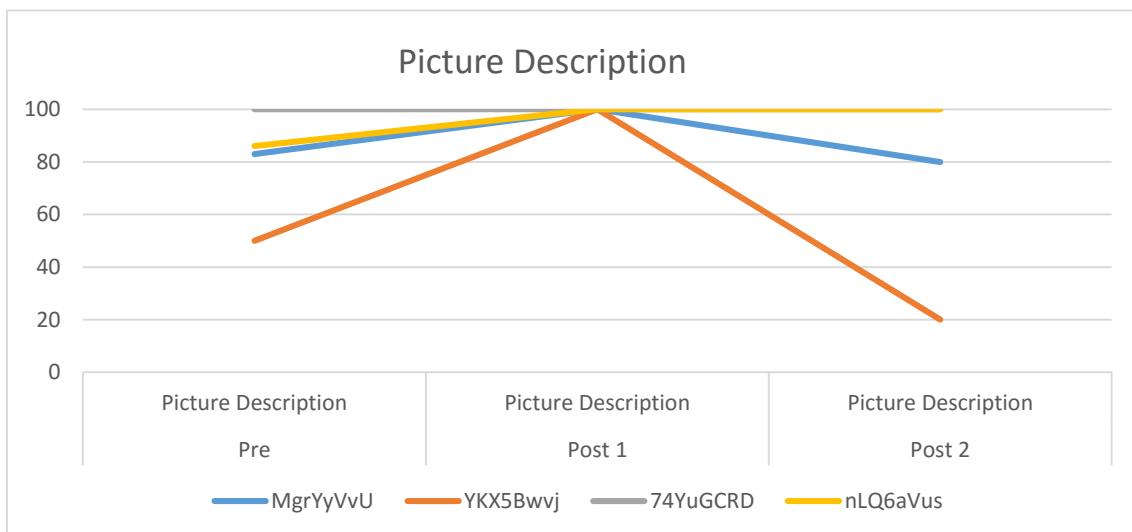


Figure 6.14: Pre- and post-test scores for the picture description task

Translation task

The translation task was insofar the most interesting task as it required students to interpret the subject and object roles based on grammatical features rather than word order, context, or event probabilities. This task also elicited the greatest differences in pre-test performance levels, as shown in Table 6.8 below. Both YKX5Bwvj and nLQ6aVus misinterpreted all seven OVS sentences before they received Processing Instruction. While the former's accuracy scores increased to 60% on the immediate post-test and 86% on the delayed post-test, the latter only showed a small improvement (29%), which she could not maintain on the delayed post-test (14%). MgrYyVvU performed rather well on the pre-test with 71% accuracy, and was the only participant to reach 100% on the immediate post-test, but could not repeat this on the delayed post-test. 74YuGCRD, however, gained steadily from test to test and eventually scored 100% on the delayed post-test. As can be seen in Table 6.8 below, YKX5Bwvj and 74YuGCRD show similar trajectories. It should be noted though that the quality of the translations varied a lot across participants and test sessions. For the sentence "*Der Tochter schaute die Mutter von der Empore aus zu.*", which could be translated as "The mother is watching the daughter from the gallery.", the same participant produced translations such as "The daughter looks like the mother because of the freckles." (post-test 1) and "The daughter looks like the mother from the Empore." (post-test 2). The bilingual participant, for the same sentence, produced the translation "The mother of the emperor watches the daughter." (post-test 2). While both high-span participants lacked the English equivalent for "Empore", they interpreted the subject and object roles correctly in all instances and left a blank where the word gallery would have fit in. This shows rather stable representations of the sentence structure. The difficulty ratings varied a lot across participants and time: While the bilingual participant rated the task difficulty in all sessions as 'easy enough' (2), MgrYyVvU and 74YuGCRD rated it as 'moderate' (3). nLQ6aVus rated the pre-test as 'moderate', but both post-tests as 'challenging' (4).

Participant	Pre Translation task	Post 1 Translation task	Post 2 Translation task
MgrYyVvU	71	100	57
YKX5Bwvj	0	60	86
74YuGCRD	43	86	100
nLQ6aVus	0	29	14

Table 6.8: Pre- and post-test scores for the translation task

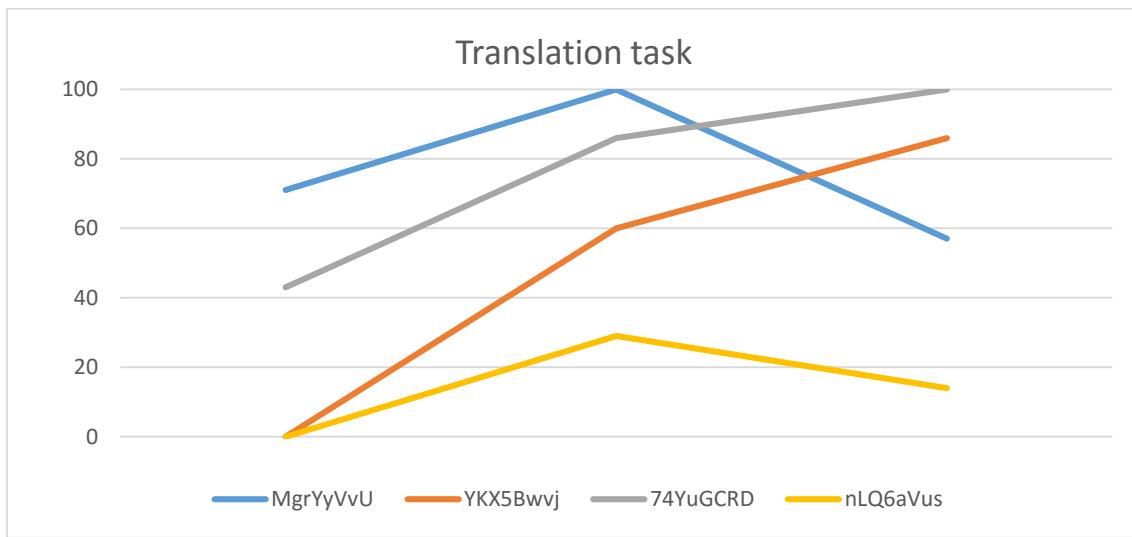


Figure 6.15: Pre- and post-test scores for the translation task

Essay task

The essay question varied from the pre-test to the post-test sessions. For the pre-test, the participants were asked to write about their favourite holiday destination and how to get there, what to do there, etc. For the post-test, the participants were asked to write about their ideal university, its location, how they would get there, what to study there, etc. In the material testing phase, both essay versions had elicited a comparable amount of switch prepositions from advanced learners as well as native speakers of German. Table 6.9 below shows the instances of correctly used switch prepositions across sessions. In the pre-test essay, MgrYyVvU produced one correct switch preposition, whereas 74YuGCRD and nLQ6aVus produced seven correct instances of switch prepositions each. The bilingual participant produced five correct switch prepositions. While 74YuGCRD produced essays of comparable length as well as the same amount of switch prepositions in all three essays, the other participants produced more switch prepositions in both post-test sessions than in the pre-test session.

While the numbers generally dropped from immediate post-test to delayed post-test, YKX5Bwvj produced more correct switch prepositions with each session. This trajectory is comparable to her performance scores for the translation task and the fill-in-the-blanks task and will be discussed later. The difficulty ratings for the essay task ranged between ‘easy enough’ (2) and ‘challenging’ (3). The bilingual participant surprisingly rated the pre-test as ‘challenging’ (4). nLQ6aVus rated the task difficulty as ‘moderate’ (3) throughout.

Participant	Pre Essay	Post 1 Essay	Post 2 Essay
MgrYyVvU	1	8	5
YKX5Bwj	5	8	12
74YuGCRD	7	7	7
nLQ6aVus	7	12	8

Table 6.9: Instances of switch prepositions in the pre- and post-test sessions

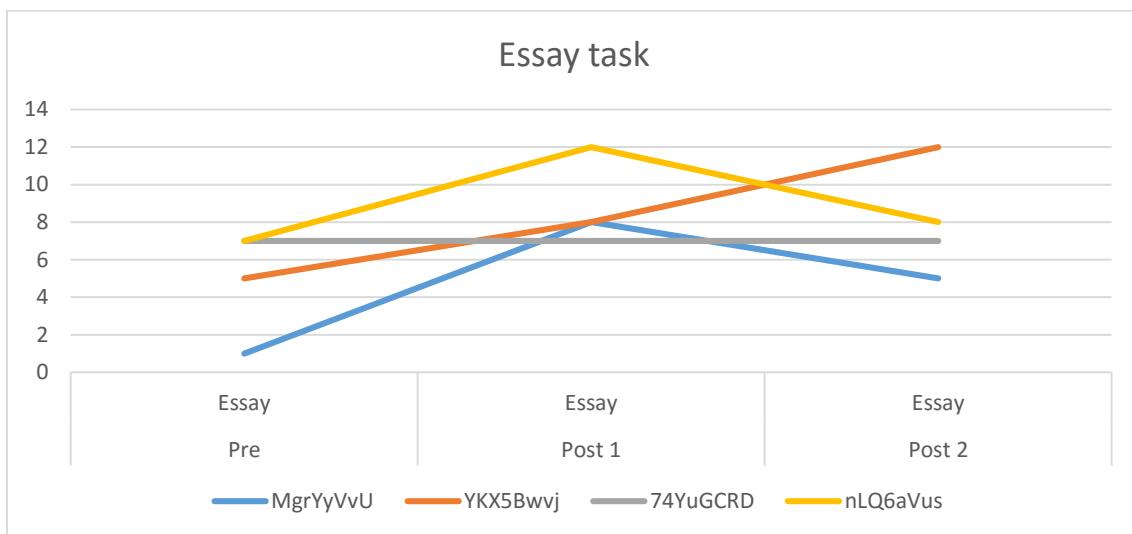


Figure 6.16: Instances of switch prepositions in the pre- and post-test sessions

The overall performance scores for the essay task, based on the complexity of the linguistic structures attempted and the overall composition of the essay, are shown in Table 6.10 below. The essays were graded in accordance with the guidelines of the participants' university, both by the researcher and a lecturer of German from a different institution who had never taught any of the participants. All four participants performed at a similar level. 74YuGCRD scored 60% on all three essays, which were remarkably similar in length and style. YKX5Bwj again improved her scoring slightly from pre- to both post-tests, whereas nLQ6aVus performed slightly less well regarding the post-test essays. MgrYyVvU's performance increased most from pre-test to post-tests, not just in terms of the number of correct switch prepositions he produced, but also in terms of the complexity of the language, text length, and essay composition. It should be noted that this essay task was rather basic for a university-level participant sample. The two participants in their final year, MgrYyVvU and 74YuGCRD, should have had an advantage, but the two first-year students might actually have been more familiar with this kind of task from their recent school days. In the end, the two participants with the

longest immersion experiences scored slightly higher than the rest, but they produced by no means sophisticated texts, not even the bilingual participant.

Participant	Pre overall score in %	Post 1 overall score in %	Post 2 overall score in %
MgrYyVvU	50	65	65
YKX5Bwvj	60	65	65
74YuGCRD	60	60	60
nLQ6aVus	65	60	60

Table 6.10: Overall scores for the Essay task in the pre- and post-test sessions

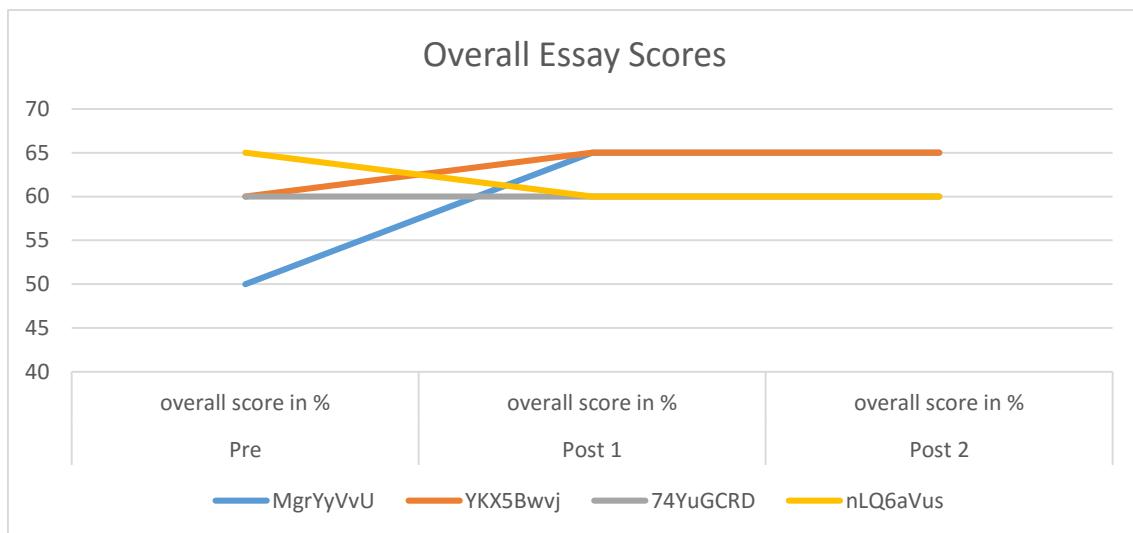


Figure 6.17: Overall scores for the Essay task in the pre- and post-test sessions

As mentioned above, the accuracy of the language used in the essays left a lot to be desired for a group of university students. Table 6.11 below shows the overall accuracy percentages, i.e. errors per 100 words. As errors counted all mistakes related to verb conjugation, tense and mood, case marking, word order, agreement between adjectives and nouns, as well as blatantly wrong word choices. The error percentages in the pre-test essay ranged between 11% and 19%, and most of the participants showed lower percentages for the immediate post-test. YKX5Bwvj was the only participant to improve across all tests, the other participants produced roughly as many errors in the delayed post-test as they had produced in the pre-test. Only MgrYyVvU, who had started with the highest error percentage, produced significantly fewer errors on the delayed post-test.

Participant	Pre Essay error rate in %	Post 1 Essay error rate in %	Post 2 Essay error rate in %
MgrYyVvU	19	4	10
YKX5Bwvj	14	9	8
74YuGCRD	13	7	11
nLQ6aVus	11	9	11

Table 6.11: Essay overall error rates for the pre- and post-test sessions

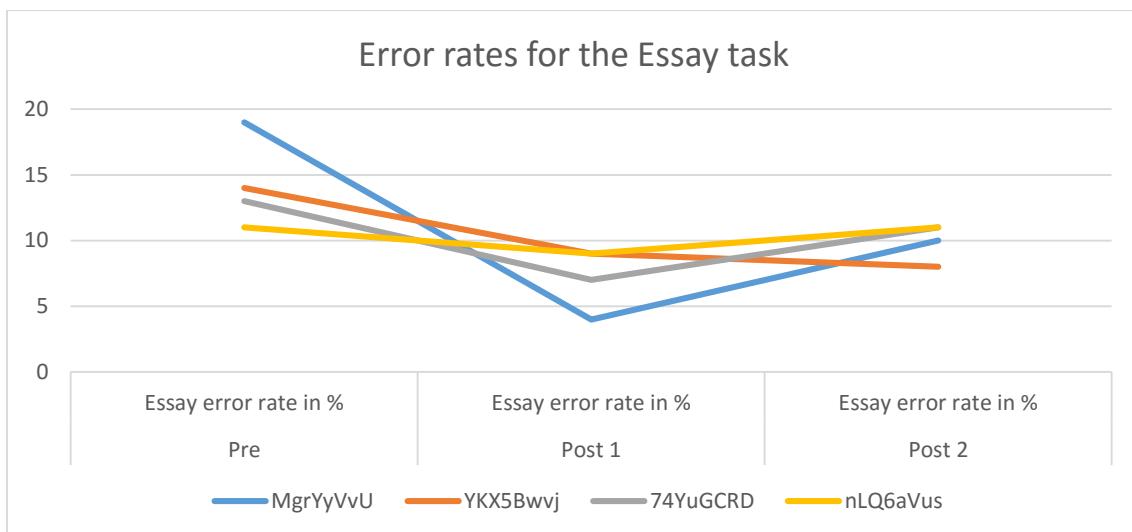


Figure 6.18: Essay overall error rates for the pre- and post-test sessions

6.3 Learner Variables

Interest in foreign language learning and attitude towards German

The interest in foreign language learning was pretty high, as could be expected from participants who are enrolled in an undergraduate language programme. The highest attitude scores were achieved by the two participants who had spent the most time in a German-speaking country and in touch with native speakers, namely YKX5Bwvj (the bilingual participant) and MgrYyVvU (who had just returned from a year abroad in Berlin). The person with the least contact to native speakers and hardly any experience of a German-speaking country, 74YuGCRD, scored lowest in terms of attitude towards German.

Participant	interest	attitude
MgrYyVvU	100	88
YKX5Bwvj	96	92
74YuGCRD	88	54
nLQ6aVus	96	79

Table 6.12: Interest and attitude scores

Anxiety

The Attitude/Motivation Test Battery differentiates between use anxiety and class anxiety. Not surprisingly, YKX5Bwvj (the bilingual) did not express any anxiety when using the language with native speakers. She did, however, express some anxiety in a classroom setting. Considering that she was in her first year of university, this score seemed rather low though. MgrYyVvU showed a similar level of anxiety, but only in a native-speaker environment. His class anxiety was lower, which could have to do with the fact that he was a final-year student and well used to handling university pressure. 74YuGCRD and nLQ6aVus both scored rather high on the anxiety measures. Interestingly, those two participants were at opposite ends of the spectrum in terms of instructional gains in this study. Especially the former's high class anxiety ratings should be discussed further in the next section as this participant was another final-year student. nLQ6aVus, on the other hand, was one of the youngest participants and in her first year of university, so (in combination with other learner variables), this might offer an explanation.

Participant	use anxiety	class anxiety
MgrYyVvU	13	4
YKX5Bwvj	0	13
74YuGCRD	29	71
nLQ6aVus	50	88

Table 6.13: Use anxiety and class anxiety scores

Motivation

The motivational intensity seemed to be highest in the case of MgrYyVvU. His instrumental motivation was the lowest of the group, which might not have been a coincidence. Overall, this participant expressed the most interest in both the language and the experiment and its design. The interest seemed genuine and not driven by instrumental motivations. YKX5Bwj also scored high on motivational intensity, and even higher in terms of instrumental motivation, which might have been influenced by a native-speaker of German in her immediate environment, but this remains speculation. 74YuGCRD and nLQ6aVus showed similar scores in terms of motivational intensity as well as instrumental orientation. A closer look at the ratings for each question might show a more nuanced picture though, as will be discussed in the next section.

Participant	motivational intensity	instrumental orientation
MgrYyVvU	92	50
YKX5Bwj	71	79
74YuGCRD	54	63
nLQ6aVus	54	67

Table 6.14: Motivational intensity and instrumental orientation scores

Grammatical sensitivity and language analytical ability

The four participants scored relatively high on the English-language grammatical sensitivity test, as could have been expected. They scored, however, remarkably low on the language analytical ability test – especially in comparison with the secondary-school participants in the pilot study. Not only were there differences in the scores, as shown in Table 6.15 below, but also in the processing time. MgrYyVvU took twice as long as the quickest participant, YKX5Bwj, and was unable to complete the task. The other participants scored at a similar level, but showed differences in processing time. None of the participants got the translations from English into Pip or vice versa right, so they ultimately were not able to demonstrate that they had grasped the structures of the artificial language. Their scores could be explained with the application of general analytical abilities rather than language-specific abilities, as will be discussed in the next section.

Participant	Grammatical Sensitivity	Language Analytical Ability	Processing Time (LAA)
MgrYyVvU	75	43	20 minutes
YKX5Bwvj	88	57	10 minutes
74YuGCRD	88	52	12 minutes
nLQ6aVus	88	57	16 minutes

Table 6.15: Grammatical sensitivity and language analytical ability scores and processing times

Extraversion

As shown in Table 6.16 below, the four participants in this study were distributed across almost the whole extraversion spectrum from a rather introverted MgrYyVvU to an extremely extraverted YKX5Bwvj. A separate score has been calculated for the questions which targeted impulsivity only. Note that the most extraverted participant in general is also the most impulsive and the most introverted participant is the least impulsive, whereas the impulsivity scores for the other two participants are equal. How this might have affected the interaction with the instructional materials and questionnaires as well as the ability to keep attention on task, will be explored in more detail in the next section.

Participant	Extraversion	Impulsivity
MgrYyVvU	19	14
YKX5Bwvj	94	86
74YuGCRD	69	43
nLQ6aVus	44	43

Table 6.16: Extraversion scores

Figure 6.19 below shows an overview of different learner variables across participants. It is quite obvious that, even in such a small sample, taken from just one university programme, at one time, there is a substantial degree of variability. Interestingly, the most variation can be observed in terms of exposure to German, extraversion/impulsivity, anxiety, and attitudes.

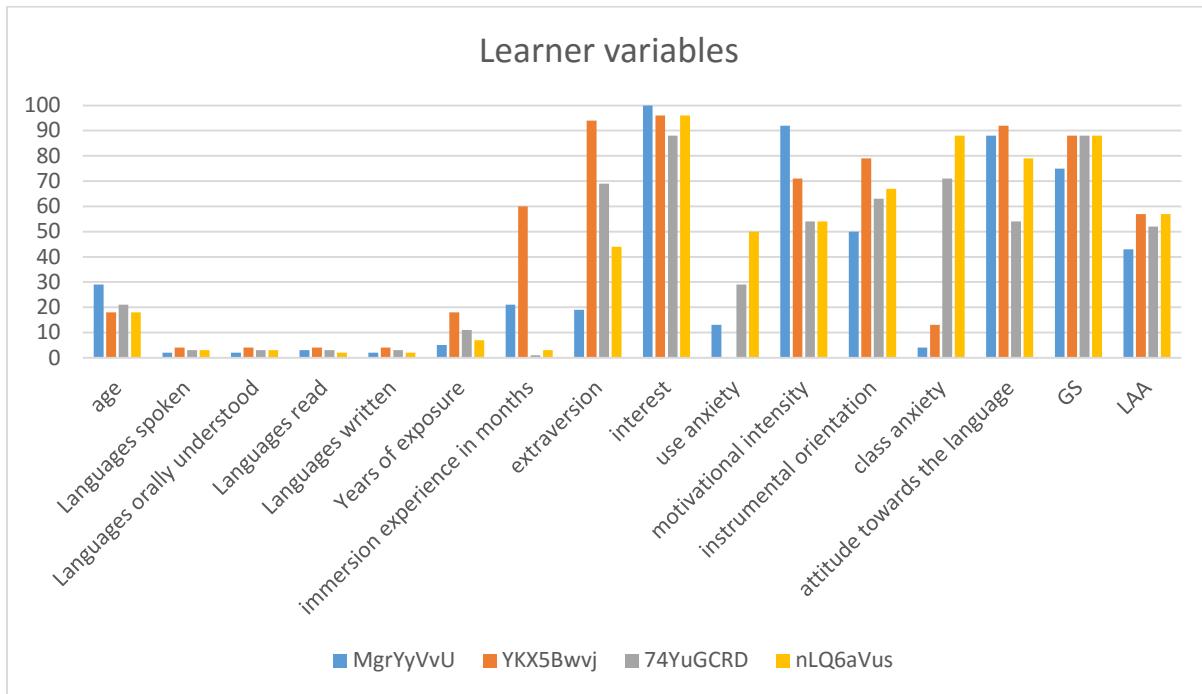


Figure 6.19: Overview of learner variables

Naturally, the participant sample in this study is too small to do any sophisticated cluster analyses. It is just the right size, however, to show how all learner variables under investigation come together as learner profiles. To this end, the raw scores for the learner variables were calculated. For those variables that had no maximum values per se (e.g. languages learnt) relative scores were calculated and converted into percentages. Figure 6.20 below shows the similarities and differences of the learner profiles. This will be discussed in more depth, especially in relationship with the instructional gains, in the next chapter. It should be noted though that the overall impression is one of similarity rather than disparity – which is in accordance with the fact that the four participants in this study were sampled from a pre-selected population of students at a selective London-based university, all of them enrolled in a language programme and selected based on their pre-test scores. There are, however, differences. Overall, the participants with the ‘denser clusters’ in terms of learner variables seem to perform better on the assessment tasks than the participant with the more diverse profile.

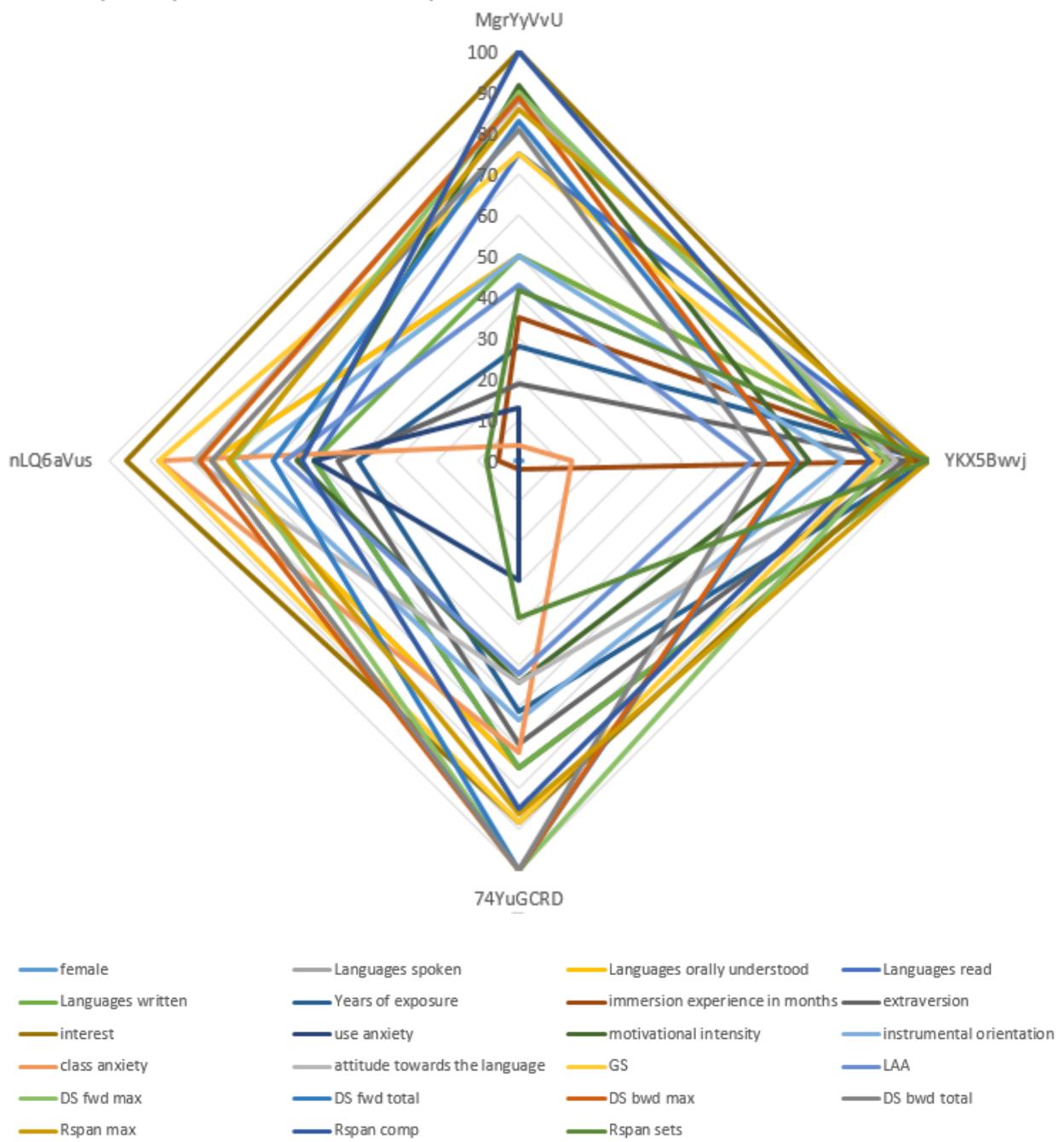


Figure 6.20: Overview of learner profiles using relative scores

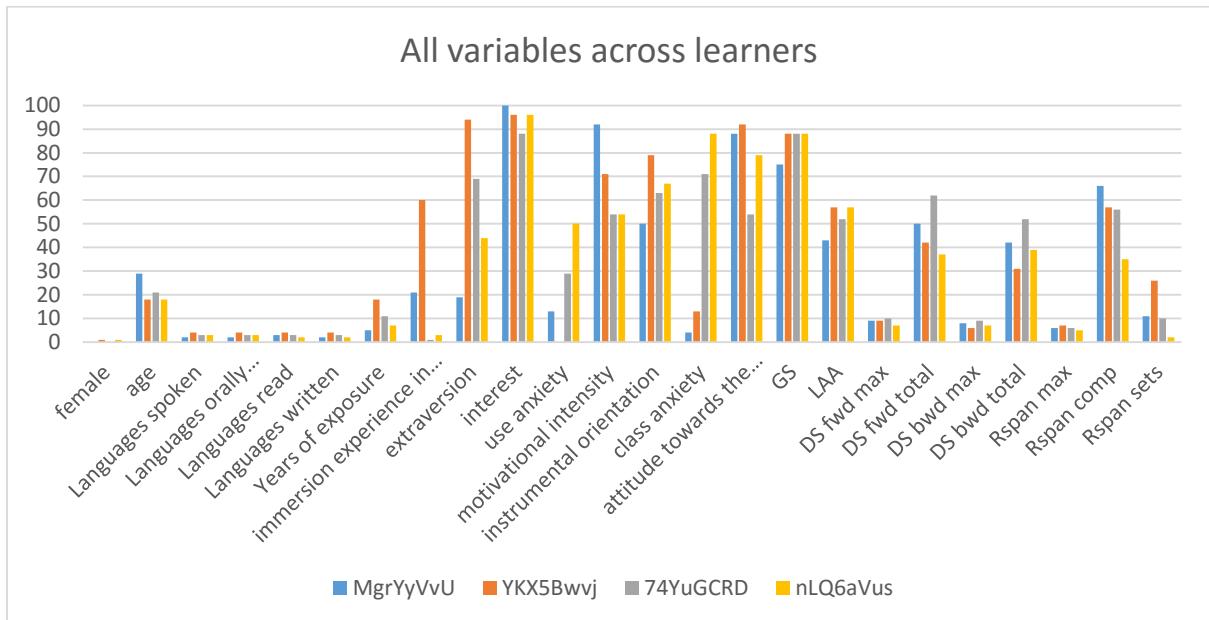


Figure 6.21: Variables across learners

Observational data

The fact that the Working Memory Span tasks were computer-delivered allowed the researcher to observe the participants while they were engaging in the tasks. 74YuGCRD showed no signs of sub-vocal rehearsal and did not look at the keyboard to visualise digit strings in the forward section of the Digit Span task, he just stared into space while listening to the digit sequences. In the backward condition, his fingers moved during the reversal of the sequence as if arranging the digits by hand. He read the instructions for both the Digit Span and the Reading Span task very carefully and made fairly quick judgements in the training trials of the Reading Span test. When the letter sets got bigger and the recall more challenging, he zoomed in on the task by leaning forward and chewing his gum more vigorously. When he could not remember a letter, he scrunched up his face, but he did not utter any sound and proceeded throughout the whole trial at a steady and quick pace. The fact that he used the space bar to indicate missing letters, shows that he was aware of how well he remembered each sequence.

YKX5BWvj also showed no outward sign of strategy use and looked straight ahead at the screen without focusing on anything in particular when listening to the digit sequences in the Digit Span task. She used her left hand exclusively and once she saw the recall prompt, she entered the digits quickly and without thinking, smiling when she could not remember them well, but entering digits anyway. All the while, she was leaning forward, with one elbow on her thigh and her face close to the laptop. In the backward condition

of the Digit Span task, the last digits to recall, i.e. the first digits she had heard, took longer to type and were cleared and re-typed more often. For the Reading Span task, she skimmed the instructions very quickly and again used only her left hand to input letters or indicate sentence judgements. She smiled and sometimes snorted when reading the non-sense sentences and leaned farther forward towards the screen when the letter sequences got longer. Like the first participant, she used the space bar to indicate a gap in the recalled letter sequence, but her sequences were correct except for the one letter she could not remember while 74YuGCRD's letter strings that contained blank spaces tended to be all jumbled.

nLQ6aVus read the instructions for the Digit Span task quickly and at first showed slight lip movements. With increasing sequence length, this stopped though and her eye movements increased. She looked at the keyboard when listening to the digit sequence, but kept the laptop at arm's length throughout the whole experiment. For the Reading Span task, she read the instructions very quickly again and also went through the training trials rather quickly. Once the test trials started, she sat up and confronted the screen full-on. After a couple of trials, she slumped and put her head in her free left hand while typing the letters and making the judgements. Periodically, as the sequence length decreased again, she would sit up straight and use both hands until the sequence length built up again and her posture became less rigid once again. She did not use the space bar to indicate gaps in the recalled letter sequence and never showed any reaction to the sentences on her face.

MgrYyVvU started the Digit Span task by reading the instructions very carefully whilst sipping herbal tea. He then got up and re-adjusted the slightly wobbly table before he started the training trials. During the forward condition trials, he closed his eyes both while listening to the digit presentation and during the recall. His lips did not move and he would occasionally massage the bridge of his nose when thinking about a digit sequence. After six trials, he stopped and reported to be confused. Then he shook his head and continued with the experiment, looking up at the ceiling when trying to recall the longer sequences. After the first two trials in the backward condition, he remarked that one could just type out the sequence one had listened to and then switch the numbers to the reverse order based on that visual aide, but that this would be cheating, so he would not do it. Instead, he nodded his head while reversing the letters in his mind and his fidgeting increased with the sequence length.

For the Reading Span task, he read the instructions very carefully and laughed about the non-sense sentences in the practice trials. He did not believe the 85% accuracy ‘trick’ that is supposed to make participants pay attention to the distractor task and continued to judge the sentences rather swiftly, as he had already done in the practice trials. When he saw the recall prompt, his eye movements indicated that he used the letters on the screen as visual cues. He kept a medium distance to the laptop, but slightly leant forward towards the later stages of the experiment. He did not use the space bar to indicate gaps in the recalled letter string.

Potential effects of the materials

To check whether the nature of the distractor sentences influenced the performance of the participants on the judgement task, the sentences were analysed regarding the location of the semantic violation within the sentence. Table 6.17 below shows the error rates for judgements of violations in sentence-initial, -medial, and -final positions as well as for correct sentences. The latter category of sentences has the lowest judgement error rate with 3%. Sentences with violations in the first phrase or last phrase surprisingly had the highest error rates with 15% and 14%, respectively. Sentence-medial violations were judged inaccurately only 9% of the time. This is somewhat unexpected as the participants’ attention should have peaked at the beginning of the sentence, then dropped around the middle, and recovered towards the end.

	sentence-initial	sentence-medial	sentence-final	no violation
sentences	7	15	17	42
occurrences	26	56	58	155
errors	3	4	8	3
time-outs	1	1	0	1
error %	15	9	14	3

Table 6.17: Position of non-sense word within the sentence and respective error rates

From the dual-task perspective, it would be interesting to see which of the violation-types described above led to the most errors in the following letter recall. Theoretically, sentences with semantic violations in the first phrase should pose the smallest processing strain on the participants’ Working Memory system as test takers would not have to read

the whole sentences in order to make a correct judgement. In contrast, sentences which contain a semantic violation in the last phrase, or are correct, should pose the greatest processing strain on the participants' resources as the whole sentence would need to be read and even then there might be doubt as to whether there was no violation or one had simply overlooked the violation. Table 6.18 below shows the position of the semantic violation within the sentence and the respective error rates in subsequent letter recall.

Note that the sentence-initial violations lead indeed to the lowest error rates. The sentence-medial violations show the highest error rates in terms of letter recall and the sentence-final violation and no-violation conditions show similar effects on the subsequent letters. It should be noted that the order of sentence presentation varied from participant to participant and that not every sentence was presented to every participant (for a detailed list, see Appendix C). So other factors, such as fatigue or the interference effects of previously recalled letters might have played a role as well – and to a different extent across participants. In addition to that, the “recall style” of the participants varied: 74YuGCRD, for example, showed an all-or-nothing approach where he recalled either long, perfect sequences, or got all letters wrong. This would have led to an inflation of errors for sentence-medial and -initial violations. Other participants only recalled individual letters incorrectly and their patterns are closer to the predictions that the more they had to process of the sentence, the more it affected their letter recall accuracy: MgrYyVvU, the most accurate scorer, perfectly recalled 100% of the letters following sentence-initial violations, 92% of letters following sentence-medial violations, and 87% and 84% of letters following sentence-final violations or correct sentences, respectively.

	sentence-initial	sentence-medial	sentence-final	no violation
sentences	7	15	17	42
occurrences	26	56	58	155
LETTER errors	5	17	15	44
error %	19	30	26	28

Table 6.18: Position of non-sense word within the sentence and respective letter recall error rates

6.4 Putting the findings in perspective

Reading Span and Digit Span scores

All in all, the rather high Working Memory Capacity scores measured in this study seem to be in accordance with the background of the participants: They are university students enrolled in a rather selective public research university. Their educational history has prepared them for test taking. The findings for the Working Memory Capacity scores are also in line with the psychometric literature insofar as they support the observation that composite scores are more sensitive to individual differences than absolute span scores as the latter can only take very few discrete values. It should also be noted that the participant rating would have been different if absolute span scores had been used. One could argue that the Reading Span task is only a complex span task if the test takers pay attention to both recall and sentence judgement. To assure this condition, psychometric experiments usually have a cut-off score of 85% accuracy on the judgement task and participants who score lower are excluded from the experiment. This then means that all other trials count, no matter whether the distractor sentences were judged correctly, or not.

Thanks to the small number of participants in this study, it was possible to have a closer look at the data. The analyses revealed that the most sensitive measure of individual differences in Working Memory Capacity in this study was the composite score of all complete sets that were also 100% accurate in terms of sentence judgements. As mentioned before, this scoring technique does not make optimal use of the available data because it dismisses those perfectly recalled letters and sets that are accompanied by inaccurate judgements. While the more rigid approach might disregard individual letters that have been recalled in their correct position, it rewards repeated complete recall of sets with optimal performance on the distractor task – the most challenging task condition. It is therefore not surprising that this scoring technique shows the individual differences between Working Memory Capacity scores best.

With regards to the trade-off between processing and storage, the results of this study are in line with Kane et al.'s (2004) observation that participants with the highest processing accuracy usually demonstrate the highest storage capacity as well. While the present participant sample was too small to gauge the statistical significance of the

differences, it can be observed that the participant rankings are the same for both storage capacity (i.e. all letters recalled) and accuracy of judgement.

For the Digit Span task, the same observation can be made in terms of scoring: The composite scores are more sensitive to individual differences than the absolute span sizes. The participant rankings, however, differ from the Working Memory scores. This is not surprising as the Digit Span task is not supposed to measure the same construct as the Reading Span task. On the Digit Span task, MgrYyVvU, the participant with the highest Working Memory Span scores, does not perform as well as 74YuGCRD, the participant with the second-highest Working Memory Span scores. This might have had implications for some of the instructional materials. Note that the Digit Span Task provided aural input and is generally considered to be linked to the Phonological Loop or Phonological Short-Term Memory. 74YuGCRD's very good performance on the Digit Span task might explain why he did either better than the other participants on the aural-input Structured Input Activities or as well as other participants – despite the fact that he had received no German input during his year abroad in France.

The participant with the lowest Working Memory Capacity scores, nLQ6Vus, also had the lowest forward Digit Span scores, but the lowest backward Digit Span scores has YKX5Bwvj. The latter scored rather high in terms of absolute Working Memory Capacity, but her scoring was overall more inconsistent. This observation from the Working Memory tasks is also reflected in her scoring throughout the aural Structured Input Activities. Considering that she is bilingual and grew up in Germany, she misinterpreted three out of ten sentences in each of the two aural activities. This might point to a weaker Phonological Loop or Phonological Short-Term Memory. The Digit Span Task, however, has also been linked to attention. The overall more inconsistent scoring before, during, and after instruction could therefore also be connected with fluctuating degrees of attentional control.

Processing Instruction

The computer-delivered Processing Instruction, especially the Structured Input Activities that revised the accusative and dative cases respectively, elicited ceiling-level performances. Those activities were also perceived as easy by the participants, which is in tune with the claims of VanPatten (2007, 2002) regarding the maximally reduced complexity of the input. The first task that required the correct interpretation of case

markings in relationship with switch prepositions, the **picture-match task**, was perceived as challenging in the first session, but easy in the second session. Apart from the instruction itself, task familiarity might thus have played a role in the participants' perception of the difficulty level as well as in the improved performance in the second session. Impulsivity also might have affected the scores as those participants with lower impulsivity scores, i.e. those who tended to think things through before reacting to them, made more incorrect picture-sentence matches than the most impulsive participant, YKX5Bwvj. It should be noted that the feedback that the participants received might also have influenced the difficulty perceptions as those participants who received massed positive feedback tended to rate the task as less difficult than those participants who had received mixed feedback. Whether this relationship is caused by the feedback or the participant's awareness of how consistently they had been scoring, is a question which cannot be answered in this study though.

The **sentence-level comprehension** activity with textual input went very smoothly in both instructional sessions and participants again performed at ceiling level. They also repeatedly rated the task as easy, adding to the impression that Structured Input Activities are indeed not very taxing for the participants' attentional and/or Working Memory resources. A similar activity with aural input, however, elicited more varied performance scores. As mentioned above, 74YuGCRD, the participant with the least contact to German native speakers and the least amount of input prior to the experiment, performed best on this task. This might have had to do with an exceptionally good Phonological Loop or Phonological Short-Term Memory on his part. The two participants with the most German input and the longest periods of immersion under their belt performed equally well overall, with MgrYyVvU scoring slightly more consistently than YKX5Bwvj. The participant with the lowest scores in terms of Working Memory and Phonological Short-Term Memory got four out of ten sentences wrong, which is only slightly above chance-level. This is only partly reflected in the difficulty ratings: nLQ6aVus and MgrYyVvU rate the task as moderately difficult, whereas the other two participants rate it as easy. Keeping in mind that both MgrYyVvU and YKX5Bwvj scored equally well, the perception might be influenced by their personality variables, with the more impulsive participant rating the task as easy and the more reflective participant being more critical.

The **discourse-level Structured Input Activity** revealed a surprising discovery: nLQ6aVus, the participant with the lowest Working Memory Capacity scores, was the only one to answer all ten questions correctly. 74YuGCRD and MgrYyVvU got one target question wrong each and YKX5Bwvj, the bilingual participant, got one target question and

one content question wrong. While the text for this exercise was presented paragraph by paragraph and potentially problematic vocabulary was translated, one would have expected a different participant ranking nevertheless. We will get back to this point when discussing the discourse-level production tests. It should be noted that for this multiple-choice comprehension activity, the participants did not receive feedback. While most participants rated the difficulty level as easy, 74YuGCRD rated it as moderately difficult. This does not seem related to his performance as he was one of the best scorers on this task. Yet the bigger portions of textual input might have influenced this perception. On the other hand, the amount of context and the coherence of the text might have helped with the interpretation and made the questions less difficult than in isolated sentences where no additional information is available. This would then also be a possible explanation for nLQ6aVus's unexpected performance.

The **Direction versus Location** task was introduced in the second instructional session and elicited a more varied picture. MgrYyVvU, who had just spent a year in Berlin, interpreted all sentences correctly and YKX5Bwj, the bilingual, made one mistake. Those two participants had the longest periods of immersion and the most intense contact with native speakers of German. nLQ6aVus chose three wrong answers out of 15 and thus demonstrates a higher performance level than in the first session. It is therefore possible that her lower Working Memory Capacity led to a lower learning rate, but did not necessarily prevent her from improvement. It is not quite clear why 74YuGCRD performed the way he did on this task, with five wrong answers out of 15. Even if he had not been familiar with the difference between directive and stative verbs, the question pronoun should have given him a reliable clue. Considering that his performance on the subsequent task, at the end of the second instructional session, was similar, it is possible that the drop in performance scores is related to fatigue.

The **Aural Sentence Puzzle**, another new task in session two, generally did not elicit ceiling-level performances like other Structured Input Activities. YKX5Bwj and nLQ6aVus got three of the 18 sentences snippets wrong and the other two participants got four snippets wrong each. The slight drop in general performance might be an indicator of fatigue, as suggested before. It should be noted, however, that the performance rates of nLQ6aVus, the low-span participant, are again above average, supporting the hypothesis that limitations in Working Memory Capacity lead to a decrease in speed of acquisition, but not necessarily to a disadvantage in terms of ultimate attainment. The differences in performances, however, are small and that again points to the conclusion that the actual Structured Input Activities in Processing Instruction elicit

similar performance scores from participants across the Working Memory Span range. This would have to be empirically tested with bigger participant samples, of course.

All in all, it seems that individual differences in Working Memory Capacity have not played a role for the performance on Structured Input Activities and for the participants of this study. Interestingly, but not entirely surprisingly, the performance dynamics seem to be linked to the level of impulsivity on the part of the participants. It is important to note though that the effects of impulsivity or reflection differed from task to task, meaning that this personality trait can work both ways and might not be as predictive for learning success in the longer run and/or across tasks. The question now is whether the participants' relatively similar performance level towards the end of the computer-delivered Processing Instruction also translates into similar post-test scores.

Instructional gains

After being exposed to two fifty-minute sessions of computer-delivered Processing Instruction, i.e. 50 tokens of the target feature in the first session and 54 tokens in the second session, the participants were given a short break and then started on the immediate post-test. The delayed post-test, as described before, was administered 15 weeks after instruction, which is a fairly late assessment. It should also be kept in mind that none of the tests resembled Processing Instruction materials as such, meaning that the test materials were potentially more challenging, ambiguous, and complex than the instructional materials.

For the **fill-in-the blanks task**, two different trajectories can be observed: Three participants demonstrated an increase in scores on the immediate post-test and then dropped below the pre-test level on the delayed post-test. MgrYyVvU and 74YuGCRD, the two participants with the highest Working Memory scores, showed almost identical trajectories, whereas nLQ6aVus (the participant with the lowest span scores) had started at a higher level, but dropped steadily from test session to test session until her scores were below those of the other participants. Interestingly, she approached the task with the most preparation, writing down case marking paradigms in the margins of the test paper and explicitly looking up the available gender information. The other trajectory, displayed by YKX5Bwvj, is one of steadily increasing scores from test session to test session. This is a picture that can be observed in the translation task as well. It might have to do with the probability that the bilingual participant YKX5Bwvj was exposed to a

different quality and/or quantity of German input between test sessions, considering that half her family consists of German native speakers. However, instruction seems to have had an impact on three out of four participants, at least in the short run immediately after instruction. In the long run, after 15 weeks without formal instruction on the target feature, there seems to be no trace of any impact. Exposure to German input at the time of instruction and/or the testing sessions, however, might have had an influence. But then, again, this is only a small sample and these impressions need empirical testing.

The **picture description task** was designed to elicit switch prepositions, but it was a less controlled task and the participants could avoid structures they did not feel comfortable with. As in the rather open essay task, 74YuGCRD performed at a steady level and with very little variance, producing very similar, if not identical output. nLQ6aVus also did well, improved from pre-test to post-test, and kept her scores at 100% even 15 weeks after instruction. For MgrYyVvU, there is similar improvement after instruction, but he cannot maintain the perfect score. YKX5Bwvj, the bilingual participant, started with 50%, improved to 100% after instruction, then dropped down to 20%. Considering the simplicity and degree of freedom of this task, this fluctuation in scores is rather puzzling. Granted, the task only comprised on average five sentences, so one error translated into a 20% drop, but getting four out of five sentences wrong is quite substantial and cannot be explained away with a lapse in concentration. Keeping in mind the nature of the task as well as the narrow range in performance scores, it is fair to say that variation in task performance is probably not caused by individual differences in Working Memory Capacity. The performance differences are more likely to stem from the participants' form on the day or even boredom, seeing that they described the same picture three times in the course of 15 weeks.

As predicted, the **translation task** elicited the greatest performance differences across all sessions. As described in the materials section, this task combined the manipulated word order of Structured Input Activities with more distracting elements, such as difficult vocabulary and ambiguous attachments. Individual differences in Working Memory Capacity were most likely to show here. Again, there were two trajectories. YKX5Bwvj and 74YuGCRD improved from session to session, the only difference being that YKX5Bwvj started with a lower score and did not quite reach 100% on the delayed post-test. Again, there were only seven target sentences, so misinterpreting one sentence already caused a drop by 14%. The other trajectory was an improvement from pre-test to post-test, followed by a drop from post-test 1 to post-test 2. nLQ6aVus misinterpreted all OVS sentences in the first testing session, then got two right before dropping back to one.

It seems quite clear that this participant, the one with the lowest Working Memory Capacity scores, still relies on word order rather than case marking when interpreting the sentences. She also produced the least faithful translations, which varied a lot over time. MgrYyVvU started with five correct OVS interpretations, then improved to seven out of seven, and dropped down to four out of seven in the last testing session.

Again, all participants showed higher performance scores after instruction, but while two participants improved from session to session and also improved in terms of the quality of their translations, the other two could not maintain their gains. It should be noted that the participant ranking in terms of Working Memory Capacity scores is identical with the ranking of performance scores on the immediate post-test. This might suggest, again, that Working Memory Capacity is more predictive of immediate effects of instruction than the learning success in the longer run. But this, of course, needs empirical testing. Another observation worth making is that the computer-delivered Processing Instruction apparently had a positive effect on the interpretation of OVS word orders even in very complex sentences.

The **essay task** was another task with a high degree of freedom for the participants. It is unfortunate that the question in the pre-test session differed from the question in the post-test sessions, but as mentioned before, the material testing sessions elicited similar numbers and uses of switch prepositions by both advanced-level learners of German and native speakers, for both questions. As in the picture description task, 74YuGCRD produced very similar essays in terms of length and complexity and he produced the same amount of correct switch prepositions in each testing session. The term “correct switch prepositions” refers to the use of the right preposition in combination with the right case for the following noun phrase. There were rather few instances of switch prepositions in combination with the wrong case overall, but there were a few wrong prepositions in use. All of the participants improved from pre-test to immediate post-test in terms of number of correct switch prepositions used, overall grade, and overall error rate. YKX5Bwj is the only one who maintains the gains and even uses more correct switch prepositions while making fewer mistakes overall in the delayed post-test session. Again, being bilingual, she might have been exposed to a different kind of input than the other participants. It could also be argued that this type of essay is used in schools rather than at university level and that the two first-year students might have been more familiar with the task.

Overall, the accuracy improved slightly (or, in the case of MgrYyVvU, dramatically) after instruction. Neither test shows significant differences in terms of the overall grades. But the error rates and the number of correctly used switch prepositions differs both between participants and across sessions, pointing to positive effects of instruction, even if they could not be maintained 15 weeks after the second session. Both essay topics should have been similarly relatable for university students and there is no reason why the university experience essay should have elicited more switch prepositions than the holiday essay.

Looking at the pre- and post-test scores in relationship with the Working Memory Capacity scores, it becomes clear that the role of Working Memory varies from task to task. This might have been due to the nature of the task, or the individual differences in terms of learner variables, or a combination of the two. If we look at the participant rankings of this small sample, as shown in Table 6.19 below, then the Working Memory Capacity ranking is most closely mirrored by the translation task's pre- and immediate post-test scores. This makes sense as the translation task was the most challenging task. The picture description task, presumably the easiest task, produces the most stable rankings, even if they are different from the Working Memory Capacity rankings. Again, this makes sense as this easy task elicited many ceiling-level performances and therefore was not suited to highlight performance differences. The essay task could have been considered as a more complex task as the participants had been required to produce lengthier texts, which should have required planning as well as the production of more sophisticated structures.

On the other hand, participants were free to avoid structures they were not feeling confident about and the questions that were aimed at eliciting switch prepositions might have pre-structured the writing. In addition to that, the act of writing makes it easier to go back and check what one has mentioned before, so it takes some of the memory load off the participants. In addition to that, there is the tendency for novice writers to reproduce content from Long-Term Memory rather than manipulating said content to make it appropriate for their communicative goals and target audience (Galbraith, 2009: 9). For those participants who only recently left school and were familiar with this sort of writing, some of the most common phrases might have been stored as chunks in Long-Term Memory and therefore might not have been limited by Working Memory Capacity constraints. Which might explain why the participant with the lowest Working Memory Capacity produced consistently well-formed essays with a high number of switch prepositions. The fact that she produced whole case marking paradigms for the fill-in-the-

blanks task is in line with the hypothesis that this learner relied more on Long-Term Memory than Working Memory.

The fill-in-the-blanks task, the most controlled of all tasks, elicited similar trajectories from the two participants with the highest Working Memory Capacity scores, but also showed a steady improvement of YKX5Bwvj across all sessions, which means that she outperformed the others on the last post-test, thus creating a different ranking. All in all, it seems that the nature of the task played a role in eliciting performance differences. It should also be noted that some of the tasks only comprised a low number of tokens, meaning that one mistake or a lapse in attention was costly and could have influenced the ranking. In addition to that, the learner variables might have influenced the performance, as will be discussed in the next section.

	Pre-test	Post-test 1	Post-test 2	Working Memory
Fill-in-the-blanks	1. nLQ6aVus 2. YKX5Bwvj 3. MgrYyVvU 74YuGCRD	1. 74YuGCRD 2. MgrYyVvU 3. YKX5Bwvj 4. nLQ6aVus	1. YKX5Bwvj 2. MgrYyVvU 74YuGCRD 3. nLQ6aVus	1. MgrYyVvU 2. 74YuGCRD 3. YKX5Bwvj 4. nLQ6aVus
Picture description	1. 74YuGCRD 2. nLQ6aVus 3. MgrYyVvU 4. YKX5Bwvj	All score 100%	1. 74YuGCRD nLQ6aVus 3. MgrYyVvU 4. YKX5Bwvj	1. MgrYyVvU 2. 74YuGCRD 3. YKX5Bwvj 4. nLQ6aVus
Translation	1. MgrYyVvU 2. 74YuGCRD 3. YKX5Bwvj nLQ6aVus	1. MgrYyVvU 2. 74YuGCRD 3. YKX5Bwvj 4. nLQ6aVus	1. 74YuGCRD 2. YKX5Bwvj 3. MgrYyVvU 4. nLQ6aVus	1. MgrYyVvU 2. 74YuGCRD 3. YKX5Bwvj 4. nLQ6aVus
Essay (switch prepositions)	1. 74YuGCRD nLQ6aVus 3. YKX5Bwvj 4. MgrYyVvU	1. nLQ6aVus 2. YKX5Bwvj MgrYyVvU 4. 74YuGCRD	1. YKX5Bwvj 2. nLQ6aVus 3. 74YuGCRD 4. MgrYyVvU	1. MgrYyVvU 2. 74YuGCRD 3. YKX5Bwvj 4. nLQ6aVus

Table 6.19: Participant rankings for test scores and overall Working Memory Capacity scores

Learner variables

One background variable that has not been discussed in detail before is the **linguistic background** of the participants. All of them are language students, but their linguistic experiences differ. YKX5Bwvj spent her early childhood in Germany and is studying Russian, so she has extensive experience with case-marking languages, considering that

she has acquired German in an immersive context and learnt Russian, a language which has no fewer than six cases. Both her linguistic abilities and her metalinguistic knowledge should have set her up for success in this experiment. But while she improves steadily on several measures, she is overall not consistent in her scoring – an observation which will be discussed further in the personality variables section. nLQ6aVus's first language is Hindi, an SOV language with case marking. She too should have had a linguistic advantage over the other two participants who are native speakers of English, which has no morphological case marking except for some residual markings on personal pronouns. 74YuGCRD studies French and German, MgrYyVvU had French in school and studies German. So both were exposed to morphological case marking for the first time when learning German. This should have made the acquisition process harder. Yet, both are rather successful, especially in the immediate post-test session. This could be an effect of their larger Working Memory Capacity and, especially in the case of 74YuGCRD, the well-developed Phonological Short-Term Memory. Why one of the scores more consistently than the other will be discussed in the personality variables section.

The **interest** scores of the participants were rather high, which makes sense as all of them are studying German. It seemed though that the **attitude** towards German was highest in those participants who had spent the most time in Germany or with German native speakers. This high attitude score did not directly translate into similar performance trajectories in this study, however. While it makes sense to assume that “those with positive [attitudes; S.P] would be more attentive in the learning situation, would take assessment more seriously, would find it more rewarding to simply experience the language, and thus achieve more” (Gardner, 1985: 41), the findings in this study point to an interaction of several variables, with attitude just being one of them. The approach to assessment, for example, seems to be influenced by impulsivity rather than attitude towards the language. Of course, interest and attitude do not automatically translate into high performance scores or proficiency levels unless learners actually put in the effort and perform goal-directed behaviour on a regular basis. In this sense, it is **motivation** that makes learners persist when faced with adversity or setbacks. Generally, the higher the **motivational intensity** was in a participant, the lower was their **instrumental orientation**. These two orientations, one directed towards the language itself, the other directed towards the language as a means to an end, seem opposed. Additionally, the instrumental orientation was highest in the participant who reported to be speaking the most languages and lowest in the participant who only speaks English and German. This seems to fit with Gardner's (1985:134) observation that native speakers of English

usually have little instrumental incentive to learn another language. So if they choose to learn another language, it is for reasons other than instrumental motivation.

Again, the analysis of individual questionnaire items reveals important differences: 74YuGCRD and nLQ6aVus score very similarly in terms of motivational intensity and instrumental orientation, but while 74YuGCRD reported that he is not working hard for German, nLQ6aVus reported to be working hard – and she does. She is the only participant who produced the whole case-marking paradigm in the tests, preparing and pre-structuring tasks. She also reported though that she gives up quickly when things do not make sense and that she procrastinates. Combined with her low Working Memory Capacity scores, this draws the picture of a hard-working student to whom learning German might not come naturally, but who does their best to compensate for a lack in processing abilities. This might be an explanation for her better-than-expected scores in the less controlled tasks with ample planning time.

The Attitude/Motivation Test Battery differentiated between **use anxiety** and **class anxiety**. Interestingly, there was a wide range of anxiety scores in this study. YKX5Bwj, the English-German bilingual participant, expressed no use anxiety at all when talking to native speakers of German, which is not surprising as some of those native speakers are part of her own family. She did, however, express some class anxiety, which also makes sense as she had just started university at the beginning of the experiment. MgrYyVvU, a rather mature final-year student, expressed very low class anxiety, but some use anxiety. According to (Dewaele, 2013), anxiety peaks in the learners' twenties, so his overall rather low scores fit. 74YuGCRD expressed moderate use anxiety, but scored rather high on the measure of class anxiety. The fact that his use anxiety is lower might be related to the fact that he has very little contact with German, so he hardly ever finds himself in anxiety-inducing situations outside the classroom. nLQ6aVus expressed moderately high use anxiety and very high classroom anxiety. She had spent some time in Germany for work internships, so this might explain the higher use anxiety scores. At the time of the questionnaire administration, she had just started university, so it makes sense that her class anxiety was extremely high at that point. It would have been interesting to see how she would have scored at a later point in time. All in all, it seems that both the measurement of anxiety and the effect of anxiety on the learning process are highly context-dependent

For the personality variables, the **extraversion** scale of the Eysenck Personality Inventory was used. Additionally, all items relating to impulsivity were analysed

separately, which led to a more complete picture. **Impulsivity** played a role in the way the participants were interacting with the materials. The most impulsive participant, YKX5Bwvj, was very quick to tackle all tasks and did not stop to think and plan. For the Reading Span task, this meant that her very quick training times led to time-outs in the trial conditions, thus directly affecting her scores. On the other hand, negative feedback during instruction did not seem to faze her. Overall, she did not take herself or the experiment too seriously. The least impulsive participant, MgrYyVvU, took a lot more time to read instructions carefully, think them through, and ask questions. Compared to other participants, he produced shorter texts in the same period of time and got more easily confused, especially on the picture-match task and Working Memory test. Maybe his high Working Memory Capacity scores explain why he still performed rather well, especially during and immediately after instruction. While 74YuGCRD and nLQ6aVus scored at the same level in terms of impulsivity, the former is a little more extraverted than the latter. On the one hand, both work through tasks at a quick pace and don't overthink their answers. On the other hand, both score rather high on class anxiety. This might be connected with their lower impulsivity scores.

As mentioned before, anxiety usually has more of an effect in oral, spontaneous conversation. Lower impulsivity means that individuals care more about their answers, how well they are doing, and how others perceive them. Together with the novelty of the university experience in the case of nLQ6aVus, this self-reflection might lead to higher class anxiety scores. But again, the personality variables do not exist in isolation. A very high impulsivity score might have had a more detrimental effect on someone with a weaker language background than YKX5Bwvj's and MgrYyVvU's reflectivity might have had a bigger impact on someone with a smaller Working Memory Capacity. While the alternative outcomes cannot be known, it still seems that personality has had an effect on the instruction and its outcomes in this study, which goes against earlier findings (Leino, 1972: 38) that led researchers to believe that personality factors only played a minor role in classroom-based language learning since the learners did not have much opportunity to express themselves in a teacher-led and -controlled environment.

The **grammatical sensitivity** scores were rather high, as could be expected from language students who had received many years of formal instruction. The **language analytical ability** scores, however, were rather low. Marked differences could also be observed in terms of the time it took the participants to solve the artificial language problems. The most impulsive participant, YKX5Bwvj, was the quickest with ten minutes, and one of the most accurate (57%). The most reflective participant, MgrYyVvU, took

twice as long and did not even complete the test. nLQ6aVus approached the artificial language very methodically and managed to score 57% within 16 minutes. 74YuGCRD was slightly quicker (12 minutes), but less accurate (52%). None of the participants managed to translate correctly from one language into the other, indicating that they had not grasped the case and number marking rules of the artificial language, but relied on word order. As mentioned before, this could be taken as evidence that the participants demonstrated general analytical rather than language analytical skills when doing the artificial language test. It has been noted though that the inductive language learning ability as operationalised in the traditional aptitude tests bears conceptual similarities to general reasoning (Gardner, 1985: 19). The fluctuations in validity scores reported in the language aptitude literature furthermore suggest that the construct measured by tests such as the Modern Language Aptitude Test (Carroll and Sapon, 1959) correlate with language learning success to different extents, depending on the participant sample, the learning context, and the assessment methods (Gardner, 1985: 22–3). Again, it would have been interesting to see whether any improvements could have been demonstrated after the Processing Instruction treatment. It seems safe to say though that the language analytical abilities that would have been necessary for the solution of the artificial-language puzzle apparently were not required to be successful during Processing Instruction.

6.5 Recap of methodology and context

This small case study with four university students enrolled in a German programme at a London-based public research university explored the effects of individual differences in Working Memory Capacity on the participants' performance before, during, and after two 50-minute sessions of computer-delivered Processing Instruction. Prior to instruction, the participants had their Working Memory and Phonological Short-Term Memory Capacities measured and provided additional information on a few selected learner variables by answering questionnaires. These variables had been selected because previous research had indicated that they might have a mediating effect on the learners' language processing resources. For this study, the following constructs were explored in relationship with the participants' background, their memory scores, and the pre- and post-test performance scores: attitude towards German and its speakers, motivation and

orientation, use and class anxiety, extraversion and impulsivity, grammatical sensitivity and language analytical aptitude.

In the week after the pre-test, the participants to be included in this study were exposed to 50 minutes of computer-delivered Processing Instruction (comprising 50 interpretations of German accusative/dative case markings). In the following week, the participants received a second 50-minute session of computer-delivered Processing Instruction (consisting of 54 interpretations of German accusative/dative case markings following switch prepositions). At no point during the instructional sessions were participants required to produce any German output. The activities were designed in accordance with the guidelines for Structured Input Activities, keeping the default processing strategies of the learners in mind and forcing the learners to process grammatical form in order to derive meaning. The Structured Input was presented aurally and visually, in sentence-level and discourse-level activities. For all sentence-level activities, the participants received immediate feedback on the accuracy of their interpretations. There were no ambiguous sentences. After each completed activity, the participants were prompted to rate the difficulty level of the exercise before moving on to the next task.

Four tests were administered one week before the first Processing Instruction session, immediately after the second session, and again 15 weeks after instruction. A fill-in-the-blanks task required the participants to fill in the missing definite articles in the correct case. A picture description task prompted the participants to produce switch prepositions to describe the locations and motions of objects and people in the picture. A translation task forced the participants to interpret the subject and object roles within sentences while translating difficult words and complex sentence structures. An essay task asked the participants to write about their ideal holiday (pre-test) or ideal university experience (post-tests) – which was supposed to elicit a number of German switch prepositions in a discourse-level context. In addition to measuring the performance levels pre and post instruction, on-task performance scores were recorded for the individual participants as they were working through the Structured Input Activities.

For the analysis of the Working Memory and Phonological Short-Term Memory Capacity scores, several scoring techniques were employed to find the one that was most sensitive to individual differences. In the next step, the performance data collected before, during, and after instruction were analysed. Following that, the additional learner variables were presented and supplemented with the observational data that the researcher gathered during the Memory Span experiments. In a last step, potential effects of the materials on

the participants' task performance were explored. A tentative explanation of the findings was provided and will be discussed with regards to previous findings reported in the relevant literature in the next chapter.

7 DISCUSSION

Working Memory is a central construct in psychology and has been linked to a variety of higher-order cognitive abilities. Especially individual differences in Working Memory Capacity have been linked to “a significant portion of variance in general intellectual ability” (Conway, et al., 2005: 769). Processing Instruction, on the other hand, has been claimed to pose only a minimal strain on the learner’s processing resources, leading to the hypothesis that this kind of pedagogical intervention should be equally beneficial for learners of different Working Memory Capacities. This study has measured Working Memory and Phonological Short-Term Memory Capacity using two different span measures and several scoring techniques in order to explore their relationship with the outcomes of instruction. A special focus has been on potential mediating learner variables.

For the first time in Processing Instruction research, a small-scale case study has been conducted, allowing the researcher to look at individual learner profiles and the interplay of different factors associated with individual differences in foreign language learning. In a mixed-methods approach, pre- and post-test scores have been collected together with Working Memory Capacity scores and qualitative data on learner variables. The findings presented in the previous chapter, while small in scope, have shed light onto the potential connections between several factors. The purpose of this chapter is to discuss the findings against the theoretical and methodological background of the study and to interpret the results in terms of the MOGUL framework presented in Chapter 1. For the participant sample in this study, the previously stated research questions can be answered as follows:

7.1 Overview of research questions and hypotheses

1. Do learners benefit from Processing Instruction in equal measures, regardless of their Working Memory Capacity?

- a) Processing Instruction is maximally reduced in complexity, yet there usually are individual differences in terms of the post-test scores. Differences in Working Memory Capacity will result in differences in post-test scores.

Yes.

The participants performed at a comparable level throughout the Processing Instruction treatment and (task-novelty effects aside) rated the difficulty level of the Structured Input Activities as ‘easy enough’. Performance scores increased from the first to the second session, with the task features (i.e. textual or aural input, sentence-level or discourse-level tasks) having differential effects on the individuals. Even the low-span participant reached ceiling-level performance scores, and scoring inconsistency does not seem to be the effect of individual differences in Working Memory Capacity. Processing Instruction therefore seems to be beneficial for learners at all points of the Working Memory Capacity spectrum. For the assessment tests, however, individual differences in Working Memory Capacity seem to play a role, especially immediately after instruction. For the most challenging task, the translation task, the Working Memory Capacity rankings seem to be reflected in the performance scores. In the longer run though, other factors, such as the exposure to German input, seem to play a bigger role in the successfully interpretation and production of German switch prepositions.

2. Are there other learner variables that potentially mediate Working Memory Capacity and thereby affect the learning outcomes?

b) There might be other variables that compensate for lower Working Memory Capacity scores. If any of the measured constructs have an impact on post-test scores, corresponding patterns should show up in the individual learner profiles.

Yes. Those factors can work both ways and are interconnected.

Considering all the learner variables that were explored in this study, it seems that the person with the densest cluster performs best in the long run. Her scores, however, lack consistency due to her extremely high degree of impulsivity. The participant with the highest degree of reflectivity, whose Working Memory resources should have been impaired by self-related processing, could afford to engage in self-related cognition thanks to his large Working Memory Capacity. The participant with the lowest Working Memory Capacity scores and the highest anxiety levels overall performs less well than her learning efforts would warrant. The participant with the least exposure to German and/or German native speakers as well as the lowest motivation scores still performs well, presumably due to a large Working Memory Capacity, very good Phonological Short-Term Memory, and a balanced nature (in terms of extraversion, impulsivity, and anxiety levels).

3. Are there participant-treatment interactions?

c) It is conceivable that different individuals react to and interact with the instructional and experimental materials in different ways. This could influence the learning outcomes. Evidence for this might be found in the participants' approach to the less controlled tasks, in the way participants work with the materials, and in the observational data the researcher collects.

Yes.

Most obvious is the fact that the degree of impulsivity expressed itself in the way the participants interacted with the materials. The speed with which the activities were performed, the consistency of the scoring, as well as the effects of positive or negative feedback on the participants' perception of the task difficulty seem to have been impacted by how impulsive or reflective the individual was. In addition to that, effects of the materials could be observed in terms of the modality of the input (textual versus aural), the length (sentence-level versus discourse-level), the degree of freedom (controlled versus open), and the complexity of the task. This has had an impact not only on the way learners interact with the materials, but it could also have directly influenced the measurement of other learner variables and the instructional gains.

7.2 Summary of theoretical and research contexts

In this section, the research contexts of the individual learner variables will be highlighted in brief and summarised to point out how the findings of this case study, though not readily generalisable to other samples, fit into the wider research context. Considering that each of the learner variables selected for this study has received substantial attention at some point in time, and often in different disciplines, the presentation of the relevant research contexts cannot be exhaustive. Overall, the findings on learner variables are in line with Robinson's (2001: 385) observation that measures of aptitude are only relevant predictors for the tasks they feature in, i.e. a global prediction, even for a specific instructional context, is not possible. "It follows that for groups of learners with more differentiated abilities, there will be more variation in learning in any one environment, or on any one task (i.e. less matching of abilities and processing demands) than for groups of learners with less differentiated abilities." (Robinson, 2001: 386)

As mentioned before, this small participant sample has exhibited a wide range of (language learning) backgrounds and predispositions, and there is a multitude of individual differences that are subtler in nature and therefore even harder to capture than the ones chosen for investigation in this study. According to the Modular On-line Growth and Use of Language (MOGUL) framework, the learner's processing of language representations is shaped by the input they have processed previously. In a classroom context, this input does not necessarily have to be authentic. Depending on the structures represented in textbooks and used by the instructors, representations can develop which are based on the overuse of certain features, for example the overt pronoun in pro-drop languages or SVO word order in German. OVS word order is a lot more common in colloquial spoken German, where the first phrase receives natural stress, which provides an advantage for learners who have spent time in a German-speaking environment and/or extensively conversed with native speakers of German.

For those participants who did not spend time in an immersive context, the reliance on word order might not have been challenged as a reliable cue for processing and therefore might have won the competition against morphological case markers. And even though the essay prompts were the same across participants, the difficulty level would have depended on what the individual made of them. Writing about a purely hypothetical holiday might have been harder than writing about a past experience that was accompanied by positive feelings and that left fond memories. Likewise, the perfect university experience might have been easier to describe if the student thought the current situation was ideal and could draw on these experiences, compared to the need of creating a good hypothetical example and building an argument from scratch.

Previous research on writing in a foreign language has provided evidence for the differences between L1 and L2 writing. When writing in their L2, writers tend to produce shorter sentence parts and revise them more often during writing (Galbraith, 2009: 16). The essays produced in this study are in line with this observation. If new ideas are generated, then the ideas have to be kept accessible in Working Memory until the sentence is completed. If content stored in Long-Term Memory is accessed, on the other hand, a great deal of the processing load is alleviated by the use of pre-structured chunks. Using the same task for all participants does not mean that they have the same experience of the task and its complexity. That being said, some patterns have emerged in this case study and will be discussed in the following sections.

Working Memory Capacity and Phonological Short-Term Memory

In previous research, there has been evidence of both Phonological Short-Term and Working Memory being linked to the ability to detect, analyse, and generalise grammar rules (Martin & Ellis, 2012: 393). The participant with the least exposure to native-like German showed a clear improvement over time with regards to his ability to interpret OVS sentences correctly. For him, Processing Instruction could have tipped the scales towards a reliance on case markers rather than word order. His high Working Memory Capacity and Phonological Short-term Memory Capacity scores could have helped to create stable form-meaning connections: Because the activation levels of the case markers were raised during instruction and because the Structured Input materials provided a clear conceptual candidate to be matched with the morphological marker, the co-indexation was successful. The advantage of the high Working Memory Capacity might have come to play in this and/or in the maintenance of goal-relevant explicit knowledge that could have sped up the process.

Working Memory has also been associated with higher-order cognitive abilities “such as comprehension, reasoning, and problem solving” and accounts for “a significant portion of variance in general intellectual ability” (Conway, et al., 2005: 769). It is therefore not surprising that individual differences in Working Memory Capacity seem to have played a role in participants’ assessment performances after receiving Processing Instruction. It is true that this pedagogical intervention works with manipulated input that is maximally reduced in complexity and ambiguity. This is reflected in the relatively high accuracy rates of participants during the instruction. It must be noted, however, that for most participants, the accuracy rate drops towards the end of the instruction sessions. This is in keeping with the observation that Working Memory Span tasks not only require test takers to switch attention and to update task-relevant information, but also quickly forget about irrelevant information from previous trials. The challenge of sustaining attention and counteract interference from previous stimuli usually leads to lower accuracy rates at the later stages of each trial block. From a MOGUL perspective, the case markers and prepositions that have been processed during instruction are now exhibiting heightened activation levels, meaning that all of them are available for processing and therefore part of the competition. What determines the outcome are the resting activation levels from which they started and the stability of the form-meaning connections (or, in MOGUL terms, the co-indexations across different modules).

Previous research on Processing Instruction and individual differences in Working Memory Capacity (Santamaría, 2007) shows different effects of Working Memory on instructional outcomes, depending on the time of testing: Results for the interpretation task revealed that in the Processing Instruction group, both high- and low-span participants improved similarly from pre-test to delayed post-test, with the high-span participants scoring higher on the immediate post-test and the low-span participants scoring slightly higher on the delayed post-test (Santamaría, 2007: 68). However, the Working Memory scores do not lead to significant differences between high-span and low-span participants for the pre-test and delayed post-test, only for the immediate post-test (Santamaría, 2007: 75–6). The findings of this case study are in line with this.

The observation that the participant with the densest variable cluster performed best in the longer run is also in line with other studies that explored several learner variables at once: Sparks et al. (2012) conducted a retrospective study with 208 participants from different research projects and explored the relationship of L1 achievement, intelligence, L2 aptitude, and L2 proficiency profiles. Using *k*-means clustering, they derived three different learner profiles: Above-average L2 learners showed above-average scores on all L1 and L2 measures, average L2 success was more or less consistent with average scores on the measures, and below-average participants scored below average on all measures. The findings from a more fine-grained analysis in this study leave the impression though that Working Memory, if operationalised as a construct which can be measured via dual-span tasks, is not the decisive factor for language learning success. Depending on the task at hand, the nature of the assessment, and the processing context, the participants produced very variable scores, which begs the question whether Working Memory Capacity, when operationalised as a more or less stable construct, has much predictive power.

Other studies have shown that Phonological Short-Term and Working Memory contributed to the detection and generalisation of grammatical structures independently of each other (Martin & Ellis, 2012: 396–479). They highlighted that in order to learn from feedback, learners had to keep their utterance active in order to compare it with the feedback and notice any differences. In this study, the participants did not produce any German output during the instruction, but they did receive immediate feedback on their interpretations. In order to learn from that feedback, they did not have to remember which interpretation they had chosen because their choice was still available for inspection on the screen, but they had to apply that knowledge to subsequent items. As shown in the previous section, the consistency with which the questions were answered, varied

depending on the task and the participant. Task modality was also an important factor, considering that the participants with the most exposure to German and the longest immersion experiences generally performed better on the aural Structured Input Activities. Providing more context, as in the case of the discourse-level interpretation and comprehension tasks, also had a positive effect on performance. The additional information might have helped with the raising of activation levels as both the input processing and the output processing of a coherent text should lead to a repeated activation of co-indexed and related representations in several modules. The global understanding of the text then would add spreading activation to the co-activation of co-indexed items and help in establishing a consistent text message: "Thus, comprehension is the assembling of a coherent massive network of information that reflects the current input and has been experienced in some similar way in the past" (Grabe, 2009: 87).

Attitude, Interest, and Motivation

According to previous research findings, motivation seems more predictive of overall language learning success, while anxiety and introversion affect the actual (speaking) situation. Motivation has often been considered to be a trait of a "good" language learner (Dewaele & Furnham, 1999). According to Skehan (1991: 281), motivation plays a role for the current learning situation as well as for the overall language learning experience. A favourable attitude towards the language and its speakers increases the effort that learners put into the learning process and makes them more likely to persist in the face of difficulty or frustration. In this study, several different profiles emerged, depending on the time that the participants had spent in a German-speaking environment and how much contact they had with native speakers. Overall, this did not seem to contribute significantly to the differences observed during and after instruction though. The Attitude/Motivation Test Battery questionnaire sections in this study were administered at the beginning of the term, so scores might have been influenced by the change in routine, relative distance of assignments, impact of new teaching staff, etc. Getting paid for the participation in a linguistic experiment might have added another layer of extrinsic motivation as well as distortion. Potential measurement errors therefore stem from both the instruments and the dynamic constructs measured. If this were to be explored further, a more dynamic measure of interest, attitude, and motivation would have to be employed.

Anxiety

There has been considerable debate as to how anxiety should be categorised, i.e. as a (negative) motivational factor (Gardner, 1989), a basic emotion, or a personality trait (Eysenck, 1979). Different aspects of anxiety have been highlighted, for example the trait versus state distinction and the differentiation between general anxiety and foreign language anxiety (for a concise overview of previous research, see Dewaele, 2013: 672–675).

Foreign Language Anxiety seemed to be linked to a multitude of other factors in previous research: “age, academic and FL achievement, previous contact with FLs, perceived scholastic competence, self-worth, intellectual ability, and job competence” (Dewaele, 2013: 674). It seems to peak in the learners’ twenties (Dewaele, 2010). Other factors are the number of languages learnt and cross-linguistic differences as well as situational factors, such as the target audience. In this study, the participants received individual computer-delivered Processing Instruction. The role of anxiety is therefore hard to judge as it mostly shows correlations with spoken output and in conversational settings (Dewaele & Furnham, 1999). However, strong anxiety might have impacted on previous language learning and might have led to less stable form-meaning connections for some learners in this study.

Some interesting questions can be asked with regards to what the MOGUL framework offers on the role of affect in language learning. If representations are automatically co-indexed with the affective state of the learner at the time of processing, and re-activated whenever this state is in place, then strong co-indexed emotions would lead to a boost in resting activation levels, which would also facilitate retrieval. This is particularly interesting for Processing Instruction as, traditionally, some of the Structured Input Activities are affective rather than referential in nature. If those affective activities are designed in a way that is meaningful, relevant and relatable for the students, potentially activating positive emotions, then this would remove any doubts regarding their functionality. The present study has only employed referential Structured Input Activities, but further exploration of their affective counterparts seems worthwhile in this context. The dominance of the perceptual output structures (POpS) might provide support for a more holistic approach to language learning and teaching, for example encouraging the use of task-based learning and/or instruction that builds on actual experiences. In these cases, several converging sensory input sources would reinforce each other and lead to fairly high resting activation levels for all those representations which are newly established or co-

indexed in new ways. This approach would make parallel processing more likely, meaning that the probability of a consistent overall representation being established (potentially leading to a stable form-meaning connection) is increased.

Extraversion

Considering that extraversion is linked to strong positive affect, this dimension could be important in terms of the motivational system of a learner (Dewaele & Furnham, 1999: 512). There is a trend for extraversion to decrease with increasing age (Eysenck & Eysenck, 1981: 15), so the oldest participant being the most introverted is in line with that, but clearly this is not the only factor as the two first-year students, which are of the same age, show very different degrees of extraversion. Participants living in an urban environment usually score slightly higher on extraversion and so do male participants compared to female participants (Eysenck & Eysenck, 1981: 15). The latter was not the case in this small sample.

Introverts supposedly need more time to access information (Eysenck, 1981: 203), which has consequences for their processing capacity, and their attention is divided between task-relevant cognition and self-related cognition (Eysenck, 1979). This could be observed in the case of MgrYyVvU, whose reflective tendencies regularly interfered with the accomplishment of the tasks. On the other hand, extraversion might have a more pronounced effect on foreign language proficiency than on the developmental trajectory (Dewaele and Furnham, 199: 518). It could not be linked to higher global achievement scores when it comes to second language acquisition – except in oral communicative speech, when introverts would feel the pressure and anxiety of the communicative setting (p. 533). As summarised by Dewaele & Furnham (1999: 532), the strongest correlations between extraversion and linguistic variables can be found in the context of complex verbal foreign language tasks. These were not employed in this study.

Grammatical Sensitivity and Language Analytical Ability

As VanPatten & Borst (2012) and VanPatten et al. (2013) point out, grammatical sensitivity could play a role for some individuals and for some target features – especially those which benefit from an application of explicit information. The participants in this

study produced fairly equal scores in terms of grammatical sensitivity, but differed vastly in terms of their immediate post-test scores. While this is only a small participant sample, grammatical sensitivity does not seem to play an important role, especially not when compared to impulsivity/reflectivity and exposure to German. However, it might have helped the participant with the least exposure to German and one of the highest Working Memory Capacities. In the absence of naturalistic input and with pedagogical interventions such as Processing Instruction, grammatical sensitivity might be a factor as learners need to be able to interpret the metalinguistic information appropriately. Information on the roles of subjects and objects is useless for someone who does not know what a subject or object is. On the other hand, the participants were tempted to rely on surface features even in English, routinely choosing “playing” instead of “Kate” in the example below.

8. Arguing never solved a problem.

Right now Kate is playing with her brother.

Figure 7.1: Example sentence in Words in Sentences section

The scores for language analytical ability show a little more variation and are lower for this university sample than they were for the secondary-school students in the pilot study, even when different word orders in the materials were accounted for. Most remarkable are the differences in the time the participants spent on this task, with the quickest (and most impulsive) participant finishing within ten minutes and the slowest (and most reflective) participant giving up after twenty minutes. None of the participants was able to demonstrate productive knowledge of the rules for case or number declension in the artificial language. Looking at the Working Memory Span test performances and the scores from the language analytical ability test involving the artificial language Pip, there seems to be no connection, which makes sense as the materials were in written form and the participants could annotate them or colour-code the words that they had already tackled, thus relieving their memory resources.

Coming back to the distinction between learners who rely on analytical abilities and learners who rely on memory, there is evidence for different approaches in the essays that the participants produced. This might be a reflection of different learning contexts (immersive versus classroom setting). Learners who rely on memory rather than analytical abilities are, according to Wesche’s (1981) findings, less satisfied and achieve

less in learning contexts which are analytical in nature and vice versa. This could not be observed in this study as no introspective data were collected. Judging from the output, however, there seems to be a different “flavour” to the texts, with some essays building on surface structures and other on content development, but this does not necessarily translate into different overall scores. Processing Instruction does neither heavily rely on memory, nor on conscious analysis. This could make the pedagogical intervention equally beneficial or equally superficial for either learner type. The assessment tasks, which were output tasks that were controlled to varying extent, might encourage the reliance on chunks stored in memory or the use of rules to generate new linguistic output – and both approaches could be equally successful. While it is tempting to speculate, without some form of introspection, it will be impossible to determine which approach was used.

7.3 Discussion of the study’s contribution to the existing body of research

The present study has followed an idiographic line of investigation and put a particular emphasis on the individuals involved in individual differences research, thereby responding to Skehan’s (1991) call for more ethnographic and fine-grained approaches in this area of research. The findings might not be generalisable to a wider population, but this small and diverse sample has allowed an in-depth analysis of both process and product data. While some aspects of the experimental setting have been controlled more rigidly than in previous Processing Instruction research, e.g. by using automated Working Memory Capacity span tasks instead of PowerPoint versions delivered to a group, more diversity was allowed with regards to the participant sample, e.g. by including a participant with a known learning impairment. Choosing university students enrolled in a German programme meant a great deal of variability in terms of age, proficiency level, language learning experiences, etc. It also meant that the participants were at an advanced level of German with extensive previous exposure to formal instruction and therefore not naïve, which might have boosted the usefulness and/or applicability of the explicit information provided. Considering that this was the first case study within the context of Processing Instruction research and that it was also the first study in this field to consider the effects and interrelationships of several learner variables at the same time, the results are part of a learning curve and generate tentative answers as well as questions for future research. The data collected in this study point to several interesting paths to be explored,

one of which is the role of extraversion, and especially the impulsivity dimension, in formal classroom instruction.

While research on personality factors has increased in recent years, this variable had not been considered in Processing Instruction research to date. In the next section, it will be seen that the establishment of a link between Eysenck's (1990) Arousal Theory and the MOGUL framework could explain some of the effects observed. Moreover, the interplay between variables such as motivation or anxiety and their links with personality dimensions such as extraversion-introversion have yet to be investigated systematically. Keeping in mind that Structured Input Activities are maximally reduced in complexity and do not require the production of second-language output, this pedagogical intervention might be especially suited for introverted and/or anxious learners.

Previous research that focused on the learning rate demonstrated by learners receiving Processing Instruction on German case marking, as presented in VanPatten & Borst (2012) and VanPatten et al. (2013), has used a trials-to-criterion approach that counted the number of incorrectly processed sentences before the criterion of three correct OVS plus one correct SVO sentence interpretations was reached. VanPatten & Borst found that while explicit information was not essential for the correct processing of German accusative case marking, it seemed to speed up the process. But even within the group of learners that had received explicit information on the target feature, some participants started processing the accusative case marking immediately, and some took longer or never reached the criterion throughout the 50 trial sentences. The weak effect of grammatical sensitivity observed in the group that had received explicit information was attributed to the possibility that straightforward, easily applicable explicit knowledge can help learners in detecting the case markings. As an explanation for the differences in learning rate, VanPatten & Borst (2012: 104) offer the following statement:

Perhaps there is a relationship between grammatical sensitivity and the ability to use EI during a processing task, a relationship that may involve a third variable such as working memory. Only future research that uses a multitude of measures of individual differences could address this issue.

This case study has investigated the effects of Working Memory Capacity and other individual differences variables, including grammatical sensitivity and language analytical ability. While the effect of explicit information has not been measured, the participants showed different instructional outcomes despite similar scores on the grammatical sensitivity and language analytical ability tests and after receiving the same amount and

type of explicit information. While the target feature in this study cannot be directly compared with the target feature in VanPatten & Borst's (2012) study and while different processing problems were targeted, it seems that grammatical sensitivity had no clear impact on the instructional outcomes of this study.

However, the findings also provide support for the argument that it is possible for adult learners of a second language to change their default processing strategies and improve, even in a classroom setting. If output processing is the reversal of input processing, then both should serve to reinforce the activation levels and co-indexations of representations across the respective stores, making them both more accessible for future processing and potentially more complex. This would explain the findings from previous Processing Instruction studies (as well as the present one) that participants who have received Processing Instruction can produce the target feature without having received output training. The intensive exposure to the feature during instruction has heightened the resting activation levels. The straightforward co-indexation of functional and conceptual representations has made them accessible and relatively easily available for production (Sharwood Smith & Truscott, 2014: 304).

7.4 Interpretation of results

In this section, the findings presented previously will be interpreted based on the MOGUL framework and for each participant individually to recognise the particular profile that they bring to the tasks. For native speakers of German, activation patterns are reinforced multiple times because of the co-indexed representations for case as well as for person and number agreement. Once these patterns are established, they act like a processing short-cut whereas learners of German as a foreign language, who already have acquired one or more languages and are exposed to very limited German input, need to establish these patterns through extensive processing and in the face of competing representations from other languages. Especially when it comes to functional categories, learners of a foreign language display prolonged periods of idiosyncratic optionality. Even if the correct feature value has been established and learners demonstrate native-like comprehension, they can still produce ungrammatical output until the activation levels of the correct feature construct are high enough to win the competition. For some, this might never be attained (Sharwood Smith & Truscott, 2014: 234–7). For others, optionality persists, which means that acquisition should be seen in probabilistic rather than dichotomic terms.

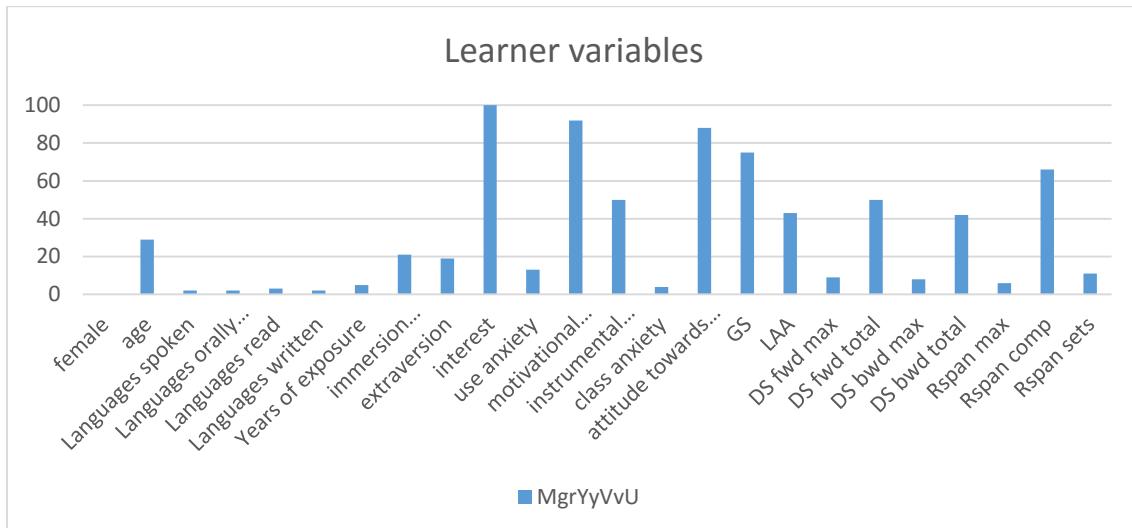


Figure 7.2: MgrYyVvU's learner profile

MgrYyVvU is the participant who scored highest in terms of Working Memory Capacity, but he is also the one showing the highest degree of reflectivity. This is evident in his longer adaptation periods for unfamiliar tasks, such as the picture selection task in the first session. Once task novelty wears off, he improves his scoring consistency, but occasionally still exhibits a lot of optionality, despite his relatively long and intense immersion experiences. According to MOGUL, it is possible that the L2 language system in the core language module is native-like while the metalinguistic and conceptual knowledge system outside the UG-restricted modules is not. His reflectivity and constant self-monitoring could therefore get in the way of processing the forms correctly – not just for production, but also for interpretation.

The difficulty ratings are not necessarily corresponding with his scores, although negative feedback seems to have an effect, but are dependent on how sure he feels about a task. Even on tasks where he performs equally well as YKX5Bwvj and gets more positive feedback towards the end of the task, he rates the task difficulty higher than she does. His first essay was very short because he could not think of something to spontaneously write about. Due to the lack in length, errors affected the overall accuracy rate quite negatively. It is quite clear that he, as someone who did not study German in school, is not comfortable with producing this type of text, which is also evident in the greater variability in his accuracy scores. While he scores 65% on both the immediate and the delayed post-test, the two texts are very different in tone and style.

MgrYyVvU scores low on impulsivity, and his reflective, questioning demeanour hinders him quite often in the accomplishment of the tasks. Even when he successfully and swiftly

works through the first half of an exercise, he can abruptly stop and think about the task until he is confused. While he usually comes up quickly with an answer, probably provided by his core language system, he then consciously stops to re-analyse, which sometimes results in changing a correct construction to an incorrect one. He is constantly evaluating his performance, which detracts from task-related processing. He is, however, a mature student and familiar with the university environment, which is reflected in his relatively low anxiety scores. He has a deep interest in the language and in German culture and wants to continue his studies in Germany. His instrumental motivation is relatively low, although German clearly serves a purpose for him as well. It seems though that he is not avoiding difficult situations and he does not choose the easiest path.

Interestingly, MgrYyVvU was recently diagnosed with dyslexia, but this is not reflected in his Working Memory Capacity scores. According to recent research findings, dyslexia is assumed to be caused by an impairment in the learning of the serial order of items (Szmałec, et al., 2013: 80). This should be reflected in lower Digit and Reading Span task scores, but this is not the case. Neither do the observational data converge with the statement that his processing speed is below average. He does, however, struggle the most with the language analytical ability task and cannot offer translations of his own for the artificial language. Considering the overall impression of his constant doubting and overthinking, it stands to reason that his learning difficulties are not necessarily caused by dyslexia as such, but by the interference of processes outside the core language module (Sharwood Smith & Truscott, 2014: 304).

YKX5Bwvj

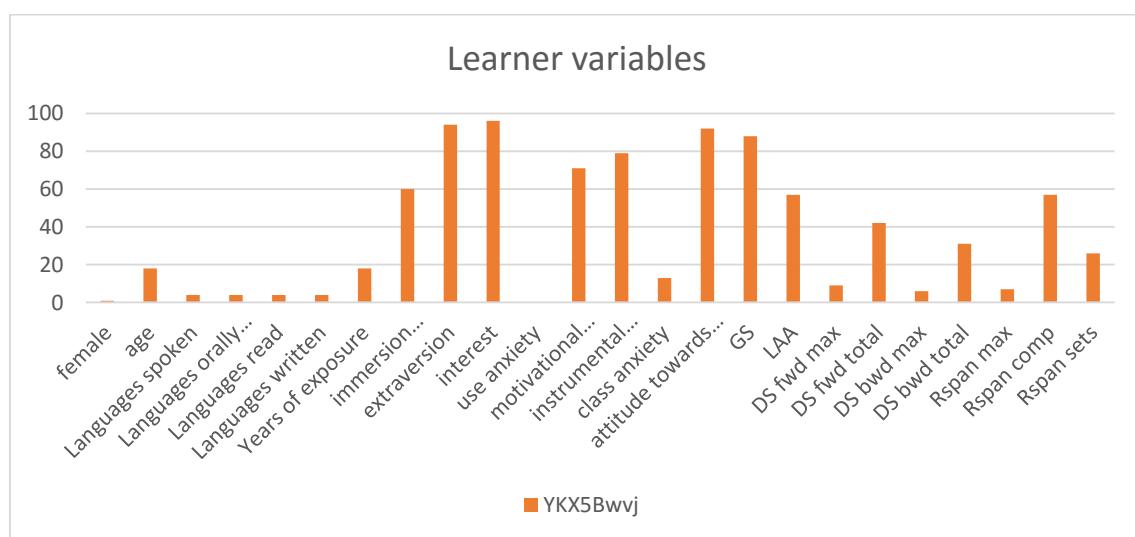


Figure 7.3: YKX5Bwvj's learner profile

Participant YKX5Bwvj, who grew up in Germany, but returned to the UK before she reached school age, is a special case in many respects. She is the only participant who improved on several measures across all testing sessions. One reason for this could be that she has German family and receives regular German input. In MOGUL terms, Processing Instruction might have helped in establishing form-meaning connections and straightforward co-indexation of representations in different stores. Consequently, the activation levels might have reached a critical point that got reinforced every time she subsequently processed switch prepositions. From a connectionist perspective, “neural computation is often highly nonlinear. This means that under certain circumstances, small differences in input may have little effect; whereas under other circumstances, small differences in input produce qualitatively different behaviors” (Elman, 1997: 19). The additional input might have made the crucial difference as 74YuGGCRD also made clear and sustainable gains in interpretation, but could not turn them into improved production yet. This is further proof that the same instruction does not have the same effects on learners (Pienemann, 1984: 189). That being said, YKX5Bwvj also exhibits the greatest variance in scores overall. Despite the few assessment items in the picture description, it still seems remarkable that she should produce 50%, 100%, and 20% on the respective tests when most participants started out at above 80% and were mostly able to hold on to their improved scores in the delayed post-test. It could be that this task was simply too easy and not interesting enough to capture her attention.

The scoring inconsistency is also noticeable on the Reading Span task, where her maximum span is up with the best, but her composite scores are lower than those of MgrYyVvU and 74YuGGCRD, respectively. While she easily generates a lengthy text, she does not always produce appropriate language in terms of register, using expressions such as *saufen* (to booze, pejorative) instead of more sophisticated alternatives. The texts are not structured in a planned way, but follow her thoughts, which connect the parts in a way that goes deeper than surface structures. Overall, she works very swiftly and without thinking, yet produces the best performances on the delayed post-test, when most other participants’ scores drop sharply on most tasks, sometimes even below the pre-test scores. Her high degree of extraversion, which supposedly is accompanied by a lack of neural arousal, might have led to relatively low activation levels, meaning that her native-like core language representations were too slow to win against the processing outside the module – until intensive processing of switch prepositions during the Processing Instruction treatment started to even the score. More empirical research would be needed to support this claim though. It is also possible that individual computer-delivered

Processing Instruction and written output production for the assessment did not play to the strengths of this extraverted learner. According to the essay, she studies several languages in order to see the world and make friends in different countries, so a face-to-face communicative task might have increased cognitive activation and elicited more consistent scores.

YKX5Bwj's attitude towards the language is shaped by the positive experiences she had growing up in Germany and when spending the summer breaks with her German family, but this does not necessarily translate into high performance scores. It does, however, affect her anxiety scores. Because she is used to communicating with native speakers of German, she does not report any use anxiety at all. Her being new to the university system, however, results in slight class anxiety. Her motivation and instrumental orientation are relatively high, but whether this is due to a genuine interest in the language or due to values that have been passed on to her through the German side of the family, or a combination of the two, is unclear on the basis of the data.

74YuGGCRD

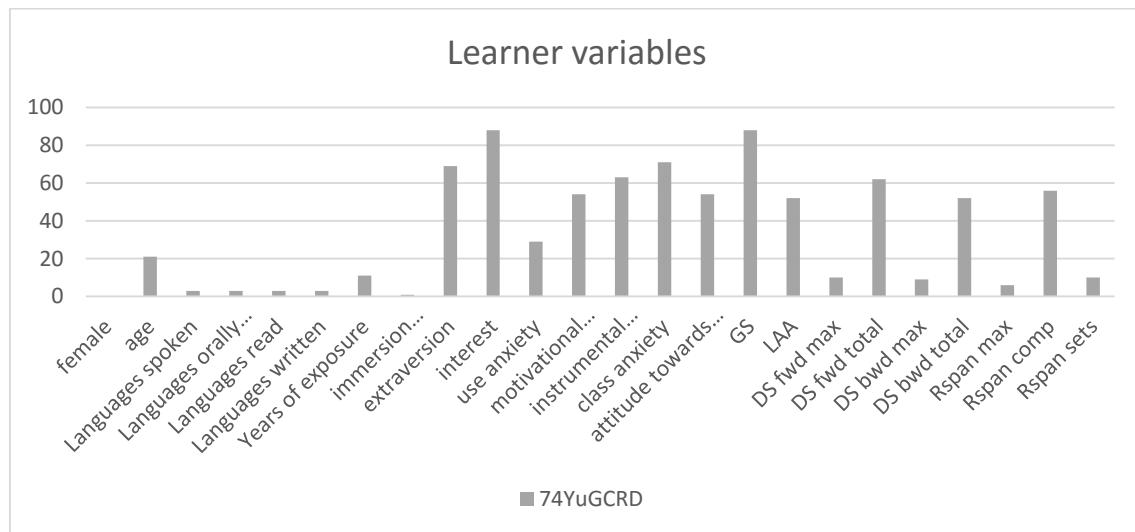


Figure 7.4: 74YuGGCRD's learner profile

Participant 74YuGGCRD scores consistently well on those tasks which allow him to use the language he is comfortable with. Especially the essays that he writes are very similar in terms of overall grades and use of switch prepositions, although he exhibits some improvement in overall accuracy directly after instruction. He had some problems during instruction with the kinds of verbs which are hard to learn unless one receives a lot of input, e.g. the distinction between to lie and to lay. His very high Digit Span scores point

to a well-developed auditory module, or phonological loop in Baddeley's terms, which might have counteracted some of the disadvantages he had compared to the two students who had just spent time in Germany. On the translation task, 74YuGGCRD shows clear improvement from pre- to post-test, which he can maintain in the delayed post-test session. The translations are not sophisticated, but he clearly does not rely on word order any longer.

74YuGGCRD is an interesting case because he shares a learning context with a lot of English students learning foreign languages, i.e. he is exposed to the language only during the class time and has had no immersive experiences when it comes to German as the only contact situations during school trips took place in the midst of a big group of English-speaking adolescents. From a MOGUL perspective, his language learning process should be rather slow because the input available for processing is relatively scarce. Due to the small amount of input being processed, representations take a long time to build up their resting activation levels and the storing of more complex, chunked representations takes longer as well as they are not encountered often enough to become a short-cut during processing. All this leads to a rather slow development, especially when it comes to processing for production. This interpretation/production divide can be observed in this study as well.

If there is no strong motivation to learn the language and/or no additional input for processing, then the learner will never become truly proficient and is likely to give up learning the language before tangible development manifests itself. To a certain degree, instruction can help language acquisition, especially if the explicit information that students receive helps them improve their processing (Sharwood Smith & Truscott, 2014: 306–7). According to MOGUL, optionality will persist even when native-like processing of a feature has been developed by the learner. As long as there is no clear winner in the competition, the processing for production of the correct feature will be probabilistic in nature rather than absolute. In terms of interpretation, this means that the conceptual store offers interpretations of the message based on representation patterns which have higher activation levels than the output of the syntactic processor and therefore beat the morphosyntactic information in the competition. If no additional explicit knowledge is available which could veto the incorrect processing, then the faster, i.e. more easily available item, will win the race (Sharwood Smith & Truscott, 2014: 78–82). At first glance, the participants did not have to process many OVS sentences, but the Structured Input Activities mostly emphasised the prepositional phrase, meaning that the subject of the

sentence appeared routinely after the verb. So the processing experience reinforced the explicit information that the subject does not have to be the first (pro)noun in the sentence.

While 74YuGGCRD is rather balanced in terms of extraversion, his class anxiety score is fairly high, which is hard to explain without further data collection and knowledge of his general classroom anxiety outside the German sphere, but it could have to do with the fact that he has had no immersive experience whatsoever and is at an age where anxiety peaks. His use anxiety might have been higher had he actually been in situations where he had to speak to native speakers outside the classroom. Judging from his immediate post-test essay, it seems that even during his year abroad in France, he was mainly spending time with Australian students rather than taking advantage of the immersion context to improve his French and in his delayed post-test essay, he expresses no desire to continue with language studies and states that he would not repeat his choices if he could start all over again. His relatively low attitude score supports this impression.

nLQ6aVus

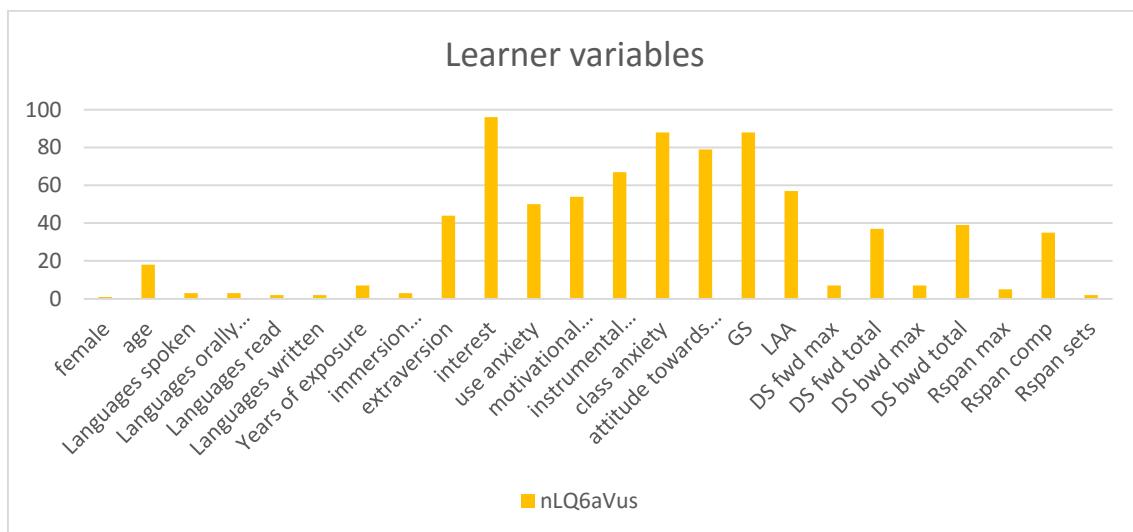


Figure 7.5: nLQ6aVus's learner profile

Participant nLQ6aVus starts well on most pre-test measures, i.e. the fill-in-the-blanks task, the picture description, and the essay task. She does not, however, do well on the interpretation task and cannot demonstrate substantial gains from pre-test to post-test, let alone from pre-test to delayed post-test. Her translations show a lot of variability and are often far from the sentence meaning, which is remarkable because she uses a wide range of vocabulary in her writing. More data would be needed to see why she does not apply any strategies, e.g. paraphrasing or inserting blanks, for the translation task. While

she might not show an emotional engagement with the materials, contrary to YKX5Bwvj and MgrYyVvU, she always produces at least as much output as required and often offers more. She also shows evidence of planning her tasks, for example by writing case morphology paradigms in the margins of the test materials, by checking the gender of the nouns, and by counting the words for each essay prompt to check that all prompts receive equal space.

On the Working Memory measures, nLQ6aVus attained the lowest scores in the whole group, no matter which scoring technique was used. Her Digit Span scores are slightly better, so it seems that it is the dual-task component that causes problems for her. This might be particularly relevant for the translation task as it depends on the integration of syntactical and semantic information, requiring the participants to hold a lot of goal-relevant information in an active state. She did very well on the discourse-level interpretation tasks during instruction and on the discourse-level production task before and after instruction. Instead of complicating the interpretation task, the additional information might have helped in activating a network of representations that reinforced each other and supported the general meaning of the text (Grabe, 2009: 40), leading to accurate answers to the multiple choice questions. Her style of writing also suggests that she has written this kind of essay many times before and it is likely that she accessed chunked language in Long-Term Memory for this. The other first-year student who recently graduated from school and should have been used to this type of essay did not produce such a measured and well-constructed text.

nLQ6aVus is very young and just started university, her class anxiety is very high, which could add to the processing problems by increasing the neural “noise” during processing. Her approach to learning, i.e. careful planning, preparation of tasks, use of memorised chunks of language, and a high degree of organisation, might be challenged in a learning environment that is new and might feel less safe than the school environment in which she was doing well. Her impulsivity is relatively high, according to her self-report. That could indeed be the case, adding to the inconsistent scoring, but from her demeanour, it could also be an instance of social desirability bias. According to her motivational intensity scores, she has the tendency to give up easily when she does not understand something. Her off-target translations might be evidence of that.

8 CONCLUSION

This chapter will summarise the findings of the present study and evaluate their significance in terms of what their theoretical and methodological contributions to research on individual differences in Processing Instruction are. The limitations of this small-scale case study will be acknowledged and suggestions for further research will be made – with the goal of addressing these limitations in the future. Considering that this project is straddling the divide between Second Language Acquisition research and Foreign Language Teaching, further suggestions for a more principled approach to teaching German switch prepositions will be made in section 8.4. These suggestions aim to improve future Processing Instruction materials on German switch prepositions, but at the same time address the question as to whether the latter are a suitable target feature for Processing Instruction at all. A new classification of locative prepositional verbs with case alternation will be suggested and the nature of the optionality in case marking following German switch prepositions will be discussed in order to show how the target feature of the present study fits within the bigger picture of teaching German case morphology. Finally, the practical applications for computer-delivered Processing Instruction will be summarised and an outlook on their potential for foreign language learning and teaching will be provided.

8.1 Summary of the findings

This case study has explored the role of individual differences in Working Memory Capacity in the context of computer-delivered Processing Instruction on German switch prepositions. The findings, which have been presented in Chapter 6 and discussed in more detail in Chapter 7, can be summarised as follows:

1. Do learners benefit from Processing Instruction in equal measures, regardless of their Working Memory Capacity?

Hypothesis: Processing Instruction is maximally reduced in complexity, yet there usually are individual differences in terms of the post-test scores. Differences in Working Memory Capacity will result in differences in post-test scores.

SUPPORTED

Individual differences in Working Memory Capacity and/or Phonological Short-Term Memory Capacity seemed to play a bigger role for some tasks than for others, but they do not seem to predict performance on Structured Input Activities. This seems to support the claim that Processing Instruction materials are maximally reduced in complexity and ambiguity and therefore allow relatively straight-forward form-meaning connections, which makes Processing Instruction a suitable intervention for learners of different Working Memory Capacities. This is also evidenced by the participants' difficulty ratings for the Structured Input Activities. The difficulty ratings for the Processing Instruction tasks are generally lower than for the assessment tasks, especially when task novelty is not an issue. Additional context, as provided in discourse-level interpretation tasks, does not necessarily have a negative effect, however. The participant with the lowest Working Memory Capacity scores seemed to benefit from discourse-level tasks more than from sentence-level tasks in terms of both production and interpretation.

The assessment scores, as predicted, showed more variability and suggested an effect of individual differences in Working Memory Capacity, at least for the most challenging task and for the immediate post-test. Working Memory might play a less pronounced role in the longer run, which is in line with the observation that learners with different profiles and predispositions might exhibit different learning trajectories and rely on different memory or processing systems, but eventually do acquire most structures. While the participants performed overall at a similar level during instruction, with individual participants having problems with individual tasks, the more complex assessment tasks showed marked differences in performance, some of which can potentially be attributed to Working Memory Capacity limitations. That being said, there seem to be other variables which mediate the influence of Working Memory Capacity and can compensate for "deficiencies":

2. Are there other learner variables that potentially mediate Working Memory Capacity and thereby affect the learning outcomes?

Hypothesis: There might be other variables that compensate for lower Working Memory Capacity scores. If any of the measured constructs have an impact on post-test scores, corresponding patterns should show up in the individual learner profiles.

SUPPORTED

While individual differences in Working Memory Capacity seem to have had a tangible impact on the immediate post-test scores of the most complex task, other variables seem to have influenced the participants' performance before, during, and after instruction as well. Exposure to German seems to have been the reason why one participant showed continuous improvement across all test sessions on several measures. Impulsivity also played a role in the consistency of the scoring and the way the participants interacted with the materials. The learner variables under investigation seem to have affected the on-task performance to different extents though, depending on the nature and modality of the task. While the discourse-level interpretation and production tasks were not as complex as they could have been, the better-than-expected performance of the low-span participant on the discourse-level tasks indicates that additional or global information as well as an increased reliance on chunks stored in Long-Term Memory can be an alternative to a reliance on Working Memory. Lack of regular exposure to German, on the other hand, was mediated by a balanced personality, a moderate amount of anxiety, and a well-developed Phonological Short-Term Memory. The superiority of the participant with the highest Working Memory Capacity scores, however, seems to have been diminished by his high degree of reflectivity, which especially affected the delayed post-test scores. Performance differences might have been more pronounced if the participants had been required to produce oral output, especially in spontaneous conversation.

Learner variables which had a positive effect on one task or one individual might have had a negative effect on a different task or a different learner. Judging from the learner profile analyses, it seems that the learner variables also influenced each other. From this small-scale case study, it is not possible to deduce cause-and-effect relationships, but the discussion of individual learner profiles in the previous chapter should have provided a qualitative insight into how these four participants came to produce the observed outcomes. Especially regular exposure to German and some personality dimensions, such as impulsivity versus reflectivity, seem to play a bigger role than anticipated. The tasks affected the participants to different extents based on this interplay of learner variables. While this might not have hindered acquisition per se, it might have decreased the learning rate for some participants, leading to lower production scores and potentially not reflecting the improvement in processing for interpretation. But this is a problem that is not unique to Processing Instruction studies, as discussed previously. Which leads to the question whether this type of instruction affected different individuals to different extents.

3. Are there participant-treatment interactions?

Hypothesis: It is conceivable that different individuals react to and interact with the instructional and experimental materials in different ways. This could influence the learning outcomes. Evidence for this might be found in the participants' approach to the less controlled tasks, in the way participants work with the materials, and in the observational data the researcher collects.

SUPPORTED

Evidence for this was found in the participants' approach to the test materials as well as the instructional materials, but also in the way they tackled the questionnaires and Working Memory Capacity measures. Depending on the personality type, task novelty could either lead to curiosity or confusion; and participants showed differences in terms of scoring consistency, reactions to feedback, perceived task difficulty, maintenance of learning gains, etc. It stands to reason that these differences would have been observed in the context of any pedagogical intervention and there is no way of telling whether Processing Instruction is any more engaging for some participants than for others. In the end, all one can do is create a variety of tasks on a topic that is relatable and of interest to the students and hope that at least a fraction of what is offered finds its way into the developing system.

8.2 Evaluation of significance of the findings

The present study has contributed to the existing body of research on Processing Instruction in several ways, see Table 8.1 at the end of this section. The use of computer-delivered instructional materials has added to findings from previous studies which demonstrated that Processing Instruction can be successfully delivered individually via a computer (Lee & Benati, 2007). The new instructional package was delivered via Adobe Captivate, showing that computer-based Structured Input Activities can be developed without the need for specialist experimental software and/or programming experience, just by using the guidelines for Processing Instruction materials (Farley, 2005; Wong, 2004). Working with a university population has provided data on more advanced learners that were struggling with the mastery of a particular feature, namely case alternation following German switch prepositions. This target feature had never been investigated

within the Processing Instruction framework. Prior to instruction, the participants were relying on word order rather than case marking when interpreting German sentences and they were unable to consistently produce the right case markings in the context of switch prepositions. After receiving two 50-minute sessions of Processing Instruction, all the participants demonstrated an improved performance in terms of interpretation and production of German case markings, suggesting that this is another feature which can be taught using Processing Instruction.

The delayed post-test, administered fifteen weeks after instruction, showed a picture that is in accordance with the Acquisition by Processing Theory (Truscott & Sharwood Smith, 2004): If the features are not being regularly processed, their resting activation levels fall back and make them less accessible for subsequent processing, especially when processing for output is concerned, as in the context of production tests. An earlier second post-test or a repeated exposure to the target feature might have resulted in different outcomes (see Hikima, 2010), but the findings overall reflect what language teachers and learners intuitively know: you either use the language, or you lose it. However, it should also be noted that the assessment tasks were designed to be as authentic as possible, i.e. the fill-in-the-blanks task, the essay writing prompt, the picture description, and the translation exercises are all types of tasks which can be found in textbooks of German as a foreign language (Domínguez Vázquez, 2005). None of them are Processing Instruction tasks. The instructional gains can therefore not be attributed to skill learning alone as the tasks as such differed vastly from the Structured Input Activities that the participants encountered during instruction. The observation that most participants maintained modest gains on most tasks over fifteen weeks is further support for Lee's (2004: 316) hypothesis that Processing Instruction has long-term effects for the processing of second-language input, but is not unique to this pedagogical intervention, of course.

What truly set the present study apart from all previous Processing Instruction research though was the emphasis on the individual in individual differences research. Instead of selecting one variable, Working Memory Capacity, and comparing the performance of high-span and low-span participant groups, this small-scale case study collected data on several potential mediating variables in addition to measuring Working Memory Capacity, allowing the analysis of learner profiles rather than isolated constructs. For the Working Memory and Phonological Short-Term Memory Capacity measures, automated span tasks were used, which is a methodological improvement on previous studies in Working Memory and Second Language Acquisition. The scoring techniques for these measures also focused on eliciting individual differences and identified the technique which was

most sensitive in this respect. Summarising the findings, individual differences in Working Memory and Phonological Short-Term Memory Capacity seem to play a role for the immediate post-test, but maybe less so for the Processing Instruction treatment and the delayed post-test. While these results are not generalisable to a wider population, they are in line with the hypothesis that the nature of Processing Instruction allows learners at different points of the Working Memory Capacity spectrum to do well during instruction. The performance differences on the post-tests are also in line with previous observations (Lee & McNulty, 2013).

The delayed post-test scores showed quite substantial attrition for some of the participants and on some tasks. However, most of the participants could demonstrate an improvement from pre-test to delayed post-test, especially on the less controlled tasks such as the essay task. This, again, is in line with previous research. VanPatten & Fernández (2004) administered a delayed post-test eight months after a Processing Instruction treatment targeting the First Noun Principle and found that overall, the participants did better on the delayed post-test than on the pre-test. Due to the morphological case marking, German word order is a lot more flexible than English word order, which means that the First Noun Principle can have an impact on successful interpretation of non-canonical word order in input sentences. The translation task tested the participants' reliance on word order. Here, most participants managed to improve substantially, but several factors seemed to have an impact on the results: Working Memory Capacity, exposure to German, and impulsivity. The latter variable, a dimension of the extraversion construct, has never been investigated within Processing Instruction research. Even in a wider context, extraversion as a personality variable has not received a lot of attention in Second Language Acquisition research lately (Dewaele & Furnham, 1999). Together with Eysenck's (1990) Arousal Theory, the findings lend support to previous observations that high anxiety and/or a high degree of introversion, which both translate as increased neural activation, decrease the chances of successful learning. The findings of the present study point to the important role that the personality of the learner plays and call for a more systematic investigation in the future.

Most of the Processing Instruction studies have employed binary options, the selection of alternative translations or pictures, and sentence-level tasks. This study is no exception. The instructional treatment was distributed over two weeks and based on the topic of Urban Life. The researcher has tried to develop a variety of Structured Input Activities which used textual as well as aural input. The tasks moved from sentence-level to discourse-level interpretation and exposed the participants to a substantial number of

tokens: The two sessions comprised 104 instances of switch prepositions in total. This is a considerable increase in number, compared to more recent Processing Instruction studies (see White, *in press*). While earlier studies measured the learning rate of participants receiving Processing Instruction by using a trials-to-criterion approach (e.g. Fernández, 2008; VanPatten & Borst, 2012; Lee, 2014), i.e. by counting how many times the participants failed to interpret the input correctly until they reached a pre-defined criterion for success, this study could pursue an even finer-grained approach: Thanks to the small n-size, it was possible to track the performance of each individual and for each Structured Input Activity. The performance scores during instruction show signs of an abundance of influences, for example effects of task novelty, task modality, lapses in attention, etc.

Not setting a criterion for success pays tribute to a processing perspective to learning as set out in the Modular On-line Growth and Use of Language framework (Sharwood Smith & Truscott, 2014), which has been adopted as a theoretical underpinning for this investigation. In this context, there is no end state of a learner's grammar as learning is a consequence of processing. The moment the individual is not exposed to the target language or the target feature any longer, the resting activation levels of the representations will decrease and make successful processing, both for interpretation and production, more difficult by increasing the competition from other languages and features. In that sense, any mastery or "acquisition" is temporary and dependent on the learner's exposure to input. This input, however, consists of only those features which the individual's (auditory or visual) processors detect. If the participants do not process a feature because their processors do not detect it, then it is unlikely to be acquired. Depending on the processing resources and internal factors such as Working Memory Capacity and overall activation levels, some participants' learning rates should be higher than others'. The study presented here used a multitude of measures, including Working Memory and Phonological Short-Term Memory Capacity measures, and can tentatively support the claim that Processing Instruction facilitates form-meaning connections, irrespective of Working Memory Capacity. The participant with the smallest Working Memory Capacity, nLQ6Vus, did not generally perform worse than the other participants during Processing Instruction.

Last but not least, this study addressed VanPatten et al.'s (2013) call for a closer investigation of German case marking and the impact of Working Memory, especially in connection with explicit information. All of the participants in this study received the same explicit information at the same time in the experiment, at the beginning of each session.

The information was kept brief and only focused on two things: 1. That word order is not a reliable cue for interpreting German sentences and 2. that in the context of German switch prepositions, the dative case expressed a location whereas the accusative case expressed a direction. While the Working Memory system would be involved in keeping this task-relevant information in an active and accessible state, the participants' performance during instruction did point to a more complex picture: Factors such as task familiarity, impulsivity versus reflectivity, and previous exposure to German input played a role. More contextual information, on the other hand, did not necessarily prove detrimental: The participant with the lowest Working Memory Capacity scores did remarkably well on discourse-level interpretation tasks. This points to the probability that mediating variables play a role in performance and can compensate for capacity 'deficiencies'. This should have practical implications for foreign language teaching, as will be further discussed in the next section.

The contributions of this study can be summarised as follows:

New findings for a small group of university students	<ul style="list-style-type: none"> • Working Memory Capacity does not predict performance on Structured Input Activities. • Working Memory Capacity does play a role in more complex tasks. • Processing Instruction leads to modest improvements from pre-test to delayed post-test, administered 15 weeks after instruction, indicating long-term effects. • Personality variables might play a bigger role than anticipated.
Added value	<ul style="list-style-type: none"> • New target feature: German switch prepositions • New variables, e.g. extraversion • Focus on interactions of learner variables • More sophisticated measures of Working Memory and Phonological Short-Term Memory Capacity • MOGUL as a theoretical framework for present and future research • Case study = opportunity for in-depth exploration • New classification of German switch prepositions • Further support for the effectiveness of Processing Instruction in virtual contexts • Further support for the effectiveness of Processing Instruction for non-PI assessment tasks

Alternative interpretations/ operationalisations	<ul style="list-style-type: none"> • Interpretation and operationalisation of Working Memory • Conceptualisation of “acquisition” • Process-orientation wherever possible • Focus on the individual in individual differences • Non-traditional methodology, e.g. case study involving qualitative and quantitative data from multiple sources
Question generation	<ul style="list-style-type: none"> • How do the variables interact exactly? • To what extent would the learner variables have a different impact on different learning and assessment tasks? • Would different modalities make a difference? To what extent? • What are optimal levels of activation, i.e. how much anxiety is beneficial for which personality type? When does activation become too much? • How do these abstract notions fit with findings from Event-Related Potential research, etc.? • Why does L2 German routinely generate findings in Processing Instruction research which are not comparable to findings involving other languages? • Would a different kind of explicit information, e.g. one based on cognitive grammar, lead to better results in combination with Processing Instruction?

Table 8.1: Contributions of this study to the field of Processing Instruction research

8.3 Limitations of this study and indications for future research

All in all, this case study on Processing Instruction and individual differences in Working Memory Capacity has employed several theoretical and methodological approaches which are fairly new to Processing Instruction research. The learning curve has been steep. As most research studies, the present one has generated new questions and problems, which merit further investigation in the future. Some of these endeavours will address the limitations of this small-scale case study, which will be discussed in the remainder of this section. This, then, is the most severe limitation of the research project: Due to the small n-size, it is not possible to generalise the findings to a wider population. The results, however, suggest that the impact and interactions of the selected learner variables are so complex that a generalisation even of a bigger sample size would not seem advisable. Or, as Snow (1991: 207) puts it: “[E]ach experiment is a case study,

limited to its time, place, and human constituents. The best way to understand its results and to link them to the results of other studies is to obtain the richest possible description of its context". This then leads to the question as to how we can further develop our systematic understanding of the effects of individual differences on Second Language Acquisition. One challenge already lies in the definition of successful acquisition:

Do the post-test scores imply that learning has taken place if the gains cannot be maintained from immediate post-test to delayed post-test in many cases? According to MOGUL, the answer is a clear yes. Learning, in MOGUL terms, is a consequence of processing. There is no fixed end state of language acquisition, so the processing of German switch prepositions and the case markers that followed will have resulted in a change of resting activation levels for these representations. Depending on the fierceness of the competition from other features and languages, this might have been enough to tip the scales towards the correct processing results, or not. Overall, the activated representations should have been more easily accessible, which is supported by the observation that there are gains from pre-test to post-test and that the participants produce overall more switch prepositions in the essay task directly after instruction. In the longer run, these gains cannot be sustained if the participants are not exposed to the features. It would be interesting to explore the long-term effects of Processing Instruction on the neural connections within the participant's brain. While Processing Instruction researchers are traditionally not interested in what exactly goes on in the learner's mind, Processing Instruction as a pedagogical intervention is suitable for Event-Related Potential measures because it relies on interpretation, not production. All responses could be conveyed with minimal motor movement, by just pressing one of two keys.

For some participants, a critical mass might have been reached with instruction, for others, the resting activation levels sunk below a level that could have made them successful in competition. The production scores exhibit the optionality that can persist long after the processing for interpretation is native-like. Ritterbusch et al. (2006) show that students can identify the gender of a noun if they are confronted with the form and explicit case information most of the time, that they can also give the case if they see the form and have gender information, but when presented with gender and case, they can only produce the accurate form in two out of three trials. The findings of the present study also indicate that students struggle more with the production of case marking than with its interpretation, although the success rate for the latter depends on the nature of the task itself. Even the bilingual participant who had spent much more time in an immersive setting than the other participants, did not interpret case markings accurately all the time.

As shown, this does not solely depend on language-related variables. Personality variables such as impulsivity impacted the consistency of the scoring. Arnett & Jernigan (2014: 72) note that “the description of the accusative prototype in English is similar to that of German. The dative case, however, is quite different in German than in English, even if we assume for the purpose of argument that English has a dative case”. While the relationships between representations are expressed by word order in English, they are expressed through case in German. The accusative case is rather easy to conceive as it relates to a direct object which undergoes some sort of change as the result of an action (=high transitivity). The performer of this action is the subject. The dative case, however, is less tangible. Dative objects are often the receivers of the direct object. They can, however, also be the experiencers or beneficiaries of an action (=low transitivity), meaning that they are to some extent affected by the action the subject performed (Arnett & Jernigan, 2014: 79–80). The findings of this study then can be seen as encouraging: Most participants improved significantly from pre-test to post-test on the translation task, which required the accurate interpretation of subject and object roles with respect to dative verbs with low-transitivity objects.

Another major question concerns the learner variables that were selected for inclusion in this study. Naturally, the variables under investigation represented only a small selection of a considerably larger complex of factors, some of which might not even be known to Second Language Acquisition research yet. The lack of interest in personality variables such as impulsivity and the impact observed in this study are a good reminder of the fact that we might be missing some crucial pieces of the language learning puzzle. For future research, it would be interesting to explore the relationship of the learner variables further and more systematically in order to see to what extent the variables are independent of each other. In the case of motivation, for example, which has not yielded ground-breaking insights in this study, one could look at the connection between learners’ motivation, their actions, and their degree of persistence in the face of difficulty. It would also be interesting to explore how stable the measured variables are. While a processing-orientation tends to focus on states rather than traits, it is the latter that are potentially problematic for language acquisition as they are not as easily influenced or changed (Snow, 1991: 205). The natural next step would therefore be to follow up this small-scale exploratory case study with a larger correlational study.

In the case of the present population, i.e. UK university students learning German as a Foreign Language, this will most likely not be possible though. A focus on other populations and/or languages seems therefore necessary. It would also be desirable to

find participant samples at different stages of their learning process and to explore whether the repeated exposure to Processing Instruction on the target feature could lead to more stable learning gains in the longer run. Another, maybe more feasible, route would be to replicate the present study with another small sample and collect additional eye-tracking data to see which parts of the slides the participants looked at and for how long. This information could then be used to develop better materials for computer-delivered Processing Instruction. The crucial point, however, is that further investigations should be integrated in a theoretical framework which allows the generation of a principled research agenda that systematically addresses the research questions generated previously.

The participant sample of this study is in many ways different from the usual participant samples in Processing Instruction, which is both an added bonus and a limitation. Instead of learners who are being exposed to the target feature for the first time or who have only a first language and the target language on their minds, the present sample consisted of mostly multilingual first-year and final-year university students enrolled in language studies. They were therefore no naïve participants and had considerable previous knowledge regarding language learning in a formal setting. The final year students, however, did not necessarily have an advantage over the first-year students as their most recent German language learning experience involved communicative tasks rather than the types of assessment tasks that were used for the pre- and post-tests. The immersion experiences were also more or less distributed over the first- and final year students, with the bilingual student being a special case. While the pre-test cut-off score of 60% for at least one measure, as is customary in Processing Instruction research, includes only participants who have not mastered the target form yet, this procedure is imprecise in determining the abilities of the participants, which can influence the outcomes in many ways and to different extents, as we have seen. For a small case study that allows an in-depth look into the complex interactions of several factors, this should be considered a source of information rather than noise in the data though.

While the computer-delivered Processing Instruction treatment served to control for teacher effects, it also created limitations: The participants received aural input, as suggested in the Processing Instruction guidelines (e.g. Farley, 2005: 33), but they had no opportunity to interact with each other. It stands to reason that the learner variables, especially anxiety and extraversion, would have had an impact on spontaneous conversation (Dewaele, 2013; Dewaele & Furnham, 1999) or other tests of oral production. Most likely, the nature of the sessions would have differed from what the

participants are used to from their usual foreign language classrooms, so ecological validity is somewhat compromised. Keeping in mind that language learning is both a cognitive and a social endeavour, the instructional setting of this study lacks social interaction between the learners and offers no opportunity to negotiate meaning. Both of these forms of dealing with input are important factors in successful (second) language acquisition. The creation of form-meaning connections could also have been supported by a more holistic language learning experience, for example by task-based or experiential approaches to teaching. According to MOGUL, the converging activation patterns in several perceptual output stores could have created very high activation levels for the representations, eventually leading to high resting activation levels and easier retrieval (Sharwood Smith & Truscott, 2014: 303–4).

It should be noted, however, that previous studies comparing classroom-based Processing Instruction and its computer-based counterpart have found no significant differences between the two modalities (Lee & Benati, 2007). As long as the research focus is on Processing Instruction, communication in the target language will not be a feature unless it is part of the assessment process. The use of affective Structured Input Activities, however, could have had a similar effect on the creation of form-meaning connections: Because affective structures have played an important part in human evolution, they have a strong influence on the individual's cognition. Following the MOGUL perspective, the representations which are activated through processing are co-indexed with the current affective state of the system (Sharwood Smith & Truscott, 2014: 176–7). Affective Structured Input Activities that make learners engage in meaningful ways and relate activities to family and friends should then inspire positive emotions which would be co-indexed with the representations and reinforce their activation levels. Future research should therefore include affective Structured Input Activities as well.

It is furthermore unclear from the data whether the lower performance scores towards the end of the second Processing Instruction session were due to fatigue or whether they can be attributed to the materials. Considering that these two tasks involved stative and directive verbs, which were slightly more opaque than the other constructions, the observed advantage of the two participants with the longest immersion experiences might point to a problem with the materials. Section 8.4 will present a new classification for locative prepositional verbs with flexible case selection, which should lead to a more principled approach to teaching German switch prepositions. The selection of verbs to be taught via Processing Instruction would then depend on the condition that there must be true optionality. For other cases, a cognitive grammar approach might yield better results.

The assessment materials would also have to be improved and better tailored to the participant sample. The essay task, for example, was not as complex and challenging as it could have been because the participants did not need to plan. The questions already pre-structured the text for them. To design a discourse-level task that requires planning and multiple, potentially competing, higher-level cognitive processes, the essay question would probably have to be more sophisticated. Participants would have to either write an argumentative essay or perform some sort of problem-solving task to elicit effects of individual differences in Working Memory Capacity. In the case of switch prepositions, the participants could receive a picture of an object and be prompted to hazard a guess as to what it could possibly be used for and how the instrument works. This should elicit switch prepositions and require higher-level reasoning at the same time. A similar task could then be designed for oral discourse-level production in order to detect potential influences of the output modality. The output, then, should not only be evaluated in terms of accuracy, but also with regards to the complexity of the discursive conventions (Galbraith, 2009: 11).

That being said, measuring performance does not allow precise deductions regarding the underlying abilities and competences, so everything that was measured in this study is indirect and mediated by a plethora of factors. Processing Instruction is claimed to lead to changes in the learner's developing system, but following a processing perspective, these changes should be seen as temporary at best. While it is conceivable that increased salience and non-redundancy makes the processing of the target feature more likely, thereby increasing the internal exposure to the feature, it is impossible to determine whether this is actually true. Looking at the bigger picture, it is conceivable that the development of discursive and pragmatic skills takes more time and exposure than a study such as this can cover. However, as Processing Instruction focuses on the form-meaning connections rather than the pragmatic context, any learning gains on the discourse level can be seen as a bonus. Further, more longitudinal research might provide better insights here.

This study did not, and did not attempt to, answer the question as to what the nature of Working Memory is. Extensive research in psychology and neurocognition has shown that it is possible to elicit capacity limitations that lie within a rather small range of potential values. The automated Memory Span tasks used in this study are a product of this line of enquiry. The question is whether it is advisable to do so, i.e. what we learn from the traditional definition of Working Memory Capacity about the role of Working Memory in second language acquisition. The MOGUL framework suggests that it is not necessary

to entertain the notion of Working Memory as a discrete system. Instead, it is integrated in the individual processing modules, which would then lead to an interpretation of Phonological Short-Term Memory as the Working Memory portion of the auditory module. The role of selective attention would then be determined by the environment and some innate primitives rather than the individual. These explanations will need to be investigated cross-disciplinarily in the future.

The instruments chosen to measure Working Memory Capacity, grammatical sensitivity, attitude and interest, personality, etc. come with their own limitations and problems, hence it can never be entirely clear whether findings (or lack thereof) are a result of the relationship of the variables or the artefacts of measurement. While the automated span tasks for the Working Memory measurement minimised errors that could have been committed by a researcher who is essentially naïve in terms of psychometric research, they were problematic in other respects: Some of the decisions which were made by the writers of the scripts were suitable for a larger sample size, but not for a case study. If the experiment were to be conducted again with a very small sample, then the researcher would make sure that the distractor sentences are presented in the same order for all participants as a randomised order only makes sense for a large sample.

From a psychometric perspective, the experiment is flawed because each task taps more than one factor, thus contaminating the results. Especially in the earlier studies in Second Language Acquisition research, there seems to be a quest for *the* one variable that predicts successful second language acquisition. Working Memory was seen as a worthy candidate because it has been conceptualised as “the place where input is perceived, attended to, and processed for subsequent representation in the developing system” (Sanz, 2005: 15). However, this study does not subscribe to the idea that Working Memory Capacity in the traditional sense can serve as a predictor of success. While it recognises the involvement of Working Memory in the tasks mentioned above, it does not support the concept of a Working Memory system that exists isolated from and independent of other (cognitive, behavioural, emotive, etc.) learner characteristics. Diversity was thus embraced rather than controlled for. This means that the study design is not as rigorous as a quasi-experimental design, but at the same time does not make optimum use of the individual’s perspective because it leaves little opportunity for introspection and is not entirely processing-oriented. Variables such as motivation, for example, are still conceptualised as static rather than dynamic.

8.4 Further suggestions for teaching German switch prepositions

German switch prepositions are notoriously difficult to learn because they compound several processing problems and they are difficult to teach because there seems to be no satisfying classification. Even generally successful methods, such as the use of metaphors or animations to teach aspects of grammar (see e.g. Roche & Scheller, 2004) fail to substantially increase the understanding of switch prepositions in some contexts. The distinction between motion and location is very common in German textbooks, but linguists such as Willems (2011) stress two problems with this categorisation: It is not as widely applicable as language teachers and learners are made to believe and the categorisation estimates a degree of semantic encoding that underestimates “the importance of inference and psychological associations of all kind” (p. 324). Based on corpus data from the Institut für Deutsche Sprache (IDS) in Mannheim, the authority in terms of German corpus research, he argues that a better categorisation of locative prepositional verbs with variable case marking would consider factors such as ambiguity, transparency versus opacity, as well as idiosyncratic versus versatile prepositional verbs. Considering the degree of linguistic analyses that went into this new categorisation, it is hard to see how this could possibly make the subject easier for language learners. It also raises the question whether input-based approaches such as Processing Instruction are suitable candidates for teaching these target features or whether explicit information is key. We will consider the factors that Willems discussed and develop a more straightforward categorisation that is based on cognitive grammar and that simplifies matters considerably.

The most straightforward category is the one of unambiguous prepositional verbs (Willems, 2011: 332–3). Those contain switch prepositions, but only select one case, such as *liegen + prep* (to lie + prep). It is not possible to use this prepositional verb with the accusative case. The next category that Willems suggests is that of ambiguous prepositional verbs, which can be transparent or opaque. The transparent ones are those which can be combined with either dative or accusative case, depending on whether they refer to a location or a destination. *Tauchen in* (to dive + prep) can refer to someone diving into the sea or diving in the sea (Willems, 2011: 334). The opaque ones are harder to explain as the verbs are more abstract in nature and no location versus direction distinction can be readily made, as in the case of *sich auflösen in* (Willems, 2011: 336). The problem with this example is that depending on the case selected, the semantics of the verb changes. *Sich auflösen in + dat* would translate as to dissolve in something,

whereas *sich auflösen* + acc would translate as to turn into something. This is the case for all the opaque examples that Willems (2011) offers. *Einführen in*, for example, can mean to introduce something (+dat) or to import something (+acc). Clearly, these verbs have developed their own semantics over time and could therefore be classified as homonyms rather than verbs with flexible case marking.

There are, however, verbs which Willems classifies as idiosyncratic and which have flexible case marking. Those accept either case without changing in meaning, the difference is rather one of perspective or focus. One example is the verb *verschwinden in* (= to disappear in/into). Choosing the accusative case emphasises the action of disappearing, whereas the dative case stresses the location of disappearance. Both choices are perfectly grammatical and motivated by the speaker/writer. This is a simplified view of source-path-goal image schemata, of course, and could be discussed within a cognitive grammar framework with advanced learners of German. In the same vein, subtle semantic differences which surface in the passive voice (*Vorgangspassiv* stressing the action, *Zustandspassiv* stressing the result) could be broached. If we exclude the locative prepositional verbs which take on different meanings depending on the case, arguing that they have developed into homonyms, then we can classify locative prepositional verbs in a rather straightforward manner: Some of them will have a fixed case and this case will depend entirely on the verb. Some have a strong preference for one case, which can regionally differ, but the motivation for this case preference is either not obvious any more or changes if the verb expresses a metaphorical meaning (e.g. *festnageln*: +dat for literally nailing something to something else, +acc for holding someone to their word). The third category, then, is the one with variable case marking.

This kind of optionality is pragmatically motivated in the sense that the case selection depends on whether the focus is on the action itself (+acc) or the location (+dat). Any action, in accordance with a source-path-goal image scheme, can be described in terms of the beginning, the change or activity, and the result. The same action can therefore be described using dative or accusative case and the choice is not motivated by the verb, i.e. motion verbs are not more likely to be expressed with accusative case than with dative case. The selection depends on the framing. This would then explain why even advanced learners of German exhibit difficulties with switch prepositions: From a MOGUL perspective, the competition is relatively even between the two cases as both combinations are instantiated in the input. This is one of the areas where explicit information might actually make a difference. If the distinction can be co-indexed with a representation in the conceptual store, outside the language module, then this conceptual

representation can provide the learners with better means of expression. The nature of the task, as always, has to be kept in mind. While the interpretation of case marking following locative prepositional verbs is relatively unambiguous, the production offers a fair amount of idiosyncratic optionality: In production, the learner creates the context. Depending on the focus or perspective that is adopted, the choice of case might differ. This optionality could lead to a delay in acquisition as no clear winner emerges from the competition.

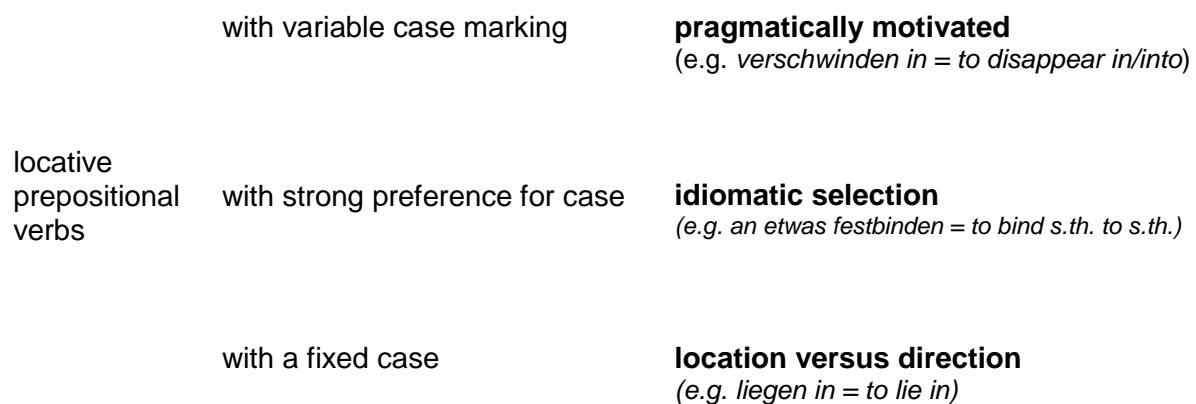


Figure 8.1: Suggested classification of locative prepositional verbs

For Processing Instruction, this means that some locative prepositional verbs are more suitable for inclusion in Structured Input Activities than others. Idiomatic expressions that involve locative prepositional verbs with a strong preference for one case make up the one category that is not suitable for Processing Instruction at all because the case selection can in many cases not be explained in terms of location versus direction. Furthermore, there is a higher degree of redundancy involved, as in the case of “*jemanden/ etwas an etwas festbinden*” (to bind someone/ something to something). Both the preposition *an* and the prefix *fest-* indicate that something will be attached to something else, facilitating overall the interpretation, but working counter to the efforts to create 1:1 form-meaning connections for Structured Input Activities. Traditionally, these constructions would be thematised in advanced-level classes of German, often through translation tasks.

Locative prepositional verbs with a fixed case, on the other hand, are no different from other verbs, they just happen to include a switch preposition. There is no optionality involved, so the case can be learnt in combination with the verb, e.g. “*liegen + da!*” (to

lie). Note that it is possible to combine this verb with any of the nine switch prepositions, but the case selected will always be dative. The opposite case selection is true for “*legen + acc*” (to lay). As explained before, a distinction of location versus direction is involved, but the verb determines the case, so this makes for slightly more complex form-meaning connections, as we have seen in the materials. Explicit information might play a role here and future research on Processing Instruction and German switch prepositions could target the question whether explicit information on boundary-crossing events could facilitate the creation of these form-meaning connections in a wider sense. In this study, the researcher avoided this explanation in order to make the explicit information involved comparable to that given in previous Processing Instruction materials. A future study could compare the two types of information and their effects on Processing Instruction.

The category of locative prepositional verbs which is best suited for Processing Instruction is clearly the one where true optionality is given, i.e. where switch prepositions can select either dative or accusative case, based on whether the focus is on direction or location. In this case, it is possible to create Structured Input Activities which keep the verb steady, but alternate the case, depending on whether the emphasis is on the location (dative case) or the direction (accusative case). This way, learners have to assign meaning to the case markers to process the sentence correctly. It is possible to supplement the explicit information, as mentioned above, and thus reinforce the notion that most, if not all, actions can be seen as a source-path-goal image. In terms of boundary-crossing events, this can be demonstrated as follows, for example in a discourse-level interpretation task that involves a series of drawings or pictures:

In one picture, a cat is walking on the pavement. To describe that action, dative case is needed because we can only describe the location in which the cat is doing the walking. A second picture could show the cat as is walks from the pavement onto the road. Now we need to use the accusative case because the cat is crossing the boundary between pavement and road. Whether we describe the direction where the cat came from or the direction in which it is going, we will need to use the accusative case. A third picture then could show the cat continuing its walk on the road. To describe this action, we would again use dative case because we are focusing on the location in which the action is being performed. An activity like that would have several advantages: it would boil down the case alternation to a relatively straightforward location versus direction distinction. It would also drive home the point that this distinction is not motivated by the characteristics of the verb, i.e. that motion verbs are not by default associated with accusative case. On a higher level, it offers the opportunity for a discussion of pragmatic issues: Depending

on what the speaker/writer focuses on, either case is possible. Someone describing the actions of the cat would follow the complete path of the cat, whereas someone driving on the road might emphasise that the cat crossed the boundary and walked onto the road rather than its previous location. None of these different framings change the actual path of the cat.

8.5 Practical applications

Ever since New Labour came into power in 1997, English schools have received massive funding for computer hardware as well as infrastructure development. As early as in the year 2000, schools were equipped with computer labs that were used for foreign language learning. But they employed the old-fashioned mechanical drills and were neither responsive to the individual learner's needs, nor communicative in nature. These days, attempts are being made to align computer-delivered instruction and assessment and there is more access to performance data than ever. But this will not help the students as long as the data accumulated are used to satisfy the government that the targets are met rather than for timely feedback to the students. There is a whole community of students from all over the world who share materials, advice, and motivation online. They supplement what they learn in school with input and help from native speakers. With more and more people learning languages independently online, and the increase of Blended Learning opportunities offered by educational establishments, the development of authentic and effective computer-delivered instruction is of paramount importance. With online language learning platforms such as Duolingo, the otherwise outdated grammar translation method becomes popular again. It is therefore crucial to develop alternative tasks which are informed by Second Language Acquisition research and designed to meet the learners' communicative needs. Processing Instruction could be such an alternative as it has been proven to be beneficial for learners of different backgrounds studying a wide range of different languages. Previous studies have employed computer-delivered Processing Instruction successfully (e.g. VanPatten & Borst, 2012) and without significantly different outcomes than teacher-delivered Processing Instruction (Lee & Benati, 2007).

The present study has shown that it is possible to use available content-delivery software such as Adobe Captivate to create Structured Input Activities which follow the guidelines set out by Farley (2005) and others. There is no need for language teachers to use the

more elaborate psycholinguistic software described in previous experiments. The participants' performance improvements from pre- to post-test are an indication of the beneficial effects of Processing Instruction, even for more traditional output-oriented tasks such as the essay task. It stands to reason that the participants might have maintained or even improved their instructional gains if they had received top-up Processing Instruction between the immediate and the delayed post-tests (Hikima, 2010). Research on cultural differences in computer-based learning contexts have shown that the use of e-learning materials can lead to changes in the way learners interact with the materials, creating a heightened interest that results in more self-directed approaches to learning (Roche & Todorova, 2010).

Making the instructional sessions on German switch prepositions available for independent study would therefore benefit the learners and add to the resources available for student-centred learning and independent revision. The participants in this study did not experience the computer-delivered Processing Instruction as strenuous or particularly challenging. This might make the aforementioned revision more likely to happen. Learners could also take a more active approach and create their own interpretation tasks by taking pictures of locations and boundary-crossing events, finding appropriate descriptions, and then quizzing each other. This would increase the range of materials and at the same time ensure that the pictures are of interest to the learners. One advantage of linking programmes such as Adobe Captivate to Learning Management Systems is that the students' progress can be monitored individually. This makes timely feedback and formative rather than summative evaluation more feasible. It also means that the exercises and quizzes can be selected based on the learner's previous performance and overall learning trajectory. The wider application of computer-delivered Processing Instruction materials could also have implications for future research. If standardised instructional sessions could be delivered across a wide range of students in different countries and different educational institutions, at no additional cost, and if the learners' on-task performance could be recorded, then this would provide some desperately needed critical mass for more sophisticated data analyses.

To end with the words of DeKeyser (2012: 197):

Extending this perspective over much longer stretches of learning, and studying how the learning processes (and their results) evolve through the interaction of individual differences such as aptitudes and age with aspects of input and instruction, should be high on our agenda for the years to come, both for the psycholinguistic insights it may yield and for the instructional strategies it may suggest, in particular now that various forms of technology are beginning to make true individualization more feasible than before.

REFERENCES

- Admissions Testing Service. (2014, November 05). *University of Oxford*. Retrieved May 01, 2016, from <https://www.ox.ac.uk/admissions/undergraduate/applying-to-oxford/tests?wssl=1>
- Agiasophiti, Z. (2011). *An empirical psycholinguistic investigation of input processing and input enhancement in L1 English/L2 German: the acquisition of V2 and accusative case marking*. Unpublished PhD dissertation.
- Agiasophiti, Z. (2013). Exploring Possible Effects of Gender and Enhanced vs. Unenhanced Processing Instruction on the Acquisition of Case Marking in L2 German. In J. F. Lee, & A. G. Benati (Eds.), *Individual Differences and Processing Instruction* (pp. 153–184). Sheffield: Equinox.
- Altarriba, J., & Isurin, L. (2013). *Memory, language, and bilingualism: theoretical and applied approaches* (1st ed.). Cambridge and New York: Cambridge University Press.
- Angelovska, T., & Benati, A. G. (2013). Processing instruction and the age factor: Can adults and school-age native speakers of German process English simple past tense correctly? In J. F. Lee, & A. G. Benati (Eds.), *Individual Differences and Processing Instruction* (pp. 131–152). Sheffield: Equinox.
- Arnett, C., & Jernigan, H. (2014). A Cognitive Grammar Account of Case for L2 Students of German. *GFL German as a foreign language journal*, 1, 68–93.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence, & J. T. Spence (Eds.), *The psychology of learning and motivation (Volume 2)* (pp. 89–195). New York: Academic Press.
- Baars, B. J. (1998). Metaphors of consciousness and attention in the brain. *Trends in Neurosciences*, 21, 58–62.
- Baddeley, A. D. (1986). *Working Memory* (1st ed.). Oxford: Clarendon Press.
- Baddeley, A. D. (1996). The Fractionation of working memory. *Proceedings of the US National Academy of Sciences*, 93(24), 13468–13472.
- Baddeley, A. D. (1998). Recent developments in working memory. *Current Opinion in Neurobiology*, 8, 234–238.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 11(4), 417–423.
- Baddeley, A. D. (2002). Fractionating the central executive. In D. Stuss, & R. T. Knight (Eds.), *Principles of Frontal Lobe function* (pp. 246–260). New York: Oxford University Press.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory (Volume 8)* (pp. 47–89). New York: Academic Press.
- Bartolotti, J., & Marian, V. (2013). Bilingual memory: structures, access, and processing. In J. Altarriba, & L. Isurin (Eds.), *Memory, Language, and Bilingualism. Theoretical and Applied Approaches* (pp. 7–47). Cambridge and New York: Cambridge University Press.

- Bates, E., & MacWhinney, B. (1981). Second language acquisition from a functionalist perspective: Pragmatic, semantic and perceptual strategies. In H. Winitz (Ed.), *Annals of the New York Academy of Sciences conference on native and foreign language acquisition* (pp. 190–214). New York: New York Academy of Sciences.
- Benati, A. G. (2001). A comparative study of the effects of processing instruction and output-based instruction on the acquisition of the Italian future tense. *Language Teaching Research*, 5(2), 95–127.
- Benati, A. G. (2005). The effects of processing instruction, traditional instruction and meaning—output instruction on the acquisition of the English past simple tense. *Language Teaching Research*, 9(1), 67–93.
- Benati, A. G. (2007). The effects of Processing Instruction on the lexical preference principle: generalizability of findings. *RiPLA Rivista di Psicolinguistica Applicata Journal of Applied Psycholinguistics*, 1(2), 93–104.
- Benati, A. G. (2009). *Issues in Second Language Proficiency* (1st ed.). London: Continuum.
- Benati, A. G. (2013). *Issues in Second Language Teaching* (1st ed.). Sheffield: Equinox.
- Benati, A. G., Laval, C., & Arche, M. J. (2014). *The Grammar Dimension in Instructed Second Language Learning. Advances in Instructed Second Language Acquisition Research* (1st ed.). London: Bloomsbury Academic.
- Benati, A. G., & Lee, J. F. (2008). *Grammar Acquisition and Processing Instruction: Secondary and Cumulative Effects*. Clevedon: Multilingual Matters.
- Benati, A. G., & Lee, J. F. (2009). *Research and Perspectives on Processing Instruction (Studies on Language Acquisition (SOLA))*. Berlin, Boston: Mouton de Gruyter.
- Benati, A. G., & Lee, J. F. (2010). *Processing instruction and discourse*. *Continuum Studies in Linguistics*. London: Continuum.
- Benati, A. G., Lee, J. F., & Houghton, S. D. (2008). From Processing Instruction on the acquisition of English past tense to secondary transfer-of-training effects on English third person singular present tense. In A. G. Benati, & J. F. Lee, *Grammar Acquisition and Processing Instruction. Secondary and Cumulative Effects* (pp. 88–120). Bristol: Multilingual Matters.
- Benati, A. G., Lee, J. F., & Laval, C. (2008). From Processing Instruction on the Acquisition of French Imparfait to Secondary Transfer-of-Training Effects on French Subjunctive and to Cumulative Transfer-of-Training Effects with French Causative Constructions. In A. G. Benati, & J. F. Lee, *Grammar Acquisition and Processing Instruction. Secondary and Cumulative Effects* (pp. 121–157). Bristol: Multilingual Matters.
- Benati, A. G., Lee, J. F., & McNulty, E. (2010). Exploring the effects of processing instruction on a discourse-level guided composition with the Spanish subjunctive after the adverb cuando. In A. G. Benati & J. F. Lee, *Processing instruction and discourse* (97–147). London: Continuum.
- Berken, J. A., Chai, X., Chen, J.-K., Gracco, V. L., & Klein, D. (2016). Effects of Early and Late Bilingualism on Resting-State Functional Connectivity. *The Journal of Neuroscience*, 36(4), 1165–1172.

- Bever, T. G. (1970). The cognitive basis for linguistic structure. In J. R. Hayes (Ed.), *Cognition and the Development of Language* (pp. 279–362). New York: John Wiley.
- Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Kroll & A. M. B. DeGroot (Eds.), *Handbook of bilingualism: Psycholinguistic perspectives* (pp. 109–127). New York: Oxford University Press.
- Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, 56(s1), 9–49.
- Buck, M. (2000). *Procesamiento del lenguaje y adquisición de una segunda lengua. Un estudio de la adquisición de un punto gramatical en inglés por hispanohablantes*. Unpublished doctoral dissertation, Universidad Nacional Autónoma de México.
- Cadierno, T. (1995). Formal instruction in processing perspective: An investigation into the Spanish past tense. *The Modern Language Journal*, 79, 179–94.
- Carroll, J. B. (1965). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training, Research, and Education* (pp. 87–136). Pittsburgh, PA: University of Pittsburgh Press.
- Carroll, J. B. (1971). *Implications for Aptitude Test Research and Psycholinguistic Theory for Foreign Language Teaching*. Princeton, NJ: Educational Testing Service.
- Carroll, J. B., & Sapon, S. M. (1959). *Modern Language Aptitude Test, Form A*. New York: The Psychological Corporation.
- Chai, X. J., Berken, J. A., Barbeau, E. B., Soles, J., Callahan, M., Chen, J.-K., & Klein, D. (2016). Intrinsic Functional Connectivity in the Adult Brain and Success in Second-Language Learning. *The Journal of Neuroscience*, 36(3), 755–761.
- Chein, J. M., Ravizza, S. M., & Fiez, J. A. (2003). Using neuroimaging to evaluate alternative models of working memory and their implications for language processing. *Journal of Neurolinguistics* 16, 315–339.
- Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: The limit in words retained without covert articulation. *The Quarterly Journal of Experimental Psychology* 62, 1420–1429.
- Cheng, A. (1995). *Grammar instruction and input processing: the acquisition of Spanish ser and estar*. Unpublished doctoral thesis: The University of Illinois at Urbana-Champaign.
- Cheng, A. (2002). The effects of processing instruction on the acquisition of ser/estar. *Hispania*, 85, 308–323.
- Cheng, A. (2004). Processing instruction and Spanish ser and estar: forms with semantic-aspectual values. In B. VanPatten (Ed.), *Processing instruction: theory, research, and commentary* (pp. 119–141). Mahwah, NJ: Erlbaum.
- Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory and Cognition*, 30(7), 1086–1095.
- Collentine, J.G. (1998). Processing Instruction and the Subjunctive. *Hispania*, 81, 576–587.

- Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786.
- Cooper, C. (2002). *Individual Differences* (2nd ed.). London: Arnold.
- Corder, P. (1967). The significance of learner's errors. *International Review of Applied Linguistics*, 5, 161–170.
- Cowan, N. (1993). Activation, attention, and short-term memory . *Memory & Cognition*, 21, 162–167.
- Cowan, N. (1995). *Attention and Memory. An Integrated Framework. Oxford Psychology Series No. 26*. New York: Oxford University Press.
- Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake, & P. Shah (Eds.), *Models of Working Memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge: Cambridge University Press.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185.
- Cowan, N. (2005). *Working memory capacity* (1st ed.). Hove: Psychology Press.
- Cowan, N. (2015). Sensational memorability: Working memory for things we see hear, feel, or somehow sense. In C. LeFebvre, P. Jolicœur, & J. Martinez-Trujillo (Eds.), *Mechanisms of Sensory Working Memory. Attention and Perfomance XXV* (pp. 5–22). Elsevier.
- Culman, H., Henry, N., & VanPatten, B. (2009). The Role of Explicit Information in Instructed SLA: An On-Line Study with Processing Instruction and German Accusative Case Inflections. *Die Unterrichtspraxis/ Teaching German*, 42(1), 19–31.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- DeKeyser, R. M. (2000). The Robustness of Critical Period Effects in Second Language Acquisition. *Studies in Second Language Acquisition*, 22, 499–533.
- DeKeyser, R. M. (2005). What makes second-language grammar difficult? A review of issues. *Language Learning*, 55, Supplement 1, 1–25.
- DeKeyser, R. M. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, 2(62), 189–200.
- DeKeyser, R. M., & Sokalski, K. J. (1996). The Differential Role of Comprehension and Production Practice. *Language Learning*, 46(4), 613–642.
- Dewaele, J.-M. (2010). *Emotions in Multiple Languages* (1st ed.). Basingstoke: Palgrave Macmillan.
- Dewaele, J.-M. (2013). The Link Between Foreign Language Classroom Anxiety and Psychoticism, Extraversion, and Neuroticism Among Adult Bi- and Multilinguals. *The Modern Language Journal*, 97(3), 670–684.
- Dewaele, J.-M., & Furnham, A. (1999). Extraversion: The Unloved Variable in Applied Linguistics Research. *Language Learning*, 49(3), 509–544.

- Domínguez Vázquez, M. J. (2005). Die Betrachtung der deutschen Präpositionen aus didaktischer Sicht. Vorschläge für eine neue Behandlung im DaF-Unterricht. *GFL German as a foreign language journal*, 1, 60–78.
- Dörnyei, Z. (2001). *Teaching and Researching Motivation. Applied Linguistics in Action Series* (1st ed.). Harlow: Pearson Education.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. (1st ed.). Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. (1st ed.). Oxford: Oxford University Press.
- Dörnyei, Z. (2009). *The Psychology of Second Language Acquisition* (1st ed.). Oxford: Oxford University Press.
- Dörnyei, Z. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd and extended ed.). London: Routledge.
- Dörnyei, Z., & Csizér, K. (2012). How to design and analyze surveys in SLA research? In A. Mackey, & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 74–94). Malden, MA: Wiley-Blackwell.
- Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty, & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 589–630). Oxford: Blackwell.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, N. C., & Sinclair, S. G. (1996). Working Memory in the Acquisition of Vocabulary and Syntax: Putting Language in Good Order. *The Quarterly Journal of Experimental Psychology*, 49A(1), 234–250.
- Elman, J. L. (1997). *Rethinking Innateness: A Connectionist Perspective on Development (Second Edition)*. Cambridge, MA: MIT Press.
- Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science*, 11(1), 19–23.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309–31.
- Ericcson, K. A., Chase, W. G., & Faloon S. (1980). Acquisition of a memory skill. *Science*, 208(4448), 1181–2.
- Erlam, R. (2003a). The Effects of Deductive and Inductive Instruction on the Acquisition of Direct Object Pronouns in French as a Second Language. *The Modern Language Journal*, 87(2), 242–60.
- Erlam, R. (2003b). Evaluating the Relative Effectiveness of Structured-Input and Output-Based Instruction in Foreign Language Learning (Results from an experimental study). *Studies in Second Language Acquisition*, 25(4), 559–82.
- Eysenck, H. J. (1967). *The biological basis of personality* (1st ed.). Springfield, IL: Charles C. Thomas.

- Eysenck, H. J. (1990). Biological dimensions of personality. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 244–276). New York: Guildford.
- Eysenck, H. J., & Eysenck, M. W. (1985). Personality and individual differences: a natural science approach. New York: Plenum Press.
- Eysenck, H. J., & Eysenck, S. B. (1981). *Manual for the Eysenck Personality Inventory*. Kent: Hodder and Stoughton.
- Eysenck, S. B. G., & Eysenck, H. J. (1964). An improved short questionnaire for the measurement of extraversion and neuroticism. *Life Sciences*, 305, 1103–1109.
- Farley, A. P. (2000). Processing and meaning output-based instruction: the acquisition of Spanish subjunctive. Unpublished doctoral dissertation, University of Illinois at Urbana Champaign.
- Farley, A. (2001a). The effects of processing instruction and meaning-based output instruction. *Spanish Applied Linguistics*, 5, 57–94.
- Farley, A. (2001b). Authentic processing instruction and the Spanish subjunctive. *Hispania*, 84, 289–99.
- Farley, A. P. (2004). The relative effects of processing instruction and meaning-based output instruction. In B. VanPatten (Ed.), *Processing instruction: theory, research, and commentary* (pp. 143–168). Mahwah, NJ: Lawrence Erlbaum & Associates.
- Farley, A. P. (2005). *Structured Input: Grammar Instruction for the Acquisition-Oriented Classroom* (McGraw-Hill Second Language Professional Series. Perspectives on theory and research). New York and Boston: McGraw-Hill.
- Fernández, C. (2008). Reexamining the Role of Explicit Information in Processing Instruction. *Studies in Second Language Acquisition*, 30, 277–305.
- Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent variable analysis. *Journal of Experimental Psychology: General*, 133, 101–135.
- Galbraith, D. (2009). Cognitive Models of Writing. *GFL German as a foreign language journal*, 2(3), 7–22.
- Gardner, R. C. (1958). *Social factors in second-language acquisition*. Unpublished Master's thesis, McGill University, Montreal, Canada.
- Gardner, R. C. (1960). *Motivational variables in second-language acquisition*. Unpublished Doctoral dissertation, McGill University, Montreal, Canada.
- Gardner, R. C. (1985). *Social Psychology and Second Language Learning. The Role of Attitudes and Motivation* (1st ed.). London: Edward Arnold.
- Gardner, R. C. & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning*. Rowley, MA: Newbury House Publishers.
- Gass, S. M. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics*, 9, 198–217.
- Gass, S. M. (1997). *Input, interaction, and the second language learner* (1st ed.). Mahwah, NJ: Lawrence Erlbaum.

- Gass, S. M., & Mackey, A. (2007). *Data Elicitation for Second and Foreign Language Research*. Mahwah, NJ: Erlbaum.
- Gass, S. M., & Mackey, A. (2007). *Data Elicitation for Second and Foreign Language Research (Second Language Acquisition Research Series)* (1st ed.). Mahwah, NJ: Lawrence Erlbaum.
- Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of the Psychology of Education*, 8, 259–272.
- Grabe, W. (2009). *Reading in a Second Language. Moving from Theory to Practice*. Cambridge applied linguistics series. New York: Cambridge University Press.
- Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General*, 136(2), 217–240.
- Henry, N., Culman, H., & VanPatten, B. (2009). More on the effects of explicit information in instructed SLA: A partial replication and a response to Fernández 2008. *Studies in Second Language Acquisition*, 31, 559–575
- Hikima, N. (2010). *The effects of processing instruction and re-exposure on interpretation discourse level tasks: the case of Japanese passive forms*. University of Greenwich. Unpublished PhD Dissertation.
- Isurin, L. (2013). Memory and first language forgetting. In J. Altarriba, & L. Isurin (Eds.), *Memory, Language, and Bilingualism. Theoretical and Applied Approaches* (pp. 319–348). Cambridge and New York: Cambridge University Press.
- Jackson, C. (2008). Proficiency Level and the Interaction of Lexical and Morphosyntactic Information During L2 Sentence Processing. *Language Learning*, 58(4), 875–909.
- Johnson, J. S., & Newport, E. L. (1989). Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language. *Cognitive Psychology*, 21, 60–99.
- Juffs, A., & Harrington, M. (1996). Garden Path Sentences and Error Data in Second Language Sentence Processing. *Language Learning*, 46(2), 283–323.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 98, 122–149.
- Just, M. A., Carpenter, P. A., Keller, T. A., Emery, L., & Zajac, H. (2001). Interdependence of nonoverlapping cortical systems in dual cognitive tasks. *Neuroimage*, 14, 417–426.
- Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, 9(4), 637–671.
- Kane, M. J., Conway, A. R., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. Conway, C. Jarrold, M. J. Kane, & A. Miyake (Eds.), *Variation in Working Memory* (pp. 21–48). New York: Oxford University Press.

- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach. *Journal of Experimental Psychology: General*, 133(2), 189–217.
- Krashen, S. D. (1982). *Principles and Practice Second Language Acquisition* (1st ed.). Oxford: Pergamon.
- Laval, C. (2008). Measuring primary, secondary and cumulative effects of processing instruction in the acquisition of French. PhD thesis, University of Greenwich.
- Laval, C. (2013) The age factor on the primary, secondary and cumulative transfer-of-training effects of processing instruction on the acquisition of French as a second language. In J. F. Lee, & A. G. Benati (Eds.), *Individual Differences and Processing Instruction* (pp. 105–130). Sheffield: Equinox.
- Lee, J. F. (1987). Morphological factors influencing pronominal reference assignment by learners of Spanish. In T. A. Morgan, J. F. Lee, & B. VanPatten (Eds.), *Language and language use: Studies in Spanish* (pp. 221–232). Lanham, MD: University Press of America.
- Lee, J. F. (2004). On the Generalizability, Limits, and Potential Future Directions of Processing Instruction Research. In B. VanPatten (Ed.), *Processing Instruction. Theory, Research, and Commentary* (pp. 311–323). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lee, J. F., & Benati, A. G. (2007a). *Delivering Processing Instruction in Classrooms and in Virtual Contexts. Research and Practice* (1st ed.). London: Equinox.
- Lee, J. F., & Benati, A. G. (2007b). *Second Language Processing. An analysis of Theory, Problems and Possible Solutions* (1st ed.). London: Continuum.
- Lee, J. F., & Benati, A. G. (2009). *Research and perspectives on processing instruction. Studies on Language Acquisition (SOLA)*, 36. Berlin and New York: Mouton de Gruyter.
- Lee, J. F., & Benati, A. G. (2013) *Individual differences and processing instruction*. Sheffield: Equinox.
- Lee, J. F., & McNulty, E. M. (2013). The Effects of Language Background on the Results of Processing Instruction on the Spanish Subjunctive/Indicative Contrast after the Adverb cuando. In J. F. Lee, & A. G. Benati, *Individual Differences and Processing Instruction* (pp. 49–81). Sheffield: Equinox.
- Lee, J. F., Benati, A. G., Aguilar-Sánchez, J., & McNulty, E. (2007). Comparing three modes of delivering Processing Instruction on preterite/imperfect distinction and negative informal commands in Spanish. In J. F. Lee, & A. G. Benati (Eds.), *Delivering Processing Instruction in Classrooms and Virtual Contexts: Research and Practice* (pp. 73–98). London: Equinox.
- Leeser, M., & DeMil, A. (2013). Investigating the Secondary Effects of Processing Instruction in Spanish: From Instruction on Accusative Clitics to Transfer-of-Training Effects on Dative Clitics. *Hispania*, 96(4), 748–62.
- Leino, A.-L. (1974). *English School Achievements and Some Student Characteristics II. On the Relationships of Motivational and Auditory Variable to English School Achievements. Final Analyses*. Helsinki: Institute of Education, University of Helsinki.
- Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

- Leow, R. P., & Bowles, M. A. (2005). Attention and awareness in SLA. In C. Sanz (Ed.), *Mind and Context in Adult Second Language Acquisition. Methods, Theory, and Practice* (pp. 179–203). Washington, DC: Georgetown University Press.
- Leow, R. P., Hsieh, H.-C., & Moreno, N. (2008). Attention to Form and Meaning Revisited. *Language Learning*, 58, 665–695.
- LoCoco, V. (1987). Learner comprehension of oral and written sentences in German and Spanish: The importance of word order. In B. VanPatten, T. R. Dvorak, & J. F. Lee (Eds.), *Foreign language learning: A research perspective* (pp. 119–129). Cambridge, MA: Newbury House.
- Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, 130(2), 199–207.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review*, 109, 35–54.
- MacWhinney, B. (1987). The competition model. *Mechanisms of language acquisition*, 249–308.
- Malovrh, P. A. (2014). Variability and systematicity in interlanguage development: An analysis of mode and its effect on L2 Spanish morphology. *Studies in Hispanic and Lusophone Linguistics*, 1(7), 43–78.
- Martin, K. I., & Ellis, N. C. (2012). The Roles of Phonological Short-term Memory and Working Memory in L2 Grammar and Vocabulary Learning. *Studies in Second Language Acquisition*, 34, 379–413.
- Mavrantoni, M., & Benati, A. G. (2013). The effects of Processing Instruction and traditional instruction on two different school-age learners: The case of English present simple tense third person singular. In J. F. Lee & A. G. Benati (Eds.), *Individual differences and Processing Instruction* (pp. 185–210). Sheffield: Equinox.
- Melby-Lervåg, M., & Hulme, C. (2013). Is Working Memory Training Effective? A Meta-Analytic Review. *Developmental Psychology*, 48(2), 270–291.
- Melby-Lervåg, M., & Lervåg, A. O. (2012). Oral Language Skills Moderate Nonword Repetition Skills in Children with Dyslexia: A Meta-analysis of the role of Nonword Repetition Skills in Dyslexia. *Scientific Studies of Reading*, 16(1), 1–34.
- Mensel, L. (2011). *Städte machen uns menschlicher*. Retrieved from The European - Das Debatten Magazin: <http://www.theeuropean.de/glaeser-edward/8470-vorteile-von-staedten>
- Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The psychological review*, 63, 81–97.
- Morgan-Short, K., & Wood Bowden, H. (2006). Processing Instruction and Meaningful Output-Based Instruction: Effects on Second Language Development. *Studies in Second Language Acquisition*, 28(1), 31–65.
- Nam, E. (1975). Child and adult perceptual strategies in second language acquisition. Paper presented at the 1975 TESOL Convention, Los Angeles.

- Pienemann, M. (1984). Psychological Constraints on the Teachability of Languages. *Studies in Second Language Acquisition*, 6(2), 186–214.
- Pienemann, M. (1998). *Language Processing and Second Language Development. Studies in Bilingualism*, 15. John Benjamins.
- Pimsleur, P., Sundland, D. M., & McIntyre, R. D. (1966). *Underachievement in foreign language learning* (1st ed.). Washington, DC: Modern Language Association.
- Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring Working Memory Capacity With Automated Complex Span Tasks. *European Journal of Psychological Assessment*, 28(3), 164–171.
- Ricker, T. J., AuBuchon, A. M., & Cowan, N. (2010). Working memory. *WIREs Cognitive Science*, 1(4), 573–585.
- Ritterbusch, R., LaFond, L., Agustin, M. (2006). Learner Difficulties with German Case: Implications from an action research study. *Die Unterrichtspraxis/ Teaching German*, 39, 30–45.
- Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, 17(4), 368–392.
- Robinson, P. (2002). Learning conditions, aptitude complexes, and SLA: A framework for research and pedagogy. In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning. Language Learning and Language Teaching*, 2 (pp. 113–133). London: John Benjamins.
- Roche, J., & Scheller, J. (2004). Zur Effizienz von Grammatikanimationen beim Spracherwerb: Ein empirischer Beitrag zu einer kognitiven Theorie des multimedialen Fremdsprachenerwerbs. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 9(1), 1–15.
- Roche, J., & Todorova, D. (2010). The Role of Cultural Diversity in E-Based Language Learning. A comparative study of Bulgarian and Lithuanian learners of German as a foreign language using an online learning platform. In F. Sudweeks, H. Hrachovec, & C. Ess (Eds.), *Proceedings of the Seventh International Conference on Cultural Attitudes towards Technology and Communication Vancouver, Canada, 15–18 June 2010* (pp. 436–445). Murdoch: Murdoch University.
- Rocklin, T., & Revelle, W. (1981). The measurement of extraversion: A comparison of the Eysenck Personality Inventory and the Eysenck Personality Questionnaire. *British Journal of Social Psychology*, 20, 279–284.
- Salaberry, R. 1997. The role of input and output practice in second language acquisition. *Canadian Modern Language Review* 53, 422–451.
- Santamaría, K. (2007). *Making Form-Meaning Connections: The Influence of Instruction and Working Memory on L2 French Clitic Acquisition*. Tallahassee. Unpublished PhD Dissertation.
- Sanz, C. (2005). Adult SLA: The interaction between external and internal factors. In C. Sanz, *Mind and context in adult second language acquisition: Methods, Theory, and Practice* (pp. 3–21). Washington, DC: Georgetown University Press.

- Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning* 54(1), 35–78.
- Schlak, T. (2004). Zur Rolle expliziter Grammatikvermittlung im universitären DaF-Unterricht: Eine qualitativ-ethnographische Fallstudie. *GFL German as a foreign language journal*, 2, 40–80.
- Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt, *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu: Second Language Teaching and Curriculum Center, University of Hawai'i.
- Sharwood Smith, M. (2004). In two minds about grammar: On the interaction of linguistic and metalinguistic knowledge in performance. *Transactions of the Philological Society*, 102(2), 255–280.
- Sharwood Smith, M. (2013). Only Connect: The Interface Debate. In K. Droźdzał-Szelest, & M. Pawlak, *Psycholinguistic and Sociolinguistic Perspectives on Second Language Learning and Teaching* (pp. 27–39). Berlin & Heidelberg: Springer.
- Sharwood Smith, M., & Truscott, J. (2005). Stages or Continua in Second Language Acquisition: A MOGUL Solution. *Applied Linguistics*, 26, 219–240.
- Sharwood Smith, M., & Truscott, J. (2014). *The Multilingual Mind. A Modular Processing Perspective* (1st ed.). Cambridge: Cambridge University Press.
- Simmering, V. R., & Perone, S. (2013). Working memory capacity as a dynamic process. *Frontiers in Developmental Psychology*, 3:567, doi: 10.3389/fpsyg.2012.00567.
- Skehan, P. (1989). *Individual Differences in second language learning* (1st ed.). London: Arnold.
- Skehan, P. (1991). Individual Differences in Second Language Learning. *Studies in Second Language Acquisition*, 13, 275–298.
- Smith, M., & VanPatten, B. (2013). Instructed SLA as parameter setting: Evidence from earliest-stage learners of Japanese as L2. In A. G. Benati, C. Laval, & M. J. Arche (Eds.), *The Grammar Dimension in Instructed Second Language Learning. Advances in Instructed Second Language Learning* (pp. 127–46). London: Bloomsbury.
- Snow, R. E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Researcher*, 18(9), 8–14.
- Snow, R. E. (1991). Aptitude-Treatment Interaction as a Framework for Research on Individual Differences in Psychotherapy. *Journal of Consulting and Clinical Psychology*, 59(2), 205–216.
- Sparks, R. L., Patton, J., & Ganschow, L. (2012). Profiles of more and less successful L2 learners: A cluster analysis study. *Learning and Individual Differences*, 22, 463–472.
- Spencer, J. P., Austin, A., Schutte, A. R. (2012). Contributions of dynamic systems theory to cognitive development. *Cognitive Development*, 27, 401–418.

- Szmałec, A., Brysbaert, M., & Duyck, W. (2013). Working memory and (second) language processing. In J. Altarriba, & L. Isurin (Eds.), *Memory, Language, and Bilingualism. Theoretical and Applied Approaches* (pp. 74–94). Cambridge and New York: Cambridge University Press.
- Tanner, D. (2013). Individual differences and streams of processing. *Linguistic Approaches to Bilingualism*, 3(3), 350–356.
- Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologica*, 56, 289–301.
- Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, 17(2), 277–293.
- Truscott, J., & Sharwood Smith, M. (2004). Acquisition by processing: A modular perspective on language development. *Bilingualism: Language and Cognition*, 7, 1–20.
- Truscott, J., & Sharwood Smith, M. (2011). Input, Intake, and Consciousness. The Quest for a Theoretical Foundation. *Studies in Second Language Acquisition*, 33, 497–528.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154.
- Ullman, M. T. (2005). A Cognitive Neuroscience Perspective on Second Language Acquisition: The Declarative/ Procedural Model. In C. Sanz (Ed.), *Mind and Context in Adult Second Language Acquisition. Methods, Theory, and Practice* (pp. 141–78). Washington, DC: Georgetown University Press.
- Uludag, O., & VanPatten, B. (2012). The Comparative Effects of Processing Instruction and Dictogloss on the Acquisition of the English Passive by Speakers of Turkish. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 50(3), 189–212.
- VanPatten, B. (1990). Attending to Form and Content in the Input. An Experiment in Consciousness. *Studies in Second Language Acquisition*, 12, 287–301.
- VanPatten, B. (1994). Evaluating the Role of Consciousness in Second Language Acquisition: Terms, linguistic features, and research methodology. In J. H. Hulstijn, & R. Schmidt (Eds.), *AILA Review 11: Consciousness in Second Language Learning*, 27–36.
- VanPatten, B. (1996). *Input Processing and Grammar Instruction in Second Language Acquisition* (2nd ed.). New Jersey: Ablex.
- VanPatten, B. (2002). Processing Instruction: An update. *Language Learning*, 52, 755–803.
- VanPatten, B. (2003). *From Input to Output. A Teacher's Guide to Second Language Acquisition. The McGraw-Hill Second Language Professional Series. Directions in Second Language Learning* (1st ed.). Boston: McGraw-Hill.
- VanPatten, B. (2004). *Processing instruction: Theory, research, and commentary*. Mahwah, NJ: Lawrence Erlbaum.
- VanPatten, B. (2005). Processing Instruction. In C. Sanz, *Mind and Context in Adult Second Language Acquisition. Methods, Theory, and Practice* (pp. 267–281). Washington, DC: Georgetown University Press.

- VanPatten, B. (2007). Input Processing in Adult Second Language Acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (pp.113–134). New York: Routledge.
- VanPatten, B., & Benati, A. G. (2010). *Key Terms in Second Language Acquisition*. London: Bloomsbury.
- VanPatten, B., & Borst, S. (2012). The Roles of Explicit Information and Grammatical Sensitivity in Processing Instruction: Nominative-Accusative Case Marking and Word Order in German L2. *Foreign Language Annals*, 45(1), 92–109.
- VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, 15, 225–243.
- VanPatten, B., & Fernández, C. (2004). The Long-Term Effects of Processing Instruction. In B. VanPatten (Ed.), *Processing Instruction. Theory, Research, and Commentary* (pp. 273–289). Mahwah, NJ: Lawrence Erlbaum Associates.
- VanPatten, B., & Oikkenon, S. (1996). Explanation versus Structured Input in Processing Instruction. *Studies in Second Language Acquisition*, 18(4), 495–510.
- VanPatten, B., & Sanz, C. (1995). From input to output: Processing instruction and communicative tasks. In F. R. Eckman, D. Highland, P. W. Lee, J. Mileham, & R. R. Weber (Eds.), *Second language acquisition theory and pedagogy* (pp. 169–185). Mahwah, NJ: Lawrence Erlbaum.
- VanPatten, B., & Uludag, O. (2011). Transfer of training and processing instruction: From input to output. *System*, 39, 44–53.
- VanPatten, B., & Wong, W. (2004). Processing instruction and the French causative: Another replication. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 97–118). Mahwah, NJ: Lawrence Erlbaum Associates.
- VanPatten, B., Collopy, E., Price, J. E., Borst, S., & Qualin, A. (2013). Explicit Information, Grammatical Sensitivity, and the First Noun Principle: A Cross-Linguistic Study in Processing Instruction. *The Modern Language Journal*, 97(2), 506–527.
- VanPatten, B., Farmer, J. L., & Clardy, C. L. (2009). Processing instruction and meaning-based output instruction: A response to Keating and Farley (2008). *Hispania*, 92, 116–126.
- VanPatten, B., Inclezan, D., Salazar, H., & Farley, A. P. (2009). Processing instruction and dictogloss: A study on object pronouns and word order in Spanish. *Foreign Language Annals*, 42, 557–575.
- Walter, D. R., & van Compernolle, R. A. (2015). Teaching German declension as meaning: a concept-based approach. *Innovation in Language Learning and Teaching*, DOI: 10.1080/17501229.2015.1041961, 1–18.
- Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, 49A, 51–75.
- Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*, 35, 550–564.

- Wesche, M. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. C. Diller (Ed.), *Individual Differences and Universals in Language Learning Aptitude* (pp. 119–154). Rowley, MA: Newbury House.
- White, J. P. (in press). Frequency effects of structured input: Primary and secondary effects. *Southern Journal on Linguistics*.
- White, J. P., DeMil, A., & Rice, M. H. (2015). Traditional Instruction and Processing Instruction: German Dative Definite Articles. *GFL German as a Foreign Language journal*, 1, 1–26.
- Whong, M. (2007). Seeking Consensus: Generative Linguistics and Language Teaching. *Leeds Working Papers in Linguistics and Phonetics*, 12, 143–155.
- Willems, K. (2011). The semantics of variable case marking (Accusative/Dative) after two-way prepositions in German locative constructions. Towards a constructionist approach. *Zeitschrift für Indogermanische Forschungen*, 116, 324–366.
- Winitz, H. (1981). The Comprehension Approach to Foreign Language Instruction. Rowley, MA: Newbury House.
- Wong, W. (2004). The Nature of Processing Instruction. In B. VanPatten (Ed.), *Processing Instruction. Theory, Research, and Commentary* (pp. 33–63). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wood Bowden, H., Sanz, C., & Stafford, C. A. (2005). Individual Differences: Age, Sex, Working Memory, and Prior Knowledge. In C. Sanz, *Mind and Context in Adult Second Language Acquisition. Methods, Theory, and Practice* (pp. 105–140). Washington, DC: Georgetown University Press.
- Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical Experiments in Neuropsychology*, 33 (1), 101–111.

APPENDIX A

Selective overview of Processing Instruction Studies

Study	Target feature	Focus
VanPatten & Cadierno (1993)	Spanish direct object pronouns	PI vs TI
Cadierno (1995)	Spanish preterite tense	PI vs TI
Cheng (1995)	Spanish copula <i>ser</i> and <i>estar</i>	PI vs TI
VanPatten & Sanz (1995)	Spanish direct object pronouns	PI sentence-level vs discourse-level
VanPatten & Oikkenon (1996)	Spanish direct object pronouns	PI vs SI vs EI
Benati (2001)	Italian future tense	PI vs TI
Farley (2001a, b)	Spanish subjunctive mood	PI vs MOI
Cheng (2002)	Spanish copula	PI vs TI
Cheng (2004)	Spanish copula	PI vs TI
Farley (2004)	Spanish subjunctive of doubt	PI vs MOI
Benati (2004a)	Italian future tense	SI vs EI
Benati (2004b)	Italian gender agreement on adjectives	SI vs EI
Sanz & Morgan-Short (2004)	Spanish word order	Explicit vs implicit feedback
Wong (2004b)	French negative & indefinite article	SI vs EI
VanPatten & Wong (2004)	French causative	PI vs TI
VanPatten & Fernández (2004)	Spanish direct object pronouns	Durative vs longitudinal
Benati (2005)	English simple past (L1 Greek + Chinese)	PI vs MOI
Gely (2005)	French imperfect tense	PI vs MOI vs CG
Morgan-Short & Bowden (2006)	Spanish direct object pronouns	PI vs MOI

Lee, Benati, Alguilar Sánchez & McNulty (2007)	Negative informal commands	Classroom vs computer vs hybrid
Lee & Benati (2007a)	Spanish preterite/ imperfect	Classroom vs computer vs hybrid
Lee & Benati, (2007d)	Italian subjunctive of doubt	PI vs MOI, classroom vs computer
Lee & Benati (2007e)	French subjunctive of doubt	PI vs MOI, classroom vs computer
Benati & Lee (2008)	Italian noun-adjective agreement and Italian future tense morphology	PI vs TI secondary effects
Benati, Lee & Houghton (2008)	English past tense and English third person singular -s	PI vs TI secondary effects
Benati, Lee & Laval (2008)	French imparfait, French subjunctive, and French causative with <i>faire</i>	PI vs TI secondary and cumulative effects
Laval (2008)	French imparfait, French Subjunctive, and French causative with <i>faire</i>	PI vs TI secondary and cumulative effects
Fernández (2008)	Spanish direct object pronouns	PI vs DG
VanPatten et al. (2009)	Spanish direct object pronouns	PI vs DG
Hikima (2010)	Japanese passive	PI vs re PI
Agiasophiti (2011)	German OVS sentences	PI vs IE
VanPatten & Uludag (2011)	English passive structures	PI vs CG
VanPatten & Borst (2012)	German nominative/ accusative distinction	+EI vs -EI
VanPatten et al. (2013)	German nominative/ accusative distinction, Spanish OVS sentences, Russian OVS sentences, French causative with <i>faire</i>	+EI vs -EI

APPENDIX B

Automated Digit Span Task Instructions

Introduction

The following procedure is used to assess DIGIT SPAN.

The DIGIT SPAN is the maximum number of digits that are correctly recalled after listening to a series of digits.

The DIGIT SPAN can be assessed for FORWARD recall (= order presented) and for BACKWARD recall (=reversed order).

EXAMPLE:

(*digits heard*) ONE SEVEN TWO FOUR NINE

FORWARD RECALL (= in order presented): 17249

BACKWARD RECALL (= reversed order): 94271

In order to familiarize you with this procedure, you will be given a short practice before starting the actual assessment.

Please, continue when you are ready.

Instructions for the Forward Span Practice

The following practice is to familiarize you with the FORWARD assessment.

- (1) You will first see a red circle in the middle of the screen.
- (2) Once the circle is erased, you will HEAR a sequence of digits from 1-9.
- (3) Another red circle signals the end of the digit sequence.
- (4) A textbox is presented in the middle of the screen:

Please type in the digit sequence you have heard in THE ORDER it was presented, e.g. if you heard (three, six, two) type 362 into the textbox

NO data will be recorded during the practice.

Please, continue when you are ready.

Instructions for the Backward Span Practice

The following practice is to familiarize you with the BACKWARD assessment:

- (1) You will first see a red circle in the middle of the screen.
- (2) Once the circle is erased, you will HEAR a sequence of digits from 1-9.

(3) Another red circle signals the end of the digit sequence.

(4) A textbox is presented in the middle of the screen:

Please type in the digit sequence you have heard in the REVERSED order it was presented, e.g. if you heard (three, six, two) type 263 into the textbox

NO data will be recorded during the practice.

Please, continue when you are ready.

Exit slide for Practice

This was a practice. The actual assessment is about to start.

Please, continue when you are ready.

Instructions for the Forward Span Trials

The FORWARD digit span assessment is about to begin.

For accurate assessment, please do NOT place your fingers on the keyboard before asked to do so.

There will be NO feedback during the assessment.

Please, continue when you are ready.

Exit slide for the Forward Span Trials

THANK YOU!

FORWARD assessment:

the maximum number of digits recalled correctly was: <% values.fML %>

the maximum number of digits recalled correctly before making two consecutive errors was: <% values.fTE_ML %>

The second measure is comparable to the traditional assessment of FORWARD digit span.

Instructions for the Backward Span trials

The BACKWARD digit span assessment is about to begin.

For accurate assessment, please do NOT place your fingers on the keyboard before asked to do so.

There will be NO feedback during the assessment.

Please, continue when you are ready.

Exit slide after the Backward Span trials

THANK YOU!

BACKWARD assessment:

the maximum number of digits recalled correctly was: <% values.bML %>

the maximum number of digits recalled correctly before making two consecutive errors was: <% values.bTE_ML %>

The second measure is comparable to the traditional assessment of BACKWARD digit span.

Exit slide for the experiment

THANK YOU for your participation!

FORWARD ASSESSMENT:

the maximum number of digits recalled correctly was: <% values.fML %>

the maximum number of digits recalled correctly before making two consecutive errors was: <% values.fTE_ML %>

BACKWARD ASSESSMENT

the maximum number of digits recalled correctly was: <% values.bML %>

the maximum number of digits recalled correctly before making two consecutive errors was: <% values.bTE_ML %>

APPENDIX C

Reading Span sentences to judge

	SENTENCE	7	Y	M	n
1	She had to cancel the appointment because she caught the flu yesterday.				
2	Catherine dressed up as a scary witch for the Halloween pencil on Friday.				Red
3	On warm sunny afternoons, I like to walk in the park.				
4	Paul likes to cry long distances in the park near his house.				
5	Jane forgot to bring her umbrella and got wet in the rain.				White
6	The firefighters sour the kitten that was trapped in the big oak tree.				
7	The class did not think the professor's lecture on history was very interesting.			White	Green
8	Doug helped his family dig in their backyard for their new swimming pool.			Green	White
9	Our dog Sammy likes to greet new people by joyful on them.			Green	Green
10	Jason's family likes to visit him in Atlanta during the cherry every year.		White		
11	I took my little purple to the ice cream store to get a cone.			Green	Green
12	The prom was only three days away, but neither girl had a dress yet.			Green	Green
13	Last year, Mike was given detention for running in the hall.	Red			
14	The dogs were very excited about going for a walk in the park.				
15	The boys knew they would have to hurry to make it to the apple on time.				
16	Most people agree that Monday is the worst stick of the week.				White
17	The gathering crowd turned to look when they heard the gun shot.				
18	Jim was so tired of studying, he could not read another page.				
19	Sara wanted her mother to read her a window before going to sleep.				
20	The only furniture Steve had in his first bowl was his waterbed.				
21	When it is cold, my mother always makes me wear a cap on my head.		White		Green
22	The doctor told my aunt that she would feel better after getting happy.		Red	Red	
23	All parents hope their list will grow up to be intelligent.			Green	Green
24	A person should never be discriminated against based on his race.			Green	Green
25	Because she gets to knife early, Amy usually gets a good parking spot.				
26	Women fall in jump with their infants at first sight or even sooner.	Yellow		Red	
27	Jason broke his arm when he fell from the tree onto the ground.			Green	Green
28	John wants to be a football player when he gets older.		White		Green
29	It was a clear night, and we could see the stars in the sky.			Green	Green
30	Raising children requires a lot of dust and the ability to be firm.				
31	Kristen dropped her parents off at the love for their annual vacation.		White		
32	We like to eat eggs and bacon for breakfast in the morning.			Green	Green
33	The huge clouds covered the morning slide and the rain began to fall.	Red		Green	Red
34	I told the class that they would get a surprise if they were orange.			Green	Green
35	Martha went to the concert, but ate to bring a thick sweater.			Green	Red
36	At the party, Randy got out the camera to take some pictures.			Green	White
37	My mother and father have always wanted to live near the cup.		White	Green	Green
38	When John and Amy moved to Canada, their wish had a huge garage sale.			Green	Green
39	He wrecked his car because he was going too fast in the rain.			Green	Green
40	After being ill, Suzy hoped to catch up on her work over the weekend.			Green	Green
41	Jill wanted a garden in her backyard, but the soil was mostly clay.			Green	Green
42	The judge gave the boy community sweat for stealing the candy bar.				Red
43	Peter and Jack ruined the family carwash when they burned the turkey.	Yellow	Green	Green	

44	Dan walked around the streets posting signs and looking for his lost puppy.				
45	Even though she was in trouble, she managed to go to the dice and shop.				Red
46	Sue opened her purse and found she did not have any money.				
47	Since it was the last game, it was hard to cope with the loss.			White	
48	Carol will ask her sneaker how much the flight to Mexico will cost.				
49	The children entered in a talent contest to win a trip to Disney World.				
50	Harry plans to play a lot of golf when he retires from his job.				
51	The sugar could not believe he was being offered such a great deal.				
52	As soon as I get done taking this envy I am going to go home.				Green
53	Wendy went to check her mail but all she received were cats .	Red		White	
54	Before Katie left for the city, she took a self-defense class at the gym.				
55	The printer sprinted when he tried to print out his report last night.	Red	White		
56	The college students went to New York in March and it snowed.				
57	Spring is her favorite time of year because flowers begin to bloom.				
58	With intense determination he overcame all obstacles and won the race.				
59	The tornado came out of nowhere and destroyed our raisin .				White
60	The girls were very excited about moving into their new house next week.				Yellow
61	His stereo was playing so loud that he blew out the speakers.				
62	At church yesterday morning, Jim's daughter made a terrible plum .				Red
63	Mary was asked to stop at the new mall to pick up several items.				
64	In the spring, the large birdfeeder outside my window attracts many birds.				
65	They were worried that all of their luggage would not fit in the car.	Red			
66	Realizing that she was late, Julia rushed to pick up her child from speaker .			White	Red
67	After one date I knew that Linda's sister simply was not my type.				
68	After yelling at the game, I knew I would have a tall voice.				
69	Joseph told his mother that he was probably going to fail sixth grade math.				
70	The sick boy had to stay home from school because he had a phone .				Red
71	In the fall, my gift and I love to work together in the yard.				
72	Nick's hockey team won their final game this past weekend at the shoes .				
73	Although Joe is sarcastic at times, he can also be very sweet.		White		
74	The lemonade players decided to play two out of three sets.		White		Red
75	When I get up in the morning, the first thing I do is feed my dog.				
76	Stacey stopped dating the light when she found out he had a wife.	White			
77	Unaware of the hunter, the deer wandered into his shotgun range.		Red		
78	Mary was excited about her new furniture that she had bought on sale.				
79	The couple decided that they wanted to have a picnic in the park.				
80	My mother has always told me that it is not polite to shine .				
81	The seventh graders had to build a volcano for their science class.				

Correctly judged	Green
Incorrectly judged	Red
Time-out error	Yellow
Not in the trial	

APPENDIX D

Automated Reading Span task instructions

Slide 1

In this experiment you will try to memorize letters you see on the screen while you also read sentences.

In the next few minutes, you will have some practice to get you familiar with how the experiment works.

We will begin by practicing the letter part of the experiment.

Click the left mouse button to begin.

Slide 2

For this practice set, letters will appear on the screen one at a time.

Try to remember each letter in the order presented.

After 2-3 letters have been shown, you will see a screen listing 12 possible letters.

Your job is to select each letter in the order presented.

To do this, use the mouse to select each letter.

The letters you select will appear at the bottom of the screen.

Click the mouse button to continue.

Slide 3

When you have selected all the letters, and they are in the correct order, hit the EXIT box at the bottom right of the screen.

If you make a mistake, hit the CLEAR box to start over.

If you forget one of the letters, click the BLANK box to mark the spot for the missing letter.

Remember, it is very important to get the letters in the same order as you see them.

If you forget one, use the BLANK box to mark the position.

Do you have any questions so far?

When you're ready, click the mouse button to start the letter practice.

Slide 4

Now you will practice doing the sentence reading part of the experiment.

A sentence will appear on the screen, like this:

"I like to run in the park."

As soon as you see the sentence, you should read it and determine if it makes sense or not. The above sentence makes sense.

An example of a sentence that does not make sense would be:

"I like to run in the sky."

When you have read the sentence and determined whether it makes sense or not, you will click the mouse button.

Click the mouse to continue.

Slide 5

You will then see "This sentence makes sense." displayed on the next screen, along with a box marked TRUE and a box marked FALSE.

If the sentence on the previous screen made sense, click on the TRUE box with the mouse.

If the sentence did not make sense, click on the FALSE box.

After you click on one of the boxes, the computer will tell you if you made the right choice.
Click the mouse to continue.

Slide 6

It is VERY important that you answer the sentence problems correctly.

It is also important that you try and read the sentences as quickly as you can.

Do you have any questions?

When you're ready, click the mouse to try some practice problems.

Slide 7

Now you will practice doing both parts of the experiment at the same time.

In the next practice set, you will be given one sentence to read.

Once you make your decision about the sentence, a letter will appear on the screen. Try and remember the letter.

In the previous section where you only read the sentences, the computer computed your average time to read the sentences.

If you take longer than your average time, the computer will automatically move you onto the next letter part, thus skipping the True or False part and will count that problem as a sentence error. Therefore it is VERY important to read the sentences as quickly and as accurately as possible.

Click the mouse to continue.

Slide 8

After the letter goes away, another sentence will appear, and then another letter.

At the end of each set of letters and sentences, a recall screen will appear.

Use the mouse to select the letters you just saw.

Try your best to get the letters in the correct order.

It is important to work QUICKLY and ACCURATELY on the sentences.

Make sure you know whether the sentence makes sense or not
before clicking to the next screen.

You will not be told if your answer regarding the sentence is correct.

After the recall screen, you will be given feedback about your performance
regarding both the number of letters recalled and the percent correct on
the sentence problems.

Do you have any questions?

Click the mouse to continue.

Slide 9

During the feedback, you will see a number in red in the top right of the screen.
This indicates your percent correct for the sentence problems
for the entire experiment.

It is VERY important for you to keep this at least at 85%.

For our purposes, we can only use data where the participant was at least
85% accurate on the sentences.

Therefore, in order for you to be asked to come back for future experiments,
you must perform at least at 85% on the sentence problems
WHILE doing your best to recall as many letters as possible.

Do you have any questions?

Click the mouse to try some practice problems.

Slide 10

That is the end of the practice.

The real trials will look like the practice trials you just completed.

First you will get a sentence to read, then a letter to remember.

When you see the recall screen, select the letters in the order presented.

If you forget a letter, click the BLANK box to mark where it should go.

Some of the sets will have more sentences and letters than others.

It is important that you do your best on both the sentence problems and
the letter recall parts of this experiment.

Remember for the sentences you must work as QUICKLY and ACCURATELY as
possible. Also, remember to keep your sentence accuracy at 85% or above.

Do you have any questions?

If not, click the mouse to begin the experiment.

Slide 11

Thank you for your participation.

APPENDIX E

Attitude and Motivation Test Battery (shortened and adapted for German)

Please indicate to which extent you agree with the following statements. Answer as quickly and honestly as you can. There are no wrong or right answers and your information remains confidential.

1 - I wish I could speak many foreign languages perfectly.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

2 - I would get nervous if I had to speak German to a tourist.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

3 - Studying foreign languages is annoying.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

4 - I make a point of trying to understand all the German I see and hear.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

5 - Studying German is important because I will need it for my career.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

6 - I feel self-conscious when I am speaking German in class.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

7 - I wish I could understand TV programmes in many foreign languages.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

8 - I feel confident when asked to speak in my German class.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

9 - I really enjoy learning German.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

10 - Studying German is not important for a good education.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

11 - I would rather spend my time on courses other than German.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

12 - It feels okay to speak German.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

13 - I put off my German coursework as much as possible.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

14 - I am calm whenever I have to speak German in class.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

15 - German is a very important part of the curriculum.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

16 - There is no need for us to learn foreign languages.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

17 - Studying German is futile because I will not need it in the future.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

18 - I give up and stop paying attention when I have trouble understanding my teacher's explanation.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

19 - I would feel uncomfortable speaking German anywhere outside the classroom.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

20 - I really work hard to learn German.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

21 - Studying German is important because it shows ambition.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

22 - I get nervous when I am speaking in my German course.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

23 - I think that learning German is dull.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

24 - I would feel comfortable speaking German in the presence of English and German speakers.

O=====O=====O=====O=====O=====O=====O
strongly moderately slightly neither agree slightly moderately strongly
disagree disagree disagree nor disagree agree agree agree

APPENDIX F

Personality Inventory

Please answer each question by ticking the “YES” or the “NO” box following the question. There are no right or wrong answers, and no trick questions. Work quickly and do not think too long about the exact meaning of the question. Please remember to answer each question.

1 - Are you usually carefree?

YES NO

2 - Do you stop and think things over before doing anything?

YES NO

3 - If you say you will do something do you always keep your promise, even if it's inconvenient?

YES NO

4 - Do you generally do and say things quickly without stopping to think?

YES NO

5 - Once in a while, do you lose your temper and get really angry?

YES NO

6 - Do you often do things on the spur of the moment?

YES NO

7 - Do you generally prefer surfing the web to meeting people?

YES NO

8 - Do you prefer to have few but special friends?

YES NO

9 - When people shout at you do you shout back?

YES NO

10 - Are all your habits good and desirable ones?

YES NO

11 - Are you mostly quiet when you are with other people?

YES NO

12 - Do you sometimes gossip?

YES NO

13 - If there is something you want to know about, would you rather look it up on the internet than talk to someone about it?

YES NO

14 - Do you like the kind of work that you need to pay close attention to?

YES NO

15 - Do you hate being with a crowd who play jokes on one another?

YES NO

16 - Do you like doing things in which you have to act quickly?

YES NO

17 - Do you like talking to people so much that you rarely miss a chance of talking to someone new?

YES NO

18 - Would you rather hang out with lots of people than spend time alone?

YES NO

19 - Would you say that you were fairly self-confident?

YES NO

20 - Do you sometimes talk about things you know nothing about?

YES NO

21 - Do you like playing pranks on others?

YES NO

Thank you :-)

APPENDIX G

Grammatical Sensitivity

The purpose of the following tests is to help me see how you work with language in general. All information remains confidential.

MATCHING WORDS

Circle the word in the second sentence that plays the same role in that sentence as the underlined word
does in the **bold sentence**.

Example: **Yesterday it rained quite heavily.**
We will go fishing at the lake tomorrow.

1. Yesterday, Mary caught a fish at the lake.

Cindy cut a cake with a knife.

2. Amy sang a pretty song to her class.

James throws big rocks into the lake.

3. Peter got an orange bike for his birthday.

My sister ate a big apple on Wednesday.

4. The furry dog barked at us as we walked by.

Did John go to the store to get bread?

5. She hopes to buy a car when she gets the money.

After you left most of the group remained in school.

6. My foot became sore from the infection.

The plant quickly grew strong in the warmth from the sun.

7. Money seems to make her happy.

Years ago most farming was done by hand.

8. Arguing never solved a problem.

Right now Kate is playing with her brother.

APPENDIX H

[Language Analytical Ability test for Pilot Study]

Now I'd like you to have a look at an invented language. It's called Pip. Read each of the examples carefully and you will be able to work out the meaning of each word. Don't rely on word order! (Note that there is no "the" or "a(n)" in Pip.)

- | | |
|----------------------|------------------------------|
| pit sak run | The dog chased the cat. |
| rin lup kat | The cat watched the mouse. |
| mup taw kid | The horse saw the teacher. |
| liip puut kat | The mice watched the dogs. |
| kid taw muuk | The horse saw the squirrels. |

pit: _____

sak: _____

run: _____

rin: _____

lup: _____

kat: _____

mup: _____

taw: _____

kid: _____

liip: _____

puut: _____

muuk: _____

Give the meaning of:

miip kat put _____

lip taw kud _____

Translate into Pip:

The mouse saw the cats. _____

[Language Analytical Ability test for Main Study]

Now I'd like you to have a look at an invented language. It's called Pip. Read each of the examples carefully and you will be able to work out the meaning of each word. Don't rely on word order! (Note that there is no "the" or "a(n)" in Pip.)

- | | |
|----------------------|------------------------------|
| pit sak run | The dog chased the cat. |
| rin lup kat | The cat watched the mouse. |
| mup taw kid | The horse saw the teacher. |
| liip puut kat | The mice watched the dogs. |
| kid taw muuk | The horse saw the squirrels. |

pit: _____

sak: _____

run: _____

rin: _____

lup: _____

kat: _____

mup: _____

taw: _____

kid: _____

liip: _____

puut: _____

muuk: _____

Give the meaning of:

kat put miip _____

taw kud lip _____

Translate into Pip:

The mouse saw the cats. _____

APPENDIX I

Input sentences and on-task performance during Processing Instruction

Sentence	74YuGCRD	YKX5Bwvj	nLQ6aVus	MgrYyVvU
Accusative recap				
Die Lehrerin sieht den Jungen.	1	1	1	1
Den Jungen mag das Maedchen.	1	1	1	1
Die Journalistin findet der Politiker.	0	1	1	1
Der Baecker ruft das Kind.	1	1	1	1
Das Maedchen besucht den Doktor.	1	1	1	1
Das Kind aergert der Vater.	1	1	1	1
Den Mann beraet die Frau.	1	1	1	1
Der Junge sucht die Mutter.	1	1	1	1
Der Trainer versteht den Spieler.	1	1	1	1
Den Grossvater besucht der Enkel.	1	1	1	1
RATING	2	2	2	2
Dative recap				
Das Maedchen dankt dem Vater.	1	1	1	1
Dem Jungen vertraut die Mutter.	1	1	1	1
Die Tochter schmeichelt dem Sohn.	1	1	1	1
Der Lehrerin widerspricht der Schueler.	1	1	0	1
Das Kind verzeiht dem Grossvater.	1	1	1	1
Der Tochter winkt die Frau.	1	1	1	1
Der Junge folgt der Mutter.	1	1	1	1
Dem Maedchen hilft der Junge.	1	1	1	1
Dem Kind glaubt die Lehrerin.	1	1	1	1
Das Maedchen gratuliert der Mutter.	1	1	1	1
RATING	2	2	2	2
Picture-matching task				
Auf dem Schrank springt die Katze.	0	1	1	1
Auf dem Sofa klettert das Kind.	1	1	1	0
In die Wohnung stuerzt der Mann.	1	1	1	1
In das Parkhaus faehrt das Auto.	1	1	0	0
In dem Zoo laufen die Leute.	1	1	0	0
In die Schule laeuft der Junge.	1	1	0	1
RATING	2	2	3	4
Sentence-level comprehension				
...in dem Park, wenn das Wetter gut ist.	1	1	1	0
... in das Uni Theater, wenn sie Zeit haben.	1	1	1	1
....in dem Museum, um sich zu bilden.	1	1	1	1
....in das 3D Kino, um den neusten Film...	1	1	1	1
....in den Park, um die Enten zu fuetttern.	1	1	1	1
...in dem Wald, weil es sehr warm ist.	1	1	1	1
....in die Bibliothek, um zu lernen.	1	1	1	1
....in der Stadt, um einzukaufen.	1	1	1	1
....in der Bibliothek auf Buechersuche.	1	1	1	1
RATING	2	2	2	2

Listening comprehension				
Auf das Konzert	1	1	0	1
In dem Café	1	1	0	1
in der Diskothek	1	0	0	0
in den Bahnhof	0	0	1	0
in dem Park	1	1	1	1
in den Supermarkt	1	1	1	1
in die Bar	1	1	1	1
in das Schwimmbad	1	0	0	1
In dem Kino	1	1	1	1
In der Bibliothek	1	1	0	0
RATING	2	2	3	3
Multiple choice				
Where do people drive cars?	0	1	1	1
Where do people save energy?	1	1	1	1
Where does public transport provide...	1	1	1	0
Where do people build their own houses?	1	1	1	1
Where does industry decrease?	1	0	1	1
Where do service providers relocate?	1	0	1	1
Where does the cultural life decline?	1	1	1	1
Where do people walk most of the time?	1	1	1	1
Where do people drink and drive more....	1	1	1	1
Where should elderly people move?	1	1	1	1
Is this picture of urban life accurate?	2	4	4	3
Would you consider yourself a city dweller?	2	4	5	5
RATING	3	2	2	2

Input sentences and on-task performance during Processing Instruction

Session 2

Sentence	74YuGCRD	YKX5Bwvj	nLQ6aVus	MgrYyVvU
Picture-matching task				
Auf den Schrank springt die Katze.	0	1	1	1
Auf das Sofa klettert das Kind.	1	1	1	1
In der Wohnung stuerzt der Mann.	1	1	1	1
In dem Parkhaus fahrt das Auto.	1	1	1	1
In den Zoo laufen die Leute.	1	1	1	0
In der Schule laeuft der Junge.	1	1	1	1
Auf dem Schrank springt die Katze.	1	1	1	1
Auf dem Sofa klettert das Kind.	1	1	1	1
In die Wohnung stuerzt der Mann.	1	1	1	1
In das Parkhaus fahrt das Auto.	1	1	1	1
In dem Zoo laufen die Leute.	1	1	1	1
In die Schule laeuft der Junge.	1	1	1	1
RATING	2	2	2	2
Sentence-level comprehension				
...in dem Park.	1	1	1	1
...in das Uni Theater.	1	1	1	1
....in dem Museum.	1	1	1	1
...in das 3D-Kino.	1	1	1	1
...in den Park.	1	0	1	1
....in dem Wald.	1	1	1	1
....in die Bibliothek.	1	1	1	1
....in die Stadt.	1	1	1	1
....in der Bibliothek.	1	1	1	1
RATING	2	2	2	2
Location versus direction				
Wohin stecke ich den Stadtplan?	0	1	0	1
Wo steht die Vase?	1	1	1	1
Wo haengt die Mona Lisa?	1	1	1	1
Wohin stellt er das Auto?	0	1	1	1
Wo sitzt die Katze?	1	1	1	1
Wo liegen die ostfriesischen Inseln?	1	1	1	1
Wohin haengst du das Poster?	0	1	1	1
Wo sitzt Benedetta?	1	1	1	1
Wohin setzt sich Martha?	1	0	0	1
Wohin stelle ich meine Tasche?	0	1	1	1
Wo steht deine neue Lampe?	1	1	1	1
Wo steckt Manuel?	1	1	1	1
Wo liegt mein Handy?	1	1	1	1
Wohin haenge ich das Bild?	0	1	0	1
Wohin lege ich meine Schluessel?	1	1	1	1
RATING	3	2	2	2

Sentence puzzle				
an der Bahnstation	1	1	0	0
in den Regalen	1	1	1	1
auf dem Sofa	1	1	1	1
unter dem Tisch	1	1	1	1
in den Park	0	0	1	0
an der Wand	1	1	0	0
auf der Zeitschrift	0	0	1	1
unter den Büchern	0	1	1	1
neben dem Tisch	1	1	1	1
über das Bett	1	1	1	1
vor dem Haus	1	1	1	1
zwischen die Regale	1	1	1	1
hinter die Kulissen	0	1	1	1
neben die Haustür	1	1	1	1
über dem Bett	1	1	1	1
vor den Schrank	1	0	0	0
zwischen den Büchern	1	1	1	1
hinter dem Sofa	1	1	1	1
RATING	3	2	2	3

APPENDIX J

Pilot Study: Interpretation task

Please read the following sentences carefully and indicate which English translation is correct. Choose “not sure” if you are uncertain.

Der Sohn dankt dem Vater. a) The son thanks the father. b) The father thanks the son. c) Not sure.	Dem Lehrer vertraut das Mädchen. a) The teacher trusts the girl. b) The girl trusts the teacher. c) Not sure.
Die Frau sieht den Hund. a) The woman sees the dog. b) The dog sees the woman. c) Not sure.	Der Mutter widerspricht der Vater. a) The mother contradicts the father. b) The father contradicts the mother. c) Not sure.
Das Mädchen verzeiht dem Jungen. a) The girl forgives the boy. b) The boy forgives the girl. c) Not sure.	Den Mann erkennt die Frau. a) The man recognises the woman. b) The woman recognises the man. c) Not sure.
Der Mann folgt der Frau. a) The man follows the woman. b) The woman follows the man. c) Not sure.	Die Mutter findet der Sohn. a) The mother finds the son. b) The son finds the mother. c) Not sure.
Dem Mädchen glaubt die Mutter. a) The girl believes the mother. b) The mother believes the girl. c) Not sure.	Das Kind gratuliert der Mutter. a) The child congratulates the mother. b) The mother congratulates the child. c) Not sure.
Das Kind ruft den Lehrer. a) The child calls the teacher. b) The teacher calls the child. c) Not sure.	Das Kind fragt der Vater. a) The child asks the father. b) The father asks the child. c) Not sure.
Die Frau schmeichelt dem Mann. a) The woman flatters the man. b) The man flatters the woman. c) Not sure.	Der Mutter winkt die Tochter. a) The mother waves to the daughter. b) The daughter waves to the mother. c) Not sure.
Der Junge hört die Mutter. a) The boy hears the mother. b) The mother hears the boy. c) Not sure.	Dem Kind hilft der Mann. a) The child helps the man. b) The man helps the child. c) Not sure.

Please indicate how difficult you found this exercise.

Too easy	Easy enough	Moderate	Challenging	Too difficult
<input type="checkbox"/>				

Main Study: Translation task with model answers (pre-test)

Translation A
<i>Der Sohn hat Hausarrest, weil er das Fenster zerschossen hat.</i> The son is grounded because he shot the window to pieces.
<i>Die Frau sieht der Hund mit den langen, braunen Schlappohren.</i> The dog with the long, brown, floppy ears sees the woman.
<i>Das Mädchen verzeiht dem Jungen den vergessenen Geburtstag.</i> The girl forgives the boy for the forgotten birthday.
<i>Der Mann folgt der Frau schon seit Stunden durch die Stadt.</i> The man has been following the woman through the city for hours already.
<i>Das Kind ruft der Lehrer nach dem Mathematikunterricht.</i> The teacher calls the child after the maths lesson.
<i>Die Mutter widerspricht dem Vater oft in politischen Fragen.</i> The mother often contradicts the father when it comes to political matters.
<i>Dem Lehrer vertraut das Mädchen schon aus Prinzip nicht.</i> The girl doesn't trust the teacher out of principle.
<i>Der Mann erkennt die Frau ohne seine Brille von Weitem nicht.</i> Without his glasses, the man does not recognise the woman from afar.
<i>Die Mutter findet der Sohn endlich im chaotischen Supermarkt.</i> The son finally finds the mother in the chaotic supermarket.
<i>Das Kind gratuliert der Mutter zum Geburtstag einen Tag verspätet.</i> The child wished the mother a happy birthday one day late.
<i>Das Kind fragt der Vater um Hilfe bei der Fahrradreparatur.</i> The father asks the child for help with the bicycle repair.
<i>Der Mutter winkt die Tochter von der anderen Straßenseite aus.</i> The daughter waves at the mother from the other side of the street.
<i>Dem Kind hilft der Mann mit den grünen Hosenträgern.</i> The man with the green braces helps the child.

Please indicate how difficult you found this exercise.				
Too easy	Easy enough	Moderate	Challenging	Too difficult
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Main Study: Translation task with model answers (post-tests)

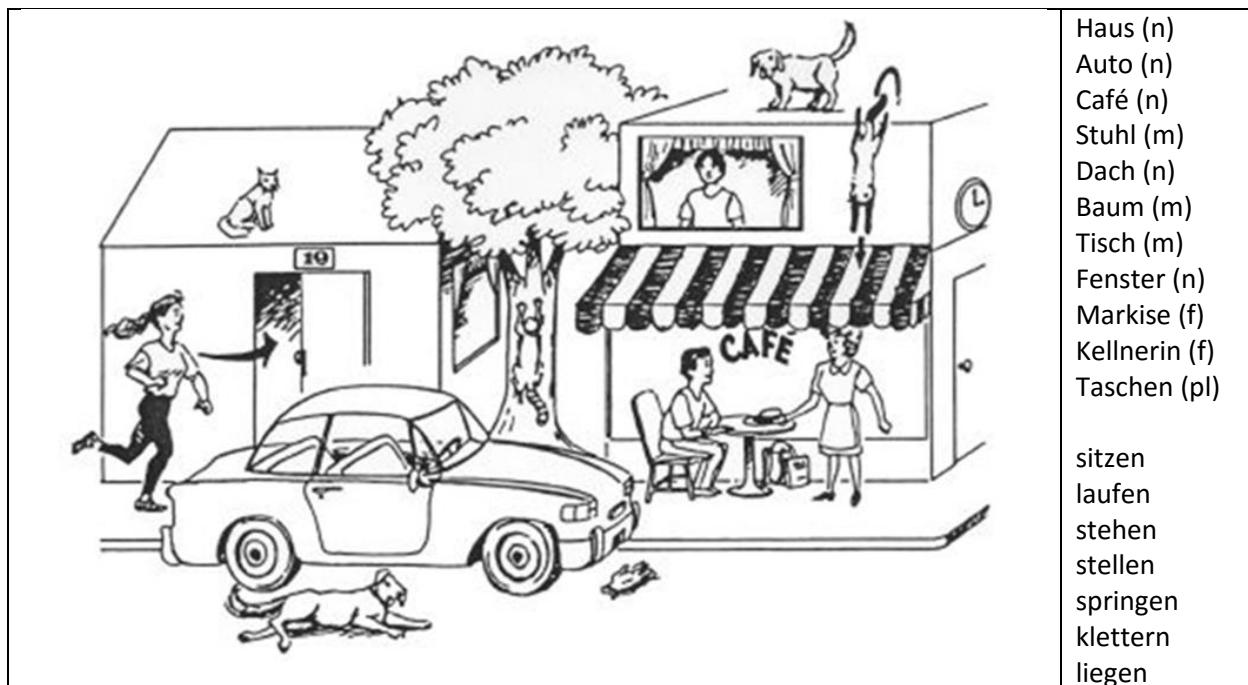
Translation B				
<i>Die Tochter findet der Vater nach dem Fußballspiel im Stadion nicht wieder.</i>				
The father doesn't relocate the daughter in the stadium after the football match.				
<i>Dem Jungen vertraut das Mädchen die Weihnachtsdekoration an.</i>				
The girl entrusts the boy with the Christmas decorations.				
<i>Das Mädchen verzeiht dem Vater den peinlichen Aufzug so schnell nicht.</i>				
The girl doesn't forgive the father for the embarrassing get-up quite so quickly.				
<i>Der Sohn folgt dem Hund ohne Leine durch den dunklen Wald.</i>				
The son follows the dog without a lead through the dark forest.				
<i>Das Mädchen ruft der Rektor noch vor dem Unterrichtsbeginn.</i>				
The headmaster calls the girl before the lesson even starts.				
<i>Die Frau widerspricht dem Mann in allen sozialen Angelegenheiten.</i>				
The woman contradicts the man on all social matters.				
<i>Die Frau sieht der Junge mit der neonblauen Pudelmütze.</i>				
The boy with the neon-blue bobble hat sees the woman.				
<i>Der Sohn erkennt die Mutter aufgrund ihrer neuen Frisur zunächst nicht.</i>				
At first, the son doesn't recognise the mother because of her new haircut.				
<i>Der Vater hat Küchendienst, weil er den Hochzeitstag vergessen hat.</i>				
The father is on kitchen duty because he forgot the wedding anniversary.				
<i>Der Sohn gratuliert der Freundin einen Tag zu früh zum Universitäts-abschluss.</i>				
The son congratulates the girlfriend on the university degree one day too early.				
<i>Das Mädchen fragt der Vater um Hilfe beim Gardinen aufhängen.</i>				
The father asks the girl for help with hanging the curtains.				
<i>Dem Mädchen hilft der Junge bei der Integralrechnung.</i>				
The boy helps the girl with the integral calculus.				
<i>Der Tochter sieht die Mutter von der Empore aus zu.</i>				
The mother watches the daughter from the gallery.				

Please indicate how difficult you found this exercise.

Too easy <input type="checkbox"/>	Easy enough <input type="checkbox"/>	Moderate <input type="checkbox"/>	Challenging <input type="checkbox"/>	Too difficult <input type="checkbox"/>
--------------------------------------	---	--------------------------------------	---	---

Picture description in both experiments and in all testing sessions

Have a look at the picture below and write at least 5 sentences about what you see.



an – in – auf – über – unter – hinter – vor – zwischen – neben

Eine Frau _____

Ein Hund _____

Eine Katze _____

Ein Mann _____

Please indicate how difficult you found this exercise.

Too easy	Easy enough	Moderate	Challenging	Too difficult
<input type="checkbox"/>				

Main Study: Essay task (pre-test)

Discourse-level production

Describe your ideal holiday: where would you like to go and why? What is the location of the place and how would you get there? What is so special about it? What can you do there?

Write at least 250 words of German.

Please indicate how difficult you found this exercise.				
Too easy	Easy enough	Moderate	Challenging	Too difficult
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Main Study: Essay task (post-tests)

Discourse-level production 2

Describe your ideal university experience: where would you like to study and why? What is the location of the place and how would you get there? What is so special about it? What can you do there?

Write at least 250 words of German.

Please indicate how difficult you found this exercise.

Too easy	Easy enough	Moderate	Challenging	Too difficult
<input type="checkbox"/>				

Fill-in-the-blanks task in both experiments (pre-test)

Sentence-level written production A

Please fill in the missing endings. The colours indicate gender and number:

masculine, singular

feminine, singular

neuter, singular

plural

1. Bei schlechtem Wetter verbringen wir viel Zeit in d__ Geschäften.
2. Vor d__ Schule steht ein alter Kastanienbaum.
3. Am Wochenende gehen wir am Maschsee auf d__ Rummel.
4. Nach der Schule gehen wir oft in d__ Einkaufszentrum.
5. Während der Sanierung fahren Züge nicht über d__ Altstadtbrücke.
6. Auf d__ Rathausplatz werden Marktstände aufgebaut.
7. An d__ Theater stehen Leute stundenlang für Tickets an.
8. Die Marathonzuschauer werfen ihren Müll unter d__ Absperrung.
9. Wir schreiben einen Beschwerdebrief an d__ Bürgermeister.
10. Meine Tante wohnt direkt über d__ Café Härtel.
11. Der neue Bahnhof kommt direkt neben d__ Parkauen.
12. Ich habe mein Fahrrad zwischen d__ Container und d__ Lampe gestellt.
13. Das Schwimmbad liegt hinter d__ Ausstellungshallen.
14. Zum Telefonieren muss man vor d__ Schultor gehen.
15. Unter d__ Stadtkirche wurde kürzlich eine Bombe entdeckt.
16. Wir machen eine Theaterführung mit, um hinter d__ Vorhang zu schauen.
17. Das neue Restaurant liegt zwischen d__ Museen und d__ Theatern.
18. Das kleine Café neben d__ Kino ist das beste Café der Stadt.

Please indicate how difficult you found this exercise.				
Too easy	Easy enough	Moderate	Challenging	Too difficult
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fill-in-the-blanks task in both experiments (post-tests)

Sentence-level written production B

Please fill in the missing endings. The colours indicate gender and number:

masculine, singular

feminine, singular

neuter, singular

plural

1. Bei gutem Wetter verbringen wir viel Zeit in d__ Parks.
2. Das kleine Kino neben d__ **Ausstellungshalle** ist besonders gut.
3. Ich habe das neue Bild zwischen d__ Regal und d__ **Schrank** gehängt.
4. Wir stellen unsere Fahrräder unter d__ Kastanienbäume.
5. Zum Rauchen müssen die Gäste vor d__ **Tür** gehen.
6. Am Wochenende gehen wir vielleicht auf d__ Maschseefest.
7. Vor der Schule gehen wir oft noch in d__ **Supermarkt**.
8. Auf d__ **Stadtkirche** brüten dieses Jahr wieder Störche.
9. Wir machen einen Stadtrundgang, um hinter d__ Fassaden zu schauen.
10. An d__ Kino stehen Leute stundenlang für Tickets an.
11. Die Konzertzuschauer werfen ihren Müll neben d__ **Container**.
12. Wir schreiben einen Dankesbrief an d__ **Bürgermeisterin**.
13. Mein Bruder wohnt direkt hinter d__ **Bahnhof**.
14. Die Ausstellung ist direkt über d__ Verkaufsräumen.
15. Während der Veranstaltung dürfen Flugzeuge nicht über d__ Gelände fliegen.
16. Das neue Café liegt zwischen d__ **Schule** und d__ Bushaltestellen.
17. Unter d__ **Parkplatz** wurde kürzlich ein Skelett gefunden.
18. Hinter d__ **Theater** parken am Wochenende viele Autos.

Please indicate how difficult you found this exercise.

Too easy	Easy enough	Moderate	Challenging	Too difficult
<input type="checkbox"/>				

APPENDIX K

Material Testing Session (Template)

Material Testing Session I (March 5th 2014)

Questionnaire	Time needed to fill in the questionnaire
Background Questionnaire	
EPI	
Attitudes and Interest	
LAA/Grammatical Sensitivity	
TOTAL	

Comments, questions and suggestions

Background Questionnaire

Attitudes and Interest

LAA/Grammatical Sensitivity

General Feedback

APPENDIX L

Pilot Study: On-task performance, Session 1

Task	A	B
Accusative recap		
Die Lehrerin sieht den Jungen.	1	1
Den Jungen mag das Maedchen.	0	1
Die Journalistin findet der Politiker.	0	0
Der Baecker ruft das Kind.	0	0
Das Maedchen besucht den Doktor.	1	1
Das Kind aergert der Vater.	1	0
Den Mann beraet die Frau.	1	1
Der Junge sucht die Mutter.	1	0
Der Trainer versteht den Spieler.	1	1
Den Grossvater besucht der Enkel.	1	1
Score	7 (of 10)	6 (of 10)
Difficulty rating	3	4
Dative recap		
Das Maedchen dankt dem Vater.	1	1
Dem Jungen vertraut die Mutter.	0	1
Die Tochter schmeichelt dem Sohn.	1	1
Der Lehrerin widerspricht der Schueler.	0	1
Das Kind verzeiht dem Grossvater.	1	1
Der Tochter winkt die Frau.	1	1
Der Junge folgt der Mutter.	1	1
Dem Maedchen hilft der Junge.	1	1
Dem Kind glaubt die Lehrerin.	1	1
Das Maedchen gratuliert der Mutter.	1	1
	8 (of 10)	10 (of 10)
Difficulty rating	4	3
Picture-matching task		
Auf dem Schrank springt die Katze.	0	1
Auf dem Sofa klettert das Kind.	1	1
In die Wohnung stuerzt der Mann.	1	1
In das Parkhaus faehrt das Auto.	0	1
In dem Zoo laufen die Leute.	1	0
In die Schule laeuft der Junge.	0	0
	3 (of 6)	4 (of 6)
Difficulty rating	5	4
Sentence-level comprehension		
in dem Park	1	0
in das Uni-Theater	1	0
in dem Museum	0	0
in das 3D-Kino	0	1
in den Park	0	0
in dem Wald	0	0

in die Bibliothek	1	0
in der Stadt	0	0
in der Bibliothek	1	1
	4 (of 9)	2 (of 9)
Difficulty rating	4	4
Listening comprehension		
auf das Konzert	1	0
in dem Café	1	0
in der Diskothek	1	0
in den Bahnhof	1	1
in dem Park	1	0
in den Supermarkt	1	0
in die Bar	1	0
in das Schwimmbad	1	1
in dem Kino	1	0
in der Bibliothek	1	0
	10 (of 10)	2 (of 10)
Difficulty rating	2	4
Discourse-level interpretation		
Where do people drive cars?	1	1
Where do people save energy? (distractor)	1	0
Where does public transport provide travel links?	1	1
Where do people build their own houses? (distractor)	1	1
Where does industry decrease? (distractor)	0	0
Where do service providers relocate?	0	1
Where does the cultural life decline? (distractor)	0	0
Where do people walk most of the time?	1	1
Where do people drink and drive more frequently? (distractor)	0	1
Where should elderly people move?	1	1
Do you think this image of urban life is accurate?	4	1
Would you say that you are a typical city dweller?	3	5
	3 (of 4)	4 (of 4)
Difficulty rating	3	4
TOTAL SCORE	35/49	28/49

Pilot Study: On-task performance, Session 2

Task	A	B
Picture-matching task		
Auf den Schrank springt die Katze.	1	0
Auf das Sofa klettert das Kind.	1	0
In der Wohnung stuerzt der Mann.	0	1
In dem Parkhaus faehrt das Auto.	1	0
In den Zoo laufen die Leute.	0	1
In der Schule laeuft der Junge.	1	0
Auf dem Schrank springt die Katze.	1	1
Auf dem Sofa klettert das Kind.	1	1
In die Wohnung stuerzt der Mann.	1	1
In das Parkhaus faehrt das Auto.	1	1
In dem Zoo laufen die Leute.	1	0
In die Schule laeuft der Junge.	0	1
	9 (of 12)	7 (of 12)
Difficulty rating	3	3
Sentence-level comprehension		
in dem Park	1	1
in das Uni-Theater	1	1
in dem Museum	1	1
in das 3D-Kino	1	1
in den Park	1	1
in dem Wald	1	1
in die Bibliothek	1	1
in der Stadt	1	1
in der Bibliothek	1	1
	9 (of 9)	9 (of 9)
Difficulty rating	2	3
Location versus direction		
in den Rucksack	1	1
auf dem Tisch	1	1
in dem Louvre	1	1
vor das Kino	1	1
auf dem Dach	0	1
in der Nordsee	1	1
in die Kueche	1	1
unter dem Apfelbaum	1	1
neben den Lehrer	1	1
neben das Bett	1	1
neben dem Bett	1	1
in der U-Bahn	1	1
Auf dem Sofa	1	1
an die Wand	1	1
auf das Tischchen	1	1
	14 (of 15)	15 (of 15)

Difficulty rating	2	3
Sentence puzzle		
stehen viele Taxis	0	0
stehen die Buecher	0	0
liegt unsere Katze	1	1
liegt der Hund	1	1
stellen sie Muelleimer	0	0
haengt das Bild	1	1
steht die Tasse	1	1
liegt das Handy	1	0
steht der Stuhl	1	1
legt sie die Decke	0	0
steht das Auto	0	1
stellt er den Schrank	0	1
gehen wir	1	0
stellt er den Muell	1	0
haengt das Poster	0	1
stellt sich Eva	0	1
steckt das Ticket	0	0
liegt die Fernbedienung	1	1
	9 (of 18)	10 (of 18)
Difficulty rating	5	4
TOTAL SCORE	41 (of 54)	41 (of 54)

APPENDIX M

Exit questionnaire

1) Did you receive any formal instruction on German switch prepositions in the time between the experiment and now?

YES NO

1a) If you answered YES in 1), what kind and how many hours of instruction did you receive?

2) Did you revise German switch prepositions in the time between the experiment and now?

YES NO

2a) If you answered YES in 2), what kind of revision did you do and how many hours did you approximately use for that?

3) How would you summarise **in one sentence** the message that you have taken away from the computer-delivered instruction you received during the experiment?

4) Did you have more or less exposure to German during the term (compared to the term break)?

MORE LESS ABOUT THE SAME

APPENDIX N

Flyer and participant information



SPEAK GERMAN? WANT TO EARN SOME MONEY?

Research participants needed

Participate in a research study on German grammar acquisition (approx. 6 hours stretched over 2 weeks) and earn £120!



intermediate to advanced level

receive free instruction

discover what type of learner you are

learn new strategies

get rewarded for your time

STEPHANIE PETER

University of Greenwich

Old Royal Naval College
30 Park Row, London
SE10 9LS

Ps31@gsu.ac.uk



INFORMATION SHEET FOR PARTICIPANTS

You are being invited to be involved in this research study. Before you decide whether you want to take part, it is important for you to understand why the research is being done and what your participation will involve. Please take time to read the following information carefully. Please contact me if anything is unclear or if you would like more information. Take time to decide whether or not you wish to take part.
Thank you for reading this.

What is the purpose of the study?

The purpose of the study is to try and understand whether a certain method of grammar instruction, i.e. Processing Instruction, is a more effective approach to grammar teaching for some learners than others. If we understand who benefits most from this kind of instruction, we can try to take better care of the individual needs of learners when teaching grammar, thus making grammar instruction a better experience for all involved.

Why have I been chosen?

You are being invited to take part in this study, as you are studying German at a UK Higher Education institution.

What will participation involve?

Participation will involve filling in a questionnaire on your personal background, completing a grammaticality judgement test and a short test on the grammatical form you will be taught, completing a few working memory tasks, receiving approximately 2 hours of grammar teaching and completing two short tests, one straight after the instructional period and one two weeks later.

All of the research data will be stored as hard copy at the University of Greenwich for 5 years.

Please note that:

- You can decide to withdraw from the study at any point
- Your name will be removed from the information and anonymised. It should not be possible to identify anyone from my reports on this study.

It is up to you to decide whether to take part or not. If you decide to take part you are still free to withdraw at any time, without giving reason. If you withdraw from the study all data will be withdrawn and destroyed. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form.

Contact for further information

Stephanie Peter, ps31@gre.ac.uk

Thank you

PARTICIPANT CONSENT FORM

To be completed by the participant. If the participant is under 18, to be completed by the parent / guardian / person acting *in loco parentis*.

<ul style="list-style-type: none">• I have read the information sheet about this study• I have had an opportunity to ask questions and discuss this study• I have received satisfactory answers to all my questions• I have received enough information about this study• I understand that I am / the participant is free to withdraw from this study:<ul style="list-style-type: none">○ At any time (until such date as this will no longer be possible, which I have been told)○ Without giving a reason for withdrawing○ (If I am / the participant is, or intends to become, a student at the University of Greenwich) without affecting my / the participant's future with the University○ Without affecting any medical or nursing care I / the participant may be receiving.• I understand that my research data may be used for a further project in anonymous form, but I am able to opt out of this if I so wish, by ticking here. <input type="checkbox"/>• I agree to take part in this study	
Signed (participant)	Date
Name in block letters	
Signature of researcher	Date
This project is supervised by: Prof Alessandro Benati, A.Benati@greenwich.ac.uk	
Researcher's contact details Stephanie Peter, ps31@gre.ac.uk 07450 191791 Old Royal Naval College 30 Park Row London SE10 9LS	

APPENDIX O

University Research Ethics Committee Application Form

SECTION 4: ETHICAL CONSIDERATIONS

4.1 What do you consider are the main ethical issues and risks that may arise in this research? (Refer to the Guidance on Ethical Approval for Research, [point 3.1](#)). What steps will be taken to address each issue?

Personal data will be collected. However, it will be treated with utmost consideration and according to the *Principles of Data Protection* of the University of Greenwich, see below. In addition to that, the aims of the study will have to be explained very carefully at the debriefing stage as the concept of working memory capacity might be associated with a more general notion, such as intelligence or aptitude – which might bias the participants' attitude and openness towards the study.

Steps will also be taken to ensure that participants are aware that participation will not have any adverse impact upon their academic performance. In those instances where the research will take place on the premises of other HE institutions, all steps will be taken to ensure that reasonable precautions are taken to protect the safety of both researcher and participant. In particular, the researcher's supervisor will be kept informed of the dates and locations of the fieldwork.

4.2 Will personal data be collected during the research (Refer to the Guidance on Ethical Approval for Research, [point 3.2](#))? Indicate YES or NO. If YES, give details of how you will deal with that data.

YES

In the course of this study on individual differences, personal data will be collected and processed in accordance with the *Principles of Data Protection* of the University of Greenwich. The information supplied by research participants will be treated as strictly confidential by the researcher and will only be used in the context of the present study. Any information published or disseminated in this context will be treated in a way that guarantees the anonymity of the participants, e.g. personal data will be separated from the identifying details of the participant and the data collected will be coded before it is processed in order to enable correlation while maintaining the participants' anonymity.

The data will be protected against unauthorised access: it will be held in a secure location, hard copies will be stored in locked cabinets and electronic copies will be password-protected. Data will be transferred in a secure manner. Movable storage media will be protected against loss, damage, or destruction of the data. Collected data will be retained for five years before it will be disposed of securely.

Only the research student and their supervisor will have access to the data in question.

APPENDIX P

Overview of the research process

Date	Event	Subsequent changes
06/ 13	1 st Postgraduate Conference <i>Presentation</i>	Choice of one Working Memory Model and a working definition for the project
03/ 14	Postgraduate Reading Group <i>Material Testing Session</i>	Improvement of questionnaires (abridged, rephrased)
04/ 14	2 nd Postgraduate Conference <i>Presentation</i>	Refinement of the preliminary experimental design
07/ 14	BAAL SIG Leeds <i>Presentation</i>	Potential change of targeted research population considered (now: secondary school students)
09/ 14	Upgrade from MPhil to PhD <i>Viva voce</i>	Changes in the presentation of the dissertation
11/ 14	L-SLARF PhD Conference <i>Presentation</i>	Improved theoretical framework for Individual Differences section (inclusion of Skehan's understanding of WM and IDs)
02/ 15	Secondary School Visits <i>Observations, meetings</i>	Identification of the target feature: German switch prepositions
04/ 15	Pilot Study	Improvement of testing materials (e.g. word order in LAA test) and development of a discourse-level task
06/ 15	TDL Groningen Conference <i>Presentation</i>	Improved justification of Working Memory Capacity measures and new perspective (cognitive grammar)
07/ 15	ATINER Athens Conference <i>Presentation</i>	Potential addition of new layers of data, e.g. collected through eye-tracking
09/ 15	Postgraduate Reading Group <i>Presentation</i>	Rejection of eye-tracking approach and re-focus on case study with different sample population (university students)
09/15	Material Testing Session 2	Development of new interpretation and production tasks based on tests with advanced learners and native speakers
10/ 15	Main Study <i>Participant Screening</i>	Change of sentence-level interpretation tests to account for higher proficiency level
11/ 15	Main Study <i>Data analyses</i>	Consideration of new Processing Instruction materials that do not include certain locative prepositional verbs
03/ 16	Triangulation <i>Essay scoring with German lecturer</i>	Further changes to discourse-level production task (e.g. higher complexity and different modality)
2016	Future research	TBC