



UNIVERSITAT DE
BARCELONA

Comparing Focus on Forms and Task-Based Language Teaching in the Acquisition of Russian as a Foreign Language

Elena Markina

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Ciència Cognitiva i Llenguatge

**Comparing Focus on Forms and Task-Based Language Teaching
in the Acquisition of Russian as a Foreign Language**

TESI DOCTORAL

Elena Markina

Director i tutor: Dr. Joan Castellví Vives

Departament de Filologia Catalana i Lingüística General

Universitat de Barcelona

Barcelona, Septembre 2018

ABSTRACT

The present study explores the impact that task-supported language teaching (TSLT) and task-based language teaching (TBLT) have on the acquisition of case forms and verbs of motion in Russian and on learners' written and oral production. From a methodological point of view, the study follows a pre-test – immediate post-test – delayed post-test design. Data have been obtained from first and second year students of the University of Barcelona (n=54) with a low level of proficiency in Russian. Each learner was assigned to a task-supported (TS) or task-based (TB) group and received 14 hours of corresponding treatment. The TSLT treatment for the first year students focused on the use of prepositions and case forms, and the treatment for the second year students involved the Russian verbs of motion with and without prefixes. For the TBLT treatment, the acquisition of the same linguistic items was analysed. These items were present in the tasks designed for the experiment. Data were obtained by means of three grammar tests (fill in the blanks, multiple choice and grammaticality judgment tests), a written task which required learners to write a letter / an email to a friend and two oral tasks (room description and map task). Measures include the target-like use of prepositions, case forms and verbs of motion in the grammar tests and in oral and written production, the number of errors per words for general accuracy, the number of clauses per T-unit and the mean length of clause for syntactic complexity, Guiraud's Index of lexical richness for lexical complexity, and speech rate for oral fluency. Statistical tests related to the use of prepositions and case forms show that both approaches provided positive results in the immediate post-test, however, neither task-supported nor task-based treatment led to a significant improvement in the long-term perspective. On the other hand, participants in both groups demonstrated a significant improvement in the target-like use of verbs of motion produced in their oral and written performance and used in their grammar tests. In written production, learners' syntactic complexity measured by means of clauses per T-unit and lexical complexity significantly increased after the treatment, whereas phrasal complexity (mean length of clause) did not change over time. Written accuracy significantly improved in both groups immediately after the treatment. However, learners in the TB group maintained this improvement three months after the treatment, whereas the accuracy of learners in the TS group decreased to the level they had before the treatment. In oral production, learners in both groups significantly improved their general accuracy in the performance of the map task. Lexical complexity and fluency improved in both oral tasks. As for syntactic complexity, no changes were found in any of the tasks. The comparison of the effects that the two types of treatment have had on different aspects of learners' production reveals that participants in the TS and TB groups have not significantly differed on accuracy, syntactic complexity and fluency of their oral and written performance. Task-supported and task-based treatments also had a similar effect on accuracy in the use of Russian prepositions and case forms. However, learners in the TB group have showed significantly better results than learners in the TS group as far as lexical complexity and the target-like use of verbs of motion are concerned.

RESUMEN

El presente estudio explora el impacto en la adquisición del ruso como lengua extranjera de dos propuestas metodológicas diferentes, las llamadas TSLT y TBLT, donde TSLT (Task-supported language teaching) se refiere a la instrucción con énfasis en las formas en la que las tareas se utilizan como soporte y TBLT (Task-based language teaching) se refiere a la enseñanza basada en tareas. El objeto de análisis es la adquisición de los casos y de los verbos de movimiento, tanto en la producción oral como la escrita, en los parámetros de la complejidad, corrección y fluidez. Desde el punto de vista metodológico, el estudio experimental sigue el diseño de pre-test – post-test inmediato – post-test diferido. Los datos han sido obtenidos a partir del trabajo con estudiantes de primer y segundo curso de la Universidad de Barcelona (n=54) con un nivel

elemental de competencia en lengua rusa. Los estudiantes se han repartido en 4 grupos: 2 grupos en el primer curso (nivel por debajo de A1) y 2 grupos en el segundo curso (A1). Tanto en el primer como en el segundo curso un grupo ha trabajado con la metodología TSLT y el otro con TBLT. Cada uno de los 4 grupos ha recibido 14 horas de formación. La instrucción en TSLT de los estudiantes del primer curso ha sido enfocada al uso de preposiciones y los casos y, en la del segundo curso, se ha centrado en los verbos de movimiento con y sin prefijos. Del mismo modo se ha estudiado en qué medida la aproximación TBLT favorece la adquisición de los mismos elementos lingüísticos. Los datos analizados se han obtenido a partir de tests gramaticales (espacios en blanco, selección múltiple y tests de gramaticalidad), una tarea escrita consistente en escribir una carta o correo electrónico a un amigo, y dos tareas orales (la descripción de una habitación y una *Map task*). Los datos obtenidos en los tests de gramática hacen referencia al uso de las preposiciones, de las formas declinadas y de los verbos de movimiento. En la producción oral y escrita, la corrección se mide a partir del número de errores respecto al número de palabras; la complejidad sintáctica, según el número de proposiciones por *T-unit* y la longitud de las proposiciones; la riqueza léxica se mide a partir del índice de Guiraud; y la fluidez oral, por el índice de sílabas respecto al tiempo de elocución. El análisis estadístico de la adquisición de las preposiciones y las formas declinadas muestra que tanto con TSLT como TBLT se consiguen resultados satisfactorios en el post-test inmediato, pero no en el post-test diferido. En cambio, independientemente del tipo de instrucción, los participantes sí han demostrado progresos significativos en su competencia de los verbos de movimiento, tanto en la producción escrita como en la oral. En la producción escrita, la complejidad sintáctica medida en proposiciones por *T-unit* y la complejidad léxica aumentan significativamente después de la instrucción, mientras que la longitud media de las proposiciones no varía. La corrección en la expresión escrita mejora significativamente en ambos grupos en los tests realizados inmediatamente después del periodo de instrucción. Sin embargo, los estudiantes del grupo TBLT mantienen ese mismo nivel de competencia tres meses después, mientras que los estudiantes del grupo TSLT presentan un nivel similar que el de antes del periodo de instrucción. En la producción oral, los estudiantes de ambos grupos mejoran su corrección en la tarea *Map task*, la complejidad léxica y fluidez en ambas tareas (*Map task* y Descripción de una habitación), pero no consiguen progresos significativos en lo que respecta a la complejidad sintáctica en ninguna de las dos tareas. La comparación de los dos tipos de instrucción muestra que los estudiantes de los grupos TSLT y TBLT no presentan grandes diferencias en la corrección, complejidad sintáctica y fluidez. Del mismo modo, tanto los estudiantes que han seguido la formación TSLT como los que han trabajado con TBLT presentan resultados parecidos respecto a la corrección en el uso de las preposiciones y de las formas declinadas. Sin embargo, los estudiantes de los grupos TBLT consiguen mejores resultados que los de TSLT en lo que se refiere a complejidad léxica y verbos de movimiento.

TABLE OF CONTENTS

ABSTRACT	i
TABLE OF CONTENTS	iii
LIST OF FIGURES	ix
LIST OF TABLES	xi
ACKNOWLEDGEMENTS	xv
INTRODUCTION	1
CHAPTER I. APPROACHES TO SYLLABUS DESIGN	7
1.1 Introduction	7
1.2 Syllabus types	7
1.3 Synthetic syllabi	10
1.3.1 Structures	10
1.3.2 Words	13
1.3.3 Skills	15
1.3.4 Situations and topics	17
1.3.5 Notions and functions	19
1.3.6 Methods supporting synthetic syllabi	22
1.3.7 Problems with synthetic syllabi	24
1.4 Analytic syllabi	25
1.4.1 Communicative Language Teaching	25
1.4.2 Natural Approach	26
1.4.3 Immersion programs	29
1.4.4 Problems with purely meaning-focused syllabi	32
1.5 Analytic task-based syllabi	33
1.5.1 The Procedural Syllabus	33
1.5.2 The Process Syllabus	36
1.5.3 Task-Based Language Teaching	38
1.5.4 Project work	45
1.5.5 CLIL	50
1.5.6 Comparison of task-based approaches	52
1.5.7 Problems with task-based syllabi	53
1.6 Summary of Chapter I	53

CHAPTER II. TASK-BASED LANGUAGE TEACHING: KEY CONCEPTS AND QUESTIONS	55
2.1 Introduction.....	55
2.2 Definitions of task.....	55
2.3 Task selection.....	58
2.4 Task sequencing.....	61
2.4.1 Early proposals for task sequencing	61
2.4.2 Two models of task complexity.....	63
2.5 Research into pedagogic task types	67
2.6 Methodological options in a task-based lesson.....	72
2.7 Task-based assessment.....	76
2.8 Summary of Chapter II	79
CHAPTER III. FOCUS ON FORMS VS. FOCUS ON FORM	81
3.1 Introduction.....	81
3.2 Approaches to teaching language forms	82
3.3 Criteria of the choice of linguistic forms	86
3.4 Focus-on-form techniques	89
3.4.1 Input flood.....	90
3.4.2 Input enhancement.....	92
3.4.3 Input elaboration.....	94
3.4.4 Task-essential language	96
3.4.5 Negotiation of meaning / form.....	98
3.4.6 Task types promoting negotiation of meaning	99
3.4.7 Corrective feedback.....	101
3.4.8 Consciousness-raising tasks.....	104
3.4.9 Input processing.....	107
3.4.10 Direct instruction	110
3.5 Summary of Chapter III.....	113
CHAPTER IV. THE EFFECTS OF TASK-SUPPORTED AND TASK-BASED LANGUAGE TEACHING ON L2 LEARNING	115
4.1 Introduction.....	115
4.2 Task-supported and task-based language teaching	115
4.3 Evaluation of task-based courses	119
4.3.1 Formative evaluations.....	120
4.3.2 Process-product evaluations	123

4.4 Research into task-supported and task-based approaches to teaching L2 Russian	137
4.5 Summary of Chapter IV	140
CHAPTER V. STUDY DESIGN AND METHODOLOGY	141
5.1 Introduction.....	141
5.2 Motivation for this study.....	141
5.3 Research goals and questions.....	142
5.4 Hypotheses	142
5.5 Justification for target items.....	144
5.6 Task elaboration: pilot	148
5.6.1 Objectives	148
5.6.2 Participants.....	148
5.6.3 Procedure	148
5.6.4 Treatment tasks.....	149
5.6.4.1 Task selection and design	149
5.6.4.2 Problems in task performance.....	151
5.6.4.3 Modification of task design	152
5.6.4.4 Conclusions for further task design and sequencing.....	153
5.6.5 Control tasks	154
5.6.5.1 Task design	154
5.6.5.2 Results.....	155
5.6.5.3 Modification of control tasks	156
5.7 The current study: design and methodology	156
5.7.1 Participants.....	156
5.7.2 Design	158
5.7.3 Procedure	158
5.7.4 Treatment sessions.....	160
5.7.4.1 Task-based phase	160
5.7.4.1.1 Task design.....	160
5.7.4.1.2 Task sequencing.....	164
5.7.4.1.3 Focus on form	164
5.7.4.2 Task-supported phase.....	166
5.7.5 Data collection	168
5.7.5.1 Grammar tests	168
5.7.5.2 Written tasks	169
5.7.5.3 Oral tasks	169

5.8 Transcription and coding	169
5.9 Evaluation procedures.....	170
5.9.1 Grammar tests	170
5.9.2 Oral and written tasks	171
5.9.2.1 Research into CAF	171
5.9.2.2 Accuracy	172
5.9.2.2.1 General accuracy	172
5.9.2.2.2 Specific accuracy	174
5.9.2.3 Complexity.....	175
5.9.2.3.1 Syntactic complexity	176
5.9.2.3.2 Lexical complexity	179
5.9.2.4 Fluency.....	179
5.9.2.5 Summary	181
5.10 Statistical analyses	181
CHAPTER VI. RESULTS	183
6.1 Introduction.....	183
6.2 Preliminary analyses	183
6.3 Group comparability	185
6.3.1 TS_Rus I and TB_Rus I.....	185
6.3.2 TS_Rus II and TB_Rus II	185
6.4 The effects of task-supported and task-based treatment on the use of prepositions and case forms	187
6.4.1 Target-like use of prepositions in written and oral production.....	188
6.4.2 Target-like use of case forms in written and oral production.....	191
6.4.3 Fill in the blanks	194
6.4.4 Multiple choice	197
6.4.5 Grammaticality judgment	198
6.5 The effects of task-supported and task-based treatment on the use of verbs of motion	201
6.5.1 Total amount of verbs of motion in written and oral production.....	201
6.5.2 Target-like use of verbs of motion in written and oral production.....	205
6.5.3 Fill in the blanks	209
6.5.4 Multiple choice	211
6.5.5 Grammaticality judgment	214
6.6 The effects of task-supported and task-based treatment on learners' accuracy, complexity and fluency.....	216
6.6.1 Written production.....	217

6.6.1.1 General accuracy.....	217
6.6.1.2 Syntactic complexity.....	219
6.6.1.3 Lexical complexity.....	221
6.6.2 Oral production.....	224
6.6.2.1 General accuracy.....	224
6.6.2.2 Syntactic complexity.....	227
6.6.2.3 Lexical complexity.....	230
6.6.2.4 Fluency.....	234
6.7 Summary of the results	237
CHAPTER VII. DISCUSSION AND CONCLUSION	243
7.1 Introduction.....	243
7.2 The effects of task-supported and task-based treatment on the use of prepositions and case forms	243
7.3 The effects of task-supported and task-based treatment on the use of verbs of motion	245
7.4 The effects of task-supported and task-based treatment on learners' accuracy, complexity and fluency.....	247
7.4.1 Accuracy.....	248
7.4.2 Syntactic complexity.....	249
7.4.3 Lexical complexity	250
7.4.4. Fluency.....	251
7.5 Conclusion	252
7.6 Implications.....	255
7.7 Limitations and further research	256
REFERENCES	259
APPENDICES	
Appendix A. Background information about participants.....	281
Appendix B. Grammar tests	283
Appendix C. Control tasks.....	287
Appendix D. Descriptive statistics on CAF and grammar scores: Means, medians, standard deviations, skewness, and kurtosis	291

LIST OF FIGURES

Figure 1.	Steps in syllabus design (Gilabert 2013)	8
Figure 2.	The levels of the process syllabus (Breen 1984, p. 57)	37
Figure 3.	Task-based syllabus design and language teaching (Long & Crookes 1992)	42
Figure 4.	Steps in project work (Legutke & Thomas 1991)	47
Figure 5.	Steps in project work (Ribé & Vidal 1993)	48
Figure 6.	Steps and processes in TBLT syllabus design (Long 2015, p. 224)	59
Figure 7.	Focus on form activities according to the degree of obtrusiveness (Doughty & Williams 1998, p. 258)	90
Figure 8.	Evidence types (Gass 2003)	94
Figure 9.	VanPatten's model of SLA (VanPatten & Sanz 1995, p. 170)	107
Figure 10.	Sub-dimensions of L2 complexity (Bulté & Housen 2012, p. 23)	176
Figure 11.	Sub-constructs of syntactic complexity (Norris & Ortega 2009)	177
Figure 12.	Median scores on TLU of prepositions (written task, Rus I)	189
Figure 13.	Mean scores on TLU of prepositions (oral task 1, Rus II)	189
Figure 14.	Mean scores on TLU of case forms (written task, Rus I)	192
Figure 15.	Mean scores on TLU of case forms (oral task 1, Rus II)	192
Figure 16.	Mean scores on fill in the blanks test (Rus I)	195
Figure 17.	Mean scores on multiple choice test (Rus I)	197
Figure 18.	Mean scores on grammaticality judgment test (Rus I)	199
Figure 19.	Mean number of VM in written production (Rus II)	202
Figure 20.	Mean number of VM in oral task 2 (Rus II)	202
Figure 21.	Mean scores on TLU of verbs of motion in written production (Rus II)	206
Figure 22.	Mean scores on TLU of verbs of motion in oral task 2 (Rus II)	206
Figure 23.	Mean scores on fill in the blanks test (Rus II)	210
Figure 24.	Mean scores on multiple choice test (Rus II)	212
Figure 25.	Mean scores on grammaticality judgment test (Rus II)	214
Figure 26.	Mean scores on errors per words (written task, Rus II)	217
Figure 27.	Mean scores on mean length of clause (written task, Rus II)	220
Figure 28.	Mean scores on clauses per T-unit (written task, Rus II)	220
Figure 29.	Mean scores on Guiraud's Index (written task, Rus II)	222
Figure 30.	Mean scores on errors per words (oral task 1)	225
Figure 31.	Mean scores on errors per words (oral task 2)	225
Figure 32.	Mean scores on mean length of clause (oral task 1)	228

Figure 33.	Mean scores on mean length of clause (oral task 2)	228
Figure 34.	Mean scores on Guiraud's Index (oral task 1)	231
Figure 35.	Mean scores on Guiraud's Index (oral task 2)	231
Figure 36.	Mean scores on speech rate (oral task 1)	234
Figure 37.	Mean scores on speech rate (oral task 2)	234

LIST OF TABLES

Table 1.	Options in language teaching (Gilabert 2013)	9
Table 2.	Example of structural syllabus in Russian textbooks	10
Table 3.	Johnson's skills syllabus (1996)	17
Table 4.	Example of structural-situational syllabus (from Esmantova 2011)	18
Table 5.	Finochiaro and Brumfit's presentation of functions (1983)	21
Table 6.	Differences in three versions of TBLT (Ellis 2012, p. 197)	42
Table 7.	TBLT, CLIL and project work (Gilabert 2013)	52
Table 8.	Skehan's model of task difficulty (1998a)	63
Table 9.	Robinson's Triadic Componential Framework (2001a)	65
Table 10.	Classifications of tasks (Ellis 2003, pp. 211–217)	68
Table 11.	Task framework (Willis 1996, p. 155)	73
Table 12.	Implicit vs. explicit instruction (De Graaff & Housen 2009)	89
Table 13.	Examples of simplified, authentic and elaborated input (Gilabert 2013)	95
Table 14.	Effect size r (Castellví & Markina 2015)	113
Table 15.	Tasks in TSLT and TBLT	117
Table 16.	Classroom behavior in a form-focused (TSLT) and task-based pedagogy (Ellis 2006, p. 88)	118
Table 17.	Distinguishing characteristics of PPP and TBLT (Long 2015, p. 349)	118
Table 18.	Evaluation of TBLT programs	134
Table 19.	Target forms for Rus I: Case forms and prepositions	146
Table 20.	Target forms for Rus II: Verbs of motion	147
Table 21.	Pilot procedure	149
Table 22.	CEFR descriptors for A2 level (2001, p. 224)	149
Table 23.	Pilot tasks	150
Table 24.	Modification of the library schedule task	152
Table 25.	Pilot oral tasks (Rus II)	154
Table 26.	Background information per group	158
Table 27.	Experimental procedure	159
Table 28.	Tasks implemented in the current study and their correlation with CEFR descriptors	162
Table 29.	Task-supported treatment	166
Table 30.	Total scores for grammar tests	170
Table 31.	Comparing participants on the written pre-test (Rus I)	185

Table 32.	Comparing participants on the written pre-test (Rus II)	186
Table 33.	Comparing participants on the oral pre-test, Task 1	186
Table 34.	Comparing participants on the oral pre-test, Task 2	187
Table 35.	Descriptive statistics: TLU of prepositions	188
Table 36.	Friedman test: TLU of prepositions (written task, Rus I)	189
Table 37.	Mann-Whitney test: TLU of prepositions (written task, Rus I)	190
Table 38.	Mixed between-within subjects ANOVA: TLU of prepositions (oral task 1, Rus II)	190
Table 39.	Descriptive statistics: TLU of case forms	191
Table 40.	Within-subjects effects: TLU of case forms	192
Table 41.	Between-subjects effects: TLU of case forms	194
Table 42.	Descriptive statistics: Fill in the blanks (Rus I)	194
Table 43.	Within-subjects effects: Fill in the blanks (Rus I)	195
Table 44.	Between-subjects effects: Fill in the blanks (Rus I)	196
Table 45.	Descriptive statistics: Multiple choice (Rus I)	197
Table 46.	Mixed between-within subjects ANOVA: Multiple choice (Rus I)	198
Table 47.	Descriptive statistics: Grammaticality judgment (Rus I)	199
Table 48.	Mixed between-within subjects ANOVA: Grammaticality judgment (Rus I)	200
Table 49.	Descriptive statistics: Total amount of VM in written and oral production (Rus II)	201
Table 50.	Mixed between-within subjects ANOVA: Total amount of VM in written production (Rus II)	203
Table 51.	Wilcoxon signed-ranks tests: Total amount of VM in oral production (Rus II)	204
Table 52.	Between-subjects effects: Total amount of VM in oral production (Rus II)	205
Table 53.	Descriptive statistics: TLU of verbs of motion in written and oral production (Rus II)	206
Table 54.	Wilcoxon signed-ranks tests: TLU of verbs of motion in written production (Rus II)	207
Table 55.	Between-subjects effects: TLU of verbs of motion in written production (Rus II)	208
Table 56.	Wilcoxon signed-ranks Tests: TLU of verbs of motion in oral production (Rus II)	208
Table 57.	Between-subjects effects: TLU of verbs of motion in oral production (Rus II)	209
Table 58.	Descriptive statistics: Fill in the blanks (Rus II)	210
Table 59.	Mixed between-within subjects ANOVA: Fill in the blanks	211

	(Rus II)	
Table 60.	Descriptive statistics: Multiple choice (Rus II)	212
Table 61.	Within-subjects effects: Multiple choice (Rus II)	213
Table 62.	Between-subjects effects: Multiple choice (Rus II)	213
Table 63.	Descriptive statistics: Grammaticality judgment (Rus II)	214
Table 64.	Within-subjects effects: Grammaticality judgment (Rus II)	215
Table 65.	Between-subjects effects: Grammaticality judgment (Rus II)	216
Table 66.	Descriptive statistics: Errors per words (written task, Rus II)	217
Table 67.	Mixed between-within subjects ANOVA: Errors per words (written task, Rus II)	218
Table 68.	Descriptive statistics: Syntactic complexity (written task, Rus II)	219
Table 69.	Mixed between-within subjects ANOVA: Mean length of clause (written task, Rus II)	220
Table 70.	Mixed between-within subjects ANOVA: Clauses per T-unit (written task, Rus II)	221
Table 71.	Descriptive statistics: Guiraud's Index (written task, Rus II)	222
Table 72.	Mixed between-within subjects ANOVA: Guiraud's Index (written task, Rus II)	222
Table 73.	Descriptive statistics: Errors per words (oral production)	224
Table 74.	Within-subjects effects: Errors per words (oral task 1)	225
Table 75.	Independent samples t-tests: Errors per words (oral task 1)	226
Table 76.	Repeated measures ANOVA: Errors per words (oral task 2)	226
Table 77.	Independent samples t-tests: Errors per words (oral task 2)	227
Table 78.	Descriptive statistics: Mean length of clause (oral production)	228
Table 79.	Repeated measures ANOVA: Mean length of clause (oral task 1)	229
Table 80.	Independent samples t-tests: Mean length of clause (oral task 1)	229
Table 81.	Within-subjects effects: Mean length of clause (oral task 2)	230
Table 82.	Between-subjects effects: Mean length of clause (oral task 2)	230
Table 83.	Descriptive statistics: Guiraud's Index (oral production)	231
Table 84.	Mixed between-within subjects ANOVA: Guiraud's Index (oral task 1)	232
Table 85.	Mixed between-within subjects ANOVA: Guiraud's Index (oral task 2)	232
Table 86.	Independent samples t-tests: Guiraud's Index (oral task 2)	233
Table 87.	Descriptive statistics: Speech rate (oral production)	234
Table 88.	Mixed between-within subjects ANOVA: Speech rate (oral task 1)	235
Table 89.	Repeated measures ANOVA: Speech rate (oral task 2)	236

Table 90.	Independent samples t-tests: Speech rate (oral task 2)	236
Table 91.	Summary of the results	238
Table 92.	Examples of learners' oral performance (oral task 2)	246

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my supervisor, Dr. Joan Castellví Vives, for his careful supervision, constant support, patience and valuable contribution to this research. Without his participation and help, this work would not be possible.

I will always be very grateful to the team of Eranet Plus programme and, in particular, to the coordinator Dr. Marc Ruiz-Zorrilla Cruzate, for selecting me as a grantee of the Eranet scholarship and for their huge support in academic and financial issues which I faced during these years.

I would also like to address words of infinite gratitude to Dr. Larisa Dunaeva, who encouraged me during my first research steps, patiently guided me while I carried out my first study at the Moscow State University and helped me to develop as a researcher and a teacher.

I would like to thank Dr. Roger Gilabert Guerrero for his valuable advice and interesting suggestions. His questions and comments notably contributed to the development of the present study.

I am very grateful to the students of the Philological Faculty of the University of Barcelona who participated in this study and survived all my pre-tests, post-tests, questionnaires and pilots. Thanks to their hard work and interest in the Russian language I managed to obtain interesting data for my research and I learnt a lot from them.

My special gratitude is to my dearest friend Dr. Elena Safonova who motivated and supported me in this long and sometimes hard trip and generously shared her research experience with me.

Finally, I address words of gratitude to Jose Raul Pérez who had to withstand my two Ph.D. research projects, who has always cheered me up and calmed me down, and who has been at my side whenever I needed him. T'estimo molt!

INTRODUCTION

Behind a language lesson there is a syllabus which contains information about what units to teach, in what order, how to teach them and how to evaluate the progress in L2¹ learning. Language teachers can choose an option among a variety of syllabus types, approaches and classroom procedures which corresponds to their general view on language teaching, and, ideally, meets the learners' needs.

The history of methodologies for the teaching of L2 Russian includes various shifts from one method to another due to changes in learners' profiles and advances in SLA research over the last half century. For example, in the 60's the number of students who had to master academic Russian in order to receive a university degree in the USSR increased, which made methodologists and teachers focus on Russian for specific purposes. It was at this time when the conscious-practical method (сознательно-практический метод) was developed (Beljaev 1965; Scherba 1974, among others) which aimed at developing metalinguistic knowledge of language forms and communicative skills at the same time.

Communicative language teaching (CLT) for Russian as second language was applied in the 80's in Russia, but widespread use of the method in its radical option (that is, focus only on meaning) was not consolidated. Instead, CLT took the form of what Russian methodologists call the functional-communicative approach (функционально-коммуникативный подход) with a focus on the development of communicative competence (Hymes 1966) and communicative skills in four areas: listening, reading, speaking, and writing (Amiantova *et al.* 2001; Vsevolodova 2000, among others). Following Ellis (2003) and Samuda and Bygate (2008), this approach will be referred to as Task-Supported Language Teaching (TSLT).

One communicative approach to teaching foreign languages that has become popular in American, European and Asian contexts is Task-Based Language Teaching (TBLT). However, task-based lessons in Russian are still sporadic (Castellví & Markina 2015; Comer 2007) due to the lack of both empirical studies and learning materials. The present study is one of the first attempts to fill this gap by addressing some theoretical and practical issues which a syllabus design and its implementation imply. The study is specifically concerned with the impact that TSLT and TBLT may have on learning² specific forms in Russian (prepositions, case forms and

¹ For the remainder of this dissertation, the term "L2" will be used to refer to both 'second language' and 'foreign language', and it will be therefore understood as the language that a speaker is learning other than his or her native language or languages.

² The distinction between the terms "L2 learning" and "L2 acquisition" will not be discussed in this study. One of the most widely used definitions of these terms belongs to Krashen (1982, p.10), who defines language learning as "conscious knowledge of a second language, knowing the rules, being aware of them, and being able to talk about

verbs of motion) which usually cause difficulties for L2 Russian learners, and on complexity, accuracy and fluency (CAF) of learners' oral and written production.

In order to compare the effects of TBLT and TSLT on the use of the target forms, the study has been divided into two separate parts. The first part focuses on the use of case forms with and without prepositions, whereas the second part is designed around verbs of motion and giving street indications in Russian. Each part includes fourteen hours of task-supported treatment and fourteen hours of task-based treatment which were implemented in two different groups of learners. The participants ($n=54$) are adult learners with low level proficiency in Russian who were enrolled in the programs "Modern Languages and Literature" and "Linguistics" at the University of Barcelona. Learners' ages range between 18 and 25 years old. More than 60% of participants are bilingual Catalan and Spanish speakers, and other participants have Spanish (and in some cases French/Italian/German) as L1. The study has been carried out under real classroom conditions as a part of the Russian language courses "Llengua Russa I" and "Llengua Russa II". The study follows a pre-test – immediate post-test – delayed post-test design, which means that data were collected three times in each group of participants.

The dissertation has been divided into seven chapters. A table of contents, lists of figures and tables, chapter introductions and summaries have been provided in order to facilitate the reading of the text. The current work also includes four appendices with additional information about study design and statistical data.

The present research consists of two main parts. The first, comprising the first four chapters, presents the state of the art by describing the general pedagogical context in which TBLT and TSLT exist, the theoretical framework and the key concepts of the two approaches. The second part of this dissertation, comprising the last three chapters, presents an experiment which attempts to find out what effects task-based and task-supported approaches to language teaching may have on learners' production and their ability to use the target forms accurately.

Chapter I deals with general questions of L2 syllabus design and instruction and contains a review of three major options in language teaching and corresponding syllabus types. Following the classifications proposed by Wilkins (1976) and Long (1991), we distinguish synthetic syllabi with focus on forms, analytic syllabi with focus only on meaning and analytic syllabi with focus on form and analyze different types of units and methods associated with each syllabus.

them". Language acquisition, on the other hand, is "a subconscious process; language acquirers are not usually aware of the fact that they are acquiring language, but are only aware of the fact that they are using the language for communication." In the present study, the term "learning" will be used to emphasize the process of developing L2 skills which require certain efforts, whereas the term "acquisition" will be understood as subconscious, effortless assimilation of linguistic knowledge.

Firstly, synthetic syllabi, which are organized around grammatical units, lexical units, skills, notions, functions, situations and topics, are described. Secondly, analytic syllabi with an exclusive focus on meaning are analyzed, including the Natural Approach (Krashen & Terrell 1983) and Canadian immersion programs. Finally, analytic syllabi that use task as a unit of syllabus design are described. This description includes both the early versions of task-based syllabi (Prabhu's Procedural syllabus (1987) and Breen's Process syllabus (1984)) and more recent task-based proposals, in particular TBLT, project work and CLIL. At the end of the review of each syllabus type, some problems with these syllabi are outlined.

In Chapter II, the key questions of TBLT research relevant for this study, are extensively reviewed. The chapter focuses on the definition and the key characteristics of the task, different methodological options in task design and the impact of a manipulation of different task features on learners' production. From an interactionist perspective, tasks have been studied according to the flow of information (one-way versus two-way tasks), the number of solutions that can be reached (open versus closed tasks), and the convergence or divergence of task goals. From an information-processing perspective, TBLT studies have focused on the dimensions of task familiarity, the number of elements, amount of planning time, and present and past time reference, among other characteristics. The problem of task sequencing and the models of task complexity proposed by Robinson (2001) and Skehan (1998b) are also discussed in this chapter. The issue of task complexity is specifically addressed with regard to linguistic difficulty of Russian. Finally, some issues of task-based assessment are discussed, including the design of assessment tasks and criteria of task performance evaluation.

Chapter III is dedicated to one of the methodological principles of TBLT – focus on form as opposed to focus on formS. Firstly, the notion of “focus on form” introduced by Long (1989, 1991) is discussed, and its differences from traditional focus on forms are outlined. Secondly, a number of criteria for choosing target forms to be focused on during instruction are discussed, including rule complexity, reliability and scope of the rule, influence of learners' L1, analysis of learners' errors, and frequency. Finally, a description of most commonly used focus-on-form techniques (input flood, input enhancement, input elaboration, explicit and implicit corrective feedback, consciousness-raising tasks, input processing, etc.) and a brief summary of studies related to their effects on learning different L2 forms are given. The analysis of findings reported in most studies shows that attention to forms within a meaningful context can be beneficial for the learning of these forms and that groups who receive treatment in which the target structures are somehow outlined (e.g., through enriched input, recasts, etc.) usually perform better than groups who receive the same treatment but with no focus on form.

Chapter IV compares TBLT and TSLT on a number of characteristics (the process of syllabus design, task selection and sequence, the learner's and teacher's role, strategies of classroom behavior, language focus, learning materials, among others) and reviews evaluative studies on the efficiency of both approaches. The results of different types of treatment are usually (but not exclusively) measured in terms of accuracy, fluency and complexity of learners' production. Comparative studies on task-based (TB) and grammar-based (GB) approaches show that learners in the TB groups demonstrate higher fluency and syntactic complexity than learners in the GB groups, with no significant differences in lexical complexity and accuracy between the groups. Finally, the results of some studies on Russian L2 teaching are discussed. The findings reported in these studies show that accuracy remains an important challenge for learners of Russian even after many hours of formal instruction.

The next three chapters present the current study design, the analysis of obtained results and conclusions drawn from the study. In Chapter V the motivation for this study, the research goals, and the specific questions and hypotheses addressed by this study are advanced. The questions that this study attempts to answer are the following:

1. Do task-supported and task-based approaches help to promote the accurate use of case forms and prepositions?
2. Which of the two approaches is more efficient for learning new lexical items (Russian verbs of motion with prefixes)?
3. How does each approach affect learners' oral and written production in terms of general accuracy, syntactic and lexical complexity, and fluency?

Chapter V also provides the justification for the target items in the present research and describes the results of the pilot study conducted prior to the experiment. Then the chapter focuses on the experiment itself by providing information about the participants, study design and procedures, treatment sessions, data collection, the transcription and coding of learners' oral production, and a variety of measures applied in the study. These measures include traditional grammar tests (fill in the blanks, multiple choice and grammaticality judgment test), on the one hand, and measures for fluency, accuracy, lexical and syntactic complexity of learners' production, on the other hand. It's important to highlight that the methodology of the present research is different from most comparative studies on the relative effectiveness of TBLT and TSLT which usually involve "easy learning targets and outcome measures that favor explicit treatments" (Long 2015, p. 320). The current research involves complex target forms (Russian case forms and verbs of motion) and a variety of outcome measures which favor both explicit and implicit treatment.

Chapter VI presents the results of statistical analyses concerning the three research questions in this study, which are summarized in Table 90. In the final chapter, Chapter VII, the obtained results are considered in relation to previous research and our hypotheses about the effects of two types of treatment on learners' production and the use of the target forms. The effects of task-supported and task-based treatment on the use of case forms, verbs of motion and on CAF dimensions are compared. Then the implications of the results of this study for teaching Russian as a foreign language are discussed. Finally, some limitations of the study are acknowledged, and some directions for further research are proposed.

CHAPTER I

APPROACHES TO SYLLABUS DESIGN

1.1 Introduction

The main objective of this chapter is to present the general context in which task-based and task-supported approaches to language teaching exist. The chapter consists of four parts. First, a general overview of different options in syllabus design (Long & Robinson 1998; Wilkins 1976; White 1988) will be given. Then, following Gilabert (2004), three major options in language teaching – focus on forms, focus on meaning and focus on form – and corresponding syllabi, synthetic and analytic, will be analyzed. This analysis will include the questions about the units of syllabus design, principles of their selection and sequencing, and methods and teaching strategies associated with each syllabus. Finally, advantages and shortcomings of each type of syllabus will be discussed.

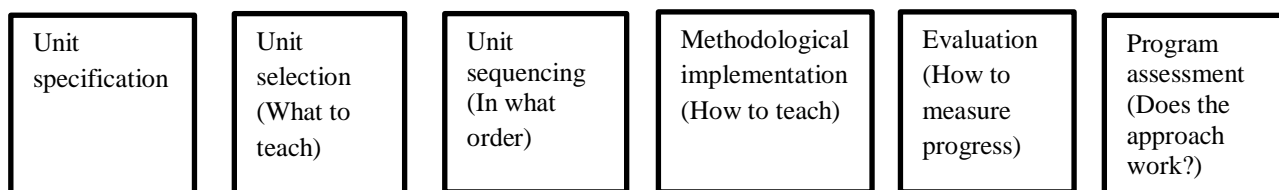
1.2 Syllabus types

Breen (1984, p. 47) defines syllabus as “a plan of what is to be achieved through our teaching and our students’ learning.” The choice of a syllabus type expresses, directly or indirectly, “certain assumptions about language, about the psychological process of learning, and about the pedagogic and social process within a classroom” (Breen 1984, p. 49).

Proposals for a syllabus type should specify at least three things: 1) units of analysis employed, 2) course content, i.e., what is to be taught in a course, and 3) the order in which it is to be taught, i.e., how the content is to be sequenced (Long 2015; Long & Crookes 1993). To be more complete, these proposals should also involve decisions about methodology design, testing, and program evaluation (see Figure 1).

In the process of syllabus design the choice of ‘unit’ plays the central role (Gilabert 2004; Long 1991, 2015; Robinson 1998) since it reflects the program designer’s and teacher’s theories about second language learning and it will inevitably affect decisions in other areas, in particular, decisions about “how and under which conditions that unit is best taught, how language is meant to be used and learned, and how learning should be evaluated” (Gilabert 2004, p. 116).

Figure 1. Steps in syllabus design (Gilabert 2013)



Historically, different approaches to language teaching have triggered various types of syllabi which differed in roles assigned to the learner and the teacher, presentation of language material, methodologies and classroom practices. According to the content presentation and the learner's role in its assimilation, a distinction was made between synthetic and analytic syllabi (Wilkins 1976) and Type A and Type B syllabi (White 1988).

In synthetic syllabi “different parts of language are taught separately and step by step so that acquisition is a process of gradual accumulation of parts until the whole structure of language has been built up” (Wilkins 1976, p. 2). In other words, these syllabi segment the target language into discrete linguistic items (e.g., grammar rules and forms, words and collocations, notions, functions) for presentation one at a time. They rely on learners' assumed ability to put together, or synthesize, in real world performance the parts of the language system they have been exposed to separately (Robinson 1998). In contrast, in analytic syllabi much greater variety of linguistic structures is permitted from the beginning, and they “present the target language whole chunks at a time, without linguistic interference or control” (Long & Crookes 1993, p. 11). They do not divide up the language to be presented in classrooms, but involve holistic use of language to perform communicative activities. Analytic approaches are organized “in terms of purposes for which people are learning language and the kinds of language performance that are necessary to meet those purposes” (Wilkins 1976, p. 13). The learner's role in these syllabi is to analyze aspects of language use and structure as the communicative activities require them and “to approximate his own linguistic behavior more and more closely to the global language” (Wilkins 1976, p. 2).

A broader conceptualization of syllabus types can be found in White (1988) who distinguished between Type A and Type B syllabi. Type A syllabi focus on what is to be learned and are interventionist and teacher-oriented. Language to be taught is preselected and divided into small pieces. Type B syllabi focus on how the language is to be learned. They are non-interventionist and learner-oriented. They don't involve artificial pre-selection of items and their objectives are determined in the process of negotiation between teacher and learner.

Finally, syllabus types can be divided into three groups according to what is in focus – form or meaning. Long (1989, 1991) distinguished between focus on formS and focus on form. Focus on form “draws students’ attention to linguistic elements as they arise incidentally in lessons whose overriding focus is on meaning or communication” (Long 1991, p. 46), while the traditional notion of formS always entails isolation or extraction of linguistic features from context or from communicative activity. As Doughty and Williams (1998) summarize, focus on form entails a focus on formal elements of language, whereas focus on formS is limited to such a focus, and focus on meaning excludes it. Table 1 presents how the key questions of syllabus design are dealt with within each approach.

Table 1. Options in language teaching (Gilabert 2013)

	Option 1 Synthetic Focus on FormS	Option 2 Analytic Focus on Meaning	Option 3 Analytic Focus on Form
Units	Structures, words and collocations, skills, notions and functions, situations and topics.	Subject-matter, conceptual units.	Tasks
Selection	Based on descriptive grammars and general consensus.	Determined by subject-matter specialists and/or linguists.	Determined by needs analysis and/or negotiation with learners.
Sequence	Experts’ intuitions about difficulty, usefulness, frequency.	Largely unspecified. Subject-matter experts’ intuitions.	Determined by task complexity/steps in the project/subject-matter.
Taught	One at a time in a linear, additive fashion.	Language is not explicitly taught. The emphasis is on the provision of sufficient quantities of positive evidence (input and models).	Taught in communicative, meaning-oriented, learner-centered classes.
Learned	Learners are meant to internalize the language system with the help of explicit instruction and practice.	Language is learned unconsciously, incidentally, while doing something else (like in L1 acquisition).	Language is meant to be learned in context, both consciously and unconsciously.
Learner’s role	To synthesize the separate units in communication.	To analyze and assimilate the language and use it in communication.	To analyze the language implicitly and by receiving feedback.
Language focus	Language is focused on.	Language is treated not as an object of study, but as a medium of communication. It is not focused on. Form is left for the learners to work out.	Language is addressed as it incidentally appears in meaning-oriented activity.
Methods	Grammar translation Audiolingual method Silent way Total physical response	Immersion Natural approach	TBLT
Evaluation	Usually in terms of accuracy through language tests or four skills model (multiple-choice, fill in the blanks, listening and reading comprehension, etc.)	Mastery of subject-matter contents	In terms of task performance / global evaluation of the project by learners, teacher and sometimes external observer/ knowledge of the subject-matter.

Below different syllabi are analyzed using parameters offered by Gilabert (2004), Long (2015), Long and Crookes (1992, 1993), Long and Robinson (1998), Robinson (1998), Skehan (1996a), among others. The analysis includes the questions about the units of syllabus design, their selection and sequencing, and methods and teaching strategies associated with each syllabus.

1.3 Synthetic syllabi

1.3.1 Structures

The structural syllabus, sometimes called grammatical (Wilkins 1976; Yalden 1983), is widely used in L2 classrooms. It was described in works by Halliday, McIntosh and Stevens (1964), Mackey (1965), Ellis (1993a, 1994), among others. The structural syllabus consists of a series of isolated linguistic forms (case forms, verb tenses, etc.), generally presented one at a time, but occasionally in contrasting pairs, e.g. singular and plural forms of nouns, perfective and imperfective aspect of verbs. The selection of the units is based on descriptive grammar and on a general consensus about the patterns of language that must be taught. As for sequencing criteria, it is one of the problematic issues of structural syllabus design. As Wilkins (1976) pointed out, a good deal of decision-making about grading and sequencing linguistic items remains subjective. Traditionally, it is based on experts' intuition about their difficulty, usefulness, and/or frequency.

Structural syllabus is presented in many modern course books of foreign languages and additional materials for students that aim to make them practice certain forms. Table 2 contains some examples of how the linguistic material is organized in textbooks for beginner learners of Russian.

Table 2. Example of structural syllabus in Russian textbooks

Ovsienko (2008). Russian for Beginners (for English-speaking learners)	Ruiz-Zorrilla Cruzate, Kornakov, & Castellví Vives (2001). Curs de lengua russa. Nivel inicial
<p>Unit 1 Text. <i>Knowing each other</i> Dialogues Grammar. The Noun. The Gender of Nouns. The Personal Pronouns он (he), она (she), оно (it). Possessive Pronouns Exercises</p>	<p>Unitat 1 Gramàtica. Substantius. Gènere i nombre. Nominatiu com a subjecte. Pronoms personals. Pronoms possessius. Primera conjugació. Segona conjugació. Acusatiu com a complement directe. Substantius inanimats i femenins en acusatiu. Adverbis de mode. Conjunccions <i>u i a</i> Text. <i>These are our friends</i> Exercicis</p>

<p>Unit 2 Text. <i>Oleg is at home</i> Dialogues Grammar. Adverbs of Place. Word Order in a Sentence. Complete and Short Answers. The construction У меня есть “I have...” Exercises</p>	<p>Unitat 2 Gramàtica. Prepositiu amb significat locatiu. Substantius singulars en prepositiu. Les preposicions <i>в</i> i <i>на</i>. La forma <i>есть</i> amb significat locatiu. El verb <i>жить</i>. Els adverbis de lloc. El verb <i>учиться</i>. L'adjectiu singular i plural en nominatiu. L'adjectiu singular en acusatiu i prepositiu. Pronoms possessius singulars en acusatiu i prepositiu Text. <i>These are our old friends</i> Exercicis</p>
--	--

Methodologically, structural syllabus is usually implemented within a Presentation-Practice-Production (PPP) approach. As it follows from the approach's name, it includes three stages: 1) a focused presentation stage where new material is introduced; 2) practice activities that are designed to enable learners to rapidly produce the material which has been presented; 3) a production stage where learners are expected to use language freely and flexibly.

The structural syllabus remains one of the most popular syllabus types in teaching foreign languages. There are a few reasons for its popularity. It is promoted and supported by leading publishing companies due to its universality. It covers general grammar and vocabulary needs and, consequently, can be used in a wide range of learning contexts. Moreover, focusing on grammar rules and forms helps students to pass standardized national or international language exams where grammar is usually a major component. Skehan (1996b) points out that a structural syllabus gives teachers a feeling of professionalism; it is easily organized into units without learner interference; it has clear learning goals and precise, well-defined evaluation systems. Ellis (1993a) argues that the structural syllabus draws learners' attention to specific formal features in the input and helps them to notice the gap between these features and the ones they produce.

Although widely used by classroom teachers and supported by publishers, structural syllabi have been severely criticized by SLA researchers. For example, Skehan (1996b, p. 18) asserts that levels of attainment in foreign language learning are poor, and students commonly leave school without being able to communicate in a foreign language: “Most language learning is associated with relative failure. Only gifted learners achieve impressive levels of proficiency.” The grammar-based approach was declared “unsatisfactory, failing as it does to reflect either the nature of language or the nature of learning” (Lewis 1996, p. 11).

The most serious problem with the structural syllabus (and with all synthetic syllabi) is that it assumes a model of language acquisition that conflicts with SLA research and with what is known about language learning (Krashen & Terrell 1983; Long & Crookes 1992, 1993; Skehan 1996b). The underlying theory for these syllabi has been based on the belief that a precise focus on a particular form leads to learning and automatization. In other words, learners are expected

to learn what is taught in the order in which it is taught. Synthetic syllabi call for immediate target-like mastery of the “form of the day”, while SLA research shows that learning is constrained by internal processes and that learners rarely move from zero to target-like mastery of new items in one step. On the contrary, they pass through developmental sequences which often include quite lengthy stages of non-target-like use of forms and seem to be impervious to instruction (Long & Crookes 1993). Ellis (1993a, p. 91) calls it the “learnability problem” of the structural syllabus: “Learners are often unable to learn the structural properties they are taught because the manner in which they are taught does not correspond to the way learners acquire them.”

Krashen (1981, 1982, 1985) also focuses on the contradictions and limitations of the structural syllabus basing his argumentation on SLA research achievements of that time. Firstly, he claims that we don’t know the correct natural order of presentation and acquisition and what we have is information about a few structures in a few languages. Secondly, the grammar syllabus only works for those students who happen to be ready for the “structure of the day.” However, all learners vary in their rate of acquisition. Thirdly, a grammatical focus invariably distorts any attempt to communicate. Other weak points of structural syllabus include: 1) the inevitable unnaturalness and artificial language of input materials; 2) misleading mixing of different functions of language as a result of using the same form; 3) negative effects on learners’ motivation (Long & Crookes 1993).

Given that there is little evidence in its favor, it is surprising how resistant to change the structural syllabus has been and that it has not been totally replaced by other syllabi. Skehan (1996b, p. 18) explains it by the teacher-focused nature of the approach that turned out to be “more influential than the approach’s lack of success, its lack of theory and its lack of explicit concern with the learner.” Some researchers and many teachers would disagree with this claim and such a negative evaluation of the traditional PPP approach. As Swan (2005, p. 387) points out, “there is no evidence that traditional methods have failed. Countless people seem to have learnt languages over the centuries through the kind of instruction currently condemned in the TBI (task-based instruction) literature. The fact is simply that such methods have not done very well. All approaches will fail in particularly unfavourable contexts such as large secondary-school classes of unmotivated adolescents.”

As a response to the obvious problems with the structural syllabus, different attempts were made to improve synthetic syllabi by making them more motivating and communication-oriented, and other units of classroom activities were offered.

1.3.2 Words

Started with various vocabulary selection studies in the 1930s (Palmer, West, Faucett, & Thorndike 1936) and resuscitated in the 1980s with the help of computer analysis of large text databases, the lexical approach to language teaching found its place among other synthetic syllabi. The lexical syllabus was discussed in Sinclair and Renouf (1988), Willis (1990) and Lewis (1996), among others.

Sinclair and Renouf (1988) point out three basic questions that must be answered to construct an adequate lexical syllabus: 1) what counts as a word, 2) which words should be included (criteria for lexical selection), and 3) what it is about a word that we want to teach. The term “word” traditionally denotes a base form, such as *give*, and an associated set of inflexions, such as *gives, giving, gave, given* (what in computer linguistics is called lemma). This view on the concept “word” is accepted for the lexical syllabus but with limitations. If the individual word forms are very different from each other in their primary meanings and central patterns of behavior (for example, *certain* and *certainly*; *real* and *really*; *use* and *used*; *one* and *ones*), they are considered as different words in a language course.

The main criterion for words selection is frequency of occurrence in a text corpus (for Sinclair and Renouf (1988) it is the Birmingham Corpus) and the typical collocations that words form. The list of the most frequent words must be extended with “items of maximal utility and power” that will include words relating to everyday domestic reality, to physical sensations and personal emotions, words for classrooms procedures, etc. The authors do not elaborate on this criterion of usefulness or provide additional explanation.

As for aspects of word that should be taught, the main focus of study should be on the commonest word forms in the language; their central patterns of usage; and the combinations which they typically form. This information can be obtained through the analysis of large, computer-held banks of text that provide evidence of typical language use. Sinclair and Renouf (1988) give an example with the word *see*. Textual evidence shows that the first and second most frequent uses of *see* are in the phrases *you see* (an indication of interactive concern in spoken discourse) and *I see* (“I understand”), although native speakers would intuitively say that it is *seeing* through one’s eyes which is the major use of this verb. Sinclair and Renouf argue that computers are more reliable than native-speaker intuition. Words in the lexical syllabus should be presented in their typical combinations that also can be singled out by computer analysis. These combinations include both lexical collocations (e.g. *happy marriage*) and collocational patterns

of common grammatical words (for example, *each* occurs significantly with units of time – *hour, day, week*; the framework *a ... of* typically encloses *lot, kind, number*).

According to Sinclair and Renouf (1988, p. 155), the lexical syllabus does not encourage the acquisition of large vocabulary. Instead, “it concentrates on making full use of the words that the learner already has, at any particular stage.” This syllabus is particularly effective for teaching English as it “makes excessive use, e.g. through phrasal verbs, of its most frequent words, and so they are well worth learning.”

The Collins COBUILD English course (Willis & Willis 1988) is an exemplary pedagogic implementation of the lexical syllabus, where word and collocation are the units of analysis. Arguing against traditional Presentation – Practice – Production sequence, D. Willis proposes taking “meaningful exposure as a starting point” (Willis 1990, p. 4). Exposure should be organized in three ways: a) language is graded in difficulty; b) language exemplifying the commonest patterns is selected; and c) the language syllabus is itemized to highlight important features. The content of the syllabus is lexically graded on the base of statistical evidence received from corpora of language use. Word frequency is identified at the 700 word, the 1,500 word, and the 2,500 word levels. Words in the corpora are itemized as collocations exemplifying each word’s typical patterns of use. D. Willis (1990, p. 7) claims that the lexical syllabus “does not dictate what will be learned and in what order,” rather “it offers the learner experience of a tiny but balanced corpus from which it is possible to make generalizations about the language as a whole.” In other words, the learner corpus which forms the basis of exposure at each level is carefully itemized, but these items are not presented individually and serially.

Willis describes the development of the COBUILD course as “a process of first intuitively deciding on interesting topics, then developing tasks and choosing texts to complement them, and then highlighting lexical items within, e.g. the first 700 words level, as they occurred in the texts” (Robinson 1998, p. 10). This series of highlighted items sequenced according to teacher intuition forms the syllabus.

The methodology accompanying the syllabus described in D. Willis (1990) and J. Willis (1996) involves a pre-task introduction to a topic and exposure to texts; a task cycle where a task is planned, drafted and rehearsed; and a final language focus including controlled practice where learners consciously focus on forms used during the task. Student work on pedagogic tasks is a two-stage classroom procedure. Tasks are first to be done in small groups and then a public report to the whole class is to follow. Since this report is potential input for other learners, it is expected to be accurate and fluent.

Robinson (1998, p. 10), analyzing the COBUILD course, comes to conclusion that it is “largely determined by the choices of texts and tasks – topics about which the lexical syllabus says nothing. This is, then, a language-focused synthetic syllabus, but with some control given to the learner about which forms to attend to and focus on, since the itemized corpora at each level function as a guide, rather than as a prospective plan.”

The late version of the lexical approach was presented by Lewis (1996). He points out that language consists of different kinds of lexical items: single words, polywords (*by the way, on the other hand*), relatively fixed collocations (*to solve the problem, an initial reaction*), fixed expressions (*I'll see what I can do*), and sentence frames (*Considerable research has been done in recent years on the question of ...*). Lewis claims that our language use is made of prefabricated chunks, often much larger than single words, hence the instruction must focus on fixed expressions that occur frequently in native speakers' speech.

The lexical syllabus managed to overcome some problems of the structural syllabus. Speaking about the advantages of Willis' COBUILD course, Long and Crookes (1993, p. 25) mention that it contains “some of the most authentic texts and dialogs in commercially published language teaching materials” and attractively presented, stimulating and carefully planned exercises and pedagogic tasks. Sinclair and Renouf (1988, p. 155) state that the main benefit of a lexical syllabus is that it emphasizes utility and “only offers to the learner things worth learning.”

However, the lexical syllabus was criticized on the grounds that it “takes no cognizance of how a second language is learned” and “it is threatened by other general problems of synthetic syllabuses” (Long & Crookes 1993, p. 26), in particular, by the wrong assumption that the unit which is presented will be what is learned and by the organization of material in an isolating fashion. Robinson (1998) adds that no account is taken of research into learnability and learning processes in selecting the collocations presented in corpora at each level of exposure.

1.3.3 Skills

Another attempt to make a synthetic syllabus more communicative was made by Johnson (1996) who offered a skills syllabus. Drawing on the distinction between declarative and procedural knowledge made by Ellis (1993b), Johnson argues that SLA and general skill learning draw on the same general cognitive mechanisms. Analyzing Johnson's approach to syllabus design, Robinson (1998, p. 10) asserts that “traditionally, skill acquisition has been viewed as a speed-up in the use of initially attention-demanding declarative knowledge. With

practice, attentional demands diminish and declarative knowledge is proceduralised,” which is valuable because it allows greater generalizability of language use, and is context-free, in contrast to procedural knowledge. Johnson (1996) concludes that his proposals support a skills syllabus, similar to, but going beyond the earlier attempts of Munby (1978) and Wilkins (1976) to specify the units of communicative syllabus design.

Johnson (1996) tries to create links between an information processing-based learning theory and communicative language teaching methodology. One of the notions he discusses is “doing more than one thing at the same time” which, in language teaching, is associated with the term “fluency”. To develop fluency, he argues, information-gap and information-transfer exercises should be used. Johnson applies this principle to both production and comprehension exercises, claiming that for selections to be made quickly, real time conditions should be stimulated and this can only occur when there is an information gap.

The purpose of skill-based teaching is to form skilled performers who possess the following characteristics: they work speedily, quickly detect and reject errors, automatize responses, can anticipate (“think ahead”), are reliable (their level is less likely to disintegrate in distracting or disturbing situations) and are able to ignore inessentials and maximize information from information available (Reed 1968).

An initial variant of skill taxonomy was presented by Munby (1978). His skills are stated in terms of an “operational element” (the action done) plus a statement of the domain over which it operates – “communicative feature.” Johnson (1996, p. 164) maintains Munby’s distinction though broadening his categories and replacing his term “communicative feature” with the label “content specifications” which implies “certain specific areas of language that the so-called communicative movement chose to focus on.”

Johnson (1996) identifies different levels of analysis at which skills identification can be made and proposes a four-tier model of syllabus design. The first level of analysis is linguistic (syntactic and phonetic). It includes language specific skills, such as forming the present perfect correctly or contrasting [i] and [i:].

The second level of analysis is semantic and pragmatic. Correspondingly, the second tier would contain semantic categories, such as notions and functions, “but only those about which pedagogically accessible generalizations can be made,” (Johnson 1996, p. 165), that is, notions and functions which can be generalized to many contexts.

A third tier that corresponds to the organizational level of skills identification includes skills often referred to in process approaches to teaching (Breen 1987a; White & Arndt 1991). For example, writing skills would involve generating ideas for an essay, drafting essays,

structuring and evaluating them. Process syllabi concentrate on how language is learned and used, rather than simply listing and teaching the product – language items. Johnson (1996) claims that his framework is capable of handling both processes and products. Thus, in Johnson’s syllabus, skill is used as a term to cover three different types of unit: language item, semantic category and writing strategy.

The fourth and final tier of Johnson’s skills syllabus concerns processing conditions that are implied by settings, topics and roles. For example, roles of “boss” and “employee” are associated with certain structural and semantic forms (*tu* and *vous* in French). In order to provide coverage of processing conditions in a meaningful way, Johnson uses tasks. The level of complexity of the classroom task should be specified and enter into sequencing decisions: “Task complexities are processing conditions, and for this reason we conclude that task specifications are an essential component for skills syllabuses” (Johnson 1996, p. 168). Johnson’s four-tier model of syllabus design is presented in Table 3.

Table 3. Johnson’s skills syllabus (1996)

	Level	Items	Examples
Content specifications	Phonetic-syntactic	Language specific skills	Distinguishing /i/ from /i:/; using the present perfect
	Semantic and pragmatic	Notions and functions	Expressing certainty (a notion); invitation (a function)
	Organizational (strategic)	Process skills	In writing: generating ideas for an essay, drafting essays, structuring and evaluating them
Operational elements		Processing conditions: settings, topics and roles	The intercontinental phone line (a setting); boss and employee (roles)

In summary, Johnson also favors a synthetic syllabus based on subskills at a number of levels: linguistic, semantic and strategic. The syllabus designer is supposed to inventory the subskills at each of these levels (as Munby (1978) attempted to do) and then sequence them.

1.3.4 Situations and topics

As Long and Crookes (1993) point out, the term “situational” has two meanings in second language syllabus design. The first use is applicable to a “structural-situational” syllabus that refers to language teaching “by means of a series of everyday situations, while at the same time grading the structures which are presented” (Alexander 1967, p. xiii). This syllabus combines the structural and situational approaches in structured dialogs, directed discourse and

situational grammar drills. In this case, structures determine the sequence of situations. As an example of such a syllabus, Long and Crookes mention L.G. Alexander’s course *First Things First* (1967) and O’Neill’s textbook *English in Situations* (1970).

The second use of “situational” implies courses which are organized around situations and deal with structures as they arise. They usually consist of “topics, which are illustrated by situations which show how an interaction in that topic area might develop” (Mohan 1977, p. 251). It doesn’t mean that topically organized materials imply the use of situations, as in *English Topics* (Cook 1974) which includes “a collection of independent units of stimulus material and discussion questions arranged in terms of topics of general interest” (Long & Crookes 1993, p. 19). However, situations and topics usually do correlate with each other. A number of modern textbooks and learning materials are organized around situations and topics. Table 4 illustrates the unit organization in the textbook of Russian “5 elements” (Esmantova 2011) which represents structural-situational syllabus. Lists of topics and situations are included in educational standards and programs such as The Common European Framework of Reference for Languages (2001) and Государственный стандарт по русскому языку как иностранному (Gosudarstvennyj standart po russkomu jazyku kak inostrannomu) (2001).

Table 4. Example of structural-situational syllabus (from Esmantova 2011)

Unit 1	Grammar	Conversation / Situations	Text
	Gender of nouns. Plural forms. Where? (here – there). Accusative case. Numerals.	In a restaurant: Tea or coffee? On the street: Where is the metro station?	Students

In order to design a situational syllabus, different features of situations should be analyzed, including “the physical context in which the language event occurs, the channel (spoken or written) of communication, whether the language activity is productive or receptive, the number and the character of the participants, the relationships between them” (Wilkins 1976, p. 16). An inventory of language situations and their features will be different for different types of learners.

The main advantage of both situational and topical syllabi is their motivational potential, especially if the material selection is based on needs identification. Nevertheless, the process of designing such syllabi causes some difficulties. First, situations and topics are so numerous and sometimes complex that it is difficult to define and distinguish them: “Most situations are too ill-defined and too broad to be used in identifying learner needs, designing materials or testing” (Long & Crookes 1993, p. 23). Moreover, topics and situations have a tendency to merge into one another.

The second problem consists of the lack of precision for materials design. The researchers claim that it is impossible to predict language from situation: “Notions and functions realized, and the lexical items and grammatical structures used to encode them are all still rather unpredictable” (Long & Crookes 1993, p. 21). This statement seems only partly true because obviously it is not possible to predict all vocabulary and grammar items that learners will need in a given situation. However, there will be grammatical and lexical forms that have a high probability of occurrence in concrete situations (like “Can I have the bill?” in a restaurant). Course designers can select a necessary minimum (keywords, typical phrases) that will help learners to achieve basic communicative goals. However, as Wilkins (1976) points out, language does not have to be related to the situation and a client of a restaurant may need to complain about the quality of the food and not to ask for the bill. He concludes that “the limited aims of a tourist, a waiter or a telephone switchboard operator might be provided for adequately in this way. However, they would, by definition, be unprepared for anything “out of the ordinary” (Wilkins 1976, p. 18).

The third problem with situational and topical syllabi is the lack of clear principles of units sequencing. They are usually ordered on the basis that the learner will encounter the situations (airport, taxi, a place to live, etc.) or according to the structural complexity of the dialogues (McKay 1980).

1.3.5 Notions and functions

The notional-functional approach to language learning places major emphasis on the communicative purposes of a speech act. It focuses on what people want to do or what they want to accomplish through speech. Notions can be either general, abstract concepts such as space, time, quantity, and quality, or specific concepts, which are usually referred to as “contexts” or “situations” (Gilbert 2004). Functions correspond to language functions, such as apologizing or complaining.

Different proposals of the notional-functional syllabus, for example those of Wilkins (1976) and Finnochiaro and Brumfit (1983), present it in contrast with grammatical and situational approaches. Wilkins (1976) views the language as groups of the linguistic devices needed to encode three types of meaning. First, “we express our perceptions of events, processes, states and abstractions” (Wilkins 1976, p. 21). Wilkins calls these types of meaning *semantico-grammatical categories*. They include expression of time, quantity, space, relational meaning

and deixis (time, place and person). While expressing his perceptions, the speaker simultaneously expresses his own attitude towards what he is saying or writing. This type of meaning is called *modality* (such as necessity, possibility, certainty, doubt). The third type of meaning that is intended to handle the use of language is a *communicative function* of an utterance. Wilkins' list of functions includes judgment and evaluation, offering, advising, requesting, denying, expressing personal emotions and emotional relations (greeting, gratitude, etc.), among others. Each group unites language sources of different levels: phonetic (intonation), lexical (words and collocations), grammatical (for example, verb tenses, aspects and modes) and syntactic (word order, impersonal structures and others).

Wilkins (1976) presents the inventory of notional categories which is intended as a tool in the construction of syllabi. The process of selecting items is based on consideration of what learners should be able to communicate in the foreign language. The question of sequencing items remains problematic as with other synthetic syllabi. There is no pedagogic significance to the ordering of the categories in Wilkins' list. The decision about sequencing of units is left to syllabus designers. As a result, the order in which communicative functions appear in different textbooks seems random.

As Wilkins (1976) points out, within the notional syllabus, units are taught in a cyclic rather than linear way. It means that a course is designed to expand the learner's semantic repertoire progressively, from the simplest and least differentiated manners to express a number of concepts and functions at the lowest level to the same functions expressed with far greater subtlety and nuance at more advanced levels. Consequently, the sequencing is a matter of the relationship between different cycles and units within each phase of the cycle. In the process of distribution of language material between different cycles, the stylistic dimensions of interpersonal relations and medium (speech or writing) will be of particular importance. A syllabus designer will have to choose how much weight to give to grammatical criteria. If he puts it first, the learner will be led to use forms that the native speaker finds inappropriate in certain situations. On the other hand, if grammatical factors are ignored, the learner will be faced with forms that are too complex for his stage of language learning, especially at the early stages. Therefore, both ways of presenting language material seem unsatisfactory.

When it comes to sequencing units within the same cycle, it is not clear if there is any intrinsic ordering to the categories and any way of linking one unit to the next. Wilkins (1976) offers an introduction of a story line as a possible solution. The process of unit grading and sequencing gets complicated by the fact that "any actual utterance inevitably contains many different kinds of grammatical meaning and may simultaneously perform more than one

function. Looked at pedagogically, there is no way in which a single element of meaning (concept or language function) could be taught without other kinds of meaning simultaneously being introduced” (Wilkins 1976, p. 24).

The expanded and updated version of the notional-functional syllabus was offered by Finnochiaro and Brumfit (1983) who tried to explain its theoretical basis and described steps in planning the curriculum content, syllabus design, methodology and evaluation.

Finnochiaro and Brumfit’s syllabus (1983) includes three interrelated elements: function, situation and notions. The basic functions depend on the communicative purposes of the speaker. The situation is characterized by three factors: people who are talking (their age, sex, attitude, role, status, distances maintained), place and time. The specific notions depend on the functions, the situation and the topic. Table 5 illustrates that the same function (apologizing) can be realized in different situations and with different language resources.

Table 5. Finnochiaro and Brumfit’s presentation of functions (1983)

Function	People	Situation setting	Topic	Communicative expressions and/or exponents
Apologizing	A mother and a son Another customer	Supermarket (causing boxes to fall off the shelf and bumping someone’s cart)	Buying food and drinks for a party	I’m so sorry. How clumsy of me! That’s all right. John, help the lady.
Apologizing	Two theater-goers	Theater (asking someone to change seats)	Watching the ballet	Excuse me. Would you mind...? I’m very grateful.

Although Long and Crookes (1993) claim that the list of functions, situations and topics is non-finite, such lists still can be found in different works (Alexander 1975; Finocchiaro & Brumfit 1983; Richterich 1980; van Ek 1980; Wilkins 1976). The textbook designers and teachers are supposed to choose items from these lists and order them according to such criteria as “frequency of occurrence and need in real-life communication, generalizability to other situations, distance between L1 and L2, complexity and familiarity of structures, importance of and interest in topics” (Finnochiaro & Brumfit 1983, p. 53). As we have seen in Section 1.3.1, the same criteria of selection and sequencing were used for the structural syllabus. However, the methods associated with the notional-functional approach are different (group work, paired practice, role play, problem solving, playing games, manipulative exercises, “free” communication, among others) and aim to involve learners in communication from the very beginning and develop their fluency in the first place.

Wilkins (1976) considered notional syllabus analytic because language teaching is organized in terms of content rather than form and there is no ordered exposure to the grammar. He argued that it is superior to the grammatical and situational syllabi because it is more communicative and motivating than the first one and because it covers the most important grammatical forms and language functions, not only those that typically occur in certain situations. However, in later reviews this syllabus was reconsidered as synthetic because it involves fragmenting the target language, and, like the structural syllabus, looks on language acquisition as a planned process of input assimilation and assumes that what is taught = what is learnt (Long & Crookes 1993; Prabhu 1984).

The merit of this approach is that “it emphasizes the fact that the students and their communicative purposes are at the very core of the teaching program. The learner’s actual and foreseeable academic, social, and vocational needs will underlie all aspects of the program’s linguistic and cultural content” (Finnochiaro & Brumfit 1983, p. 17). Although showing some improvements, the notional-functional syllabus still faces the problems of synthetic syllabi. One of them is that its preparation involves presenting one notion or function at a time, whereas functions actually co-occur in discourse (Crombie 1985; Widdowson 1978b). This type of syllabus has also been criticized for lacking a sound psychological basis, paying no attention to SLA theory and for being based purely on reasoning, not on empirical evidence (Brumfit 1981; Paulston 1981). There are also some practical problems with the syllabus design caused by the fact that many individual notions and functions are difficult to define or distinguish.

1.3.6 Methods supporting synthetic syllabi

The structural syllabus serves as a basis for different methods, such as grammar translation, the audiolingual method, Total Physical Response (Asher 1969), the Silent Way (Gattegno 1963), and the conscious-practical method (Beljaev 1965).

The grammar-translation method is derived from the classical method of teaching Greek and Latin. In grammar-translation classes, students learn grammatical rules and then practice them by doing grammar drills and translating sentences or texts to and from the target language. This method focuses on reading and writing.

In the audiolingual method, on the contrary, emphasis is on the teaching of oral skills (listening and speaking). The classroom activities consist of: 1) new material, both lexical and grammatical, presented in the form of dialogs which are supposed to represent pieces of real

communication, 2) a series of pattern drills in which the structures and vocabulary introduced in the dialog are manipulated until they become unconscious habits for students, and 3) recombination response material in which the students try to apply the newly acquired structure in guided semi-free conversations (Krashen & Terrell 1983). There is no explicit grammar instruction; everything is simply memorized in form. The idea is to practice the particular structure until the students can use it spontaneously. Drills and pattern practice include repetition, inflection (one word in a sentence appears in another form when repeated), replacement (one word is replaced by another) and restatement (the student re-phrases an utterance) (Richards & Rodgers 1986).

In Total Physical Response, students are required to obey the instructor's commands, given in the target language. These commands require a "total physical response", beginning with simple imperatives *Sit down!* and leading to more complex sentences, like *If John ran to the blackboard, run after him*. Asher (1969) maintains that an enormous amount of grammar can be embedded in the input in this way. Students are not forced to talk in the early stages and only start producing (in the form of commands to the teacher) after about 10 hours of input.

The Silent Way method makes extensive use of silence as a teaching technique. The teacher uses silence for multiple purposes: it is used to focus students' attention, to elicit student responses, and to encourage them to correct their own errors. The method emphasizes the autonomy of the learner; the teacher's role is to monitor the students' efforts. Silent Way teachers use specialized teaching materials, for example, color association to help teach pronunciation. The teacher typically introduces one new language structure at a time, and old structures are continuously reviewed and recycled.

One of the most popular methods in teaching L2 Russian is *сознательно-практический метод* which can be translated verbatim as a conscious-practical method. This method is seen as a realization of communicative approach (Antonova & Khristiansen 2003) and it is used by many universities and language schools both in Russia and in a foreign language context (including the Russian course at the University of Barcelona). Since this method was used in the experimental part of the current study, it deserves special attention.

The name of the method was offered in the 1960s by Beljaev (1965) who provided a physiological rationale for this method based on Vygotskij's (1966) theory of activity. It is considered conscious since it supposes learners' awareness and metalinguistic knowledge of language forms. It is practical since it assumes that language practice plays a crucial role in L2 learning. The linguistic rationale for the conscious-practical method is based on works of Scherba (1974) who distinguished three objects of learning: *язык* (language), *речь* (speech), and *речевая*

деятельность (speech activity). He gave priority to consciousness and practical orientation of learning and outlined the role of learners' L1 in the process of second language acquisition. Methodological principles of conscious-practical teaching include: (1) simultaneous acquisition of four language skills (listening, reading, writing and speaking); (2) organization of learning process following the sequence from language knowledge to language skills and abilities; (3) distinction of passive (for comprehension) and active (for production) language forms; (4) taking into account learners' L1; (5) introduction of new language material based on rules, examples and language models; (6) the unit of syllabus design is a sentence.

Course books which support this method usually include three types of exercises: *языковые* (language), *речевые* (speech) and *коммуникативные* (communicative) exercises. Language exercises consist of focus-on-forms activities and they are based on decontextualized practice of determined language forms (e.g. fill in the blanks; make a sentence from a list of words; translation exercises, etc.). Speech exercises suppose controlled communication aimed at expression of certain intentions within a given situation (e.g. agree or disagree with the following statements using a model; make a dialogue using the expressions below). Finally, communicative exercises may have features of focused tasks (Ellis 2003), that is, they are designed to promote use of specific language forms while solving a communication problem and achieving a desired outcome.

1.3.7 Problems with synthetic syllabi

The review of different synthetic syllabus types has shown that they suffer from some generic problems, “most obviously their static, target language, product orientation. Syllabus content is ultimately based on an analysis of the language to be learned” which is conducted on “an idealized native-speaker version of that language” (Long & Crookes 1992, p. 33). The model of language acquisition that underlies the synthetic syllabi assumes that linguistic items are acquired separately, singly, in linear fashion, and they can be learned prior to and separate from language use, that learners can and will learn what they are taught when they are taught it, and that “if learners are exposed to ready-made target versions of L2 structures, one at a time, then, after enough intensive practice, they will add the new target versions, one at a time, to their growing native-like repertoire” (Long 2015, p. 21). However, SLA research provides overwhelming evidence against all these assumptions. First, there are very few grammatical features or constructions that can be taught in isolation. Second, research findings on

interlanguage development have demonstrated that “accuracy in a given grammatical domain typically progresses in a zigzag fashion, with backslidings, occasional U-shaped behavior, over-suppliance and under-suppliance of target forms. Advances in one area sometimes cause temporary declines in accuracy in another because the increased processing demands created by control of a new feature result in diminished attentional resources being available elsewhere during production” (Long 2015, p. 22).

Among other problems with synthetic syllabi are weak sequencing criteria, difficulties in definitions and the lack of “borders” of some units (notions, functions, situations and topics), negative influence on learners’ motivation and low possibilities for communication in the classroom in case of structures as units of the syllabus design. However, synthetic syllabi have some advantages. They are easily organized and supplied with numerous published materials, they have well-defined evaluation system and they may help learners to notice the gap in their interlanguage by drawing their attention to specific formal features.

1.4 Analytic syllabi

1.4.1 Communicative Language Teaching

Communicative Language Teaching (CLT) has been widely used in teaching foreign languages for the last forty years. It was proposed and introduced into L2 classrooms in the late 1970s as an alternative to the previous L2 teaching methods, mainly in opposition to the audiolingual method (Johnson 1979, 1982). Whereas these earlier methods were based on a view of language as a set of linguistic systems (phonological, lexical, and grammatical), CLT drew on a functional model of language (Halliday 1973) and a theory of communicative competence (Hymes 1966) that included grammatical, sociolinguistic, discourse and strategic competence (Canale & Swain 1980). The ability to use language in real communication has been given priority over an accurate modeling of the target language patterns (Nizegorodcew 2007). One of the main theorists of the communicative approach, Henry Widdowson (1978a), claimed that instead of teaching the language code, teachers should develop communicative skills. CLT focused on L2 listening comprehension and speaking activities. Input and interaction were seen as crucial for L2 development and played a fundamental role in classroom practice.

Howatt (1984) distinguished a weak and a strong version of CLT. A weak version is based on the assumption that the components of communicative competence can be identified

and systematically taught (Ellis 2003). It is represented by notional-functional syllabus (Wilkins 1976) and task-supported language teaching which combines task use with traditional pedagogy. In contrast, a strong version of CLT claims that language is acquired through communication. As Ellis (2003, p. 28) puts it, “learners do not first acquire language as a structural system and then learn how to use this system in communication but rather actually discover the system itself in the process of learning how to communicate.” A strong version of CLT is reflected in Krashen and Terrell’s Natural Approach (1983) and in proposals for syllabi in which task is a central unit (Candlin 1987; Long 1989; Prabhu 1984) that will be discussed in the following sections.

1.4.2 Natural Approach

The Natural Approach (NA) was offered by S. Krashen and T. Terrell (1983) who claimed that this approach was strongly supported by the latest research in first and second language acquisition and was successfully tested and used by many classroom teachers in different circumstances. In particular, they reference the report of professor W. Voge who presented the results of comparison of two groups of college students learning German in 1981. One group was taught with the NA and in another group the grammar approach was used. It was reported that NA students, after one semester, outperformed controls on tests of speaking and writing. On grammar test two groups performed equally, which allowed Krashen and Terrell (1983, p. 2) to claim that “we do not necessarily sacrifice accuracy for fluency.”

The Natural Approach found its place in a row of other communicative approaches and methodologies of that time (for example, the Direct Method and Asher’s Total Physical Response). All these methods were characterized by the exceptional use of the target language in the classroom from the very beginning of learning. The Natural Approach and the Direct Method differed in that the latter laid more emphasis on teacher monologues, formal questions and answers, and error correction, while the NA focused on providing students with comprehensible input. The Natural Approach, like TPR, was regarded as a comprehension-based approach but, unlike Asher’s method, it focused on exposure to input instead of grammar practice, and on emotional preparedness for acquisition to take place. The advantage of the Natural Approach, in comparison with other approaches, was its adaptability to many teaching contexts for students of all ages and flexibility with regard to different teaching techniques.

The basic principles of the NA were formulated by Krashen in his previous works (1981, 1982). The first principle states that comprehension (listening and reading) precedes production

(writing or speaking). Secondly, production is allowed to emerge in stages (from nonverbal communication and a response with a single word to sentences and more complex discourse). Although grammatical accuracy is very low in early stages, it will increase on its own, without a special focus on form. The students are not forced to speak before they are ready³. The improvement comes “from supplying communicative and comprehensible input and not from forcing and correcting production” (Krashen 1982, p. 6–7). Thirdly, the course syllabus consists of communicative goals. The focus of each classroom activity is organized by topic, not grammatical structure. Fourthly, all activities focus on topics which are interesting and relevant to the students. The teacher must create a special atmosphere in the classroom that will help learners to have low anxiety level, good rapport with the teacher and a friendly relationship with other students.

In order to support these principles, Krashen presents five hypotheses that serve as a theoretical base of the NA:

1. The Input Hypothesis, central to the the NA, claims that “listening comprehension and reading are of primary importance in the language program, and that the ability to speak (or write) fluently in a second language will come on its own with time” (Krashen & Terrell 1983, p. 32). Spoken fluency emerges gradually and is not taught directly. An acquirer can move from a stage i (where i is the acquirer’s level of competence) to a stage $i + 1$ (where $i + 1$ is the stage immediately following i along some natural order) by understanding language containing $i + 1$. In other words, we acquire by understanding language that contains structure a bit beyond our current level of competence ($i + 1$). It can be achieved with the help of context and extra-linguistic information (e.g. visual aids).

2. The Acquisition-Learning Hypothesis claims that “adults have two distinct ways of developing competence in a second language” via acquisition (using language for real communication) and learning (formal knowledge of language). Traditional grammar-based methods (audiolingual or grammar-translation) are designed primarily for learning, whereas the NA aims at acquisition activities.

3. The Natural Order Hypothesis states that certain structures tend to be acquired early or to be acquired late. It is also possible that structures can be acquired in groups, i.e. several at about the same time. Different studies (Dulay & Burt 1974; Dulay, Burt, & Krashen 1982)

³ The methodological implementation of this principle is that the teacher speaks the target language exclusively, whereas the students may answer in their mother tongue or in the target language. They are free to use their L1 until they feel ready to try speaking the L2.

proved that subjects (children and adults) who speak different first languages showed a remarkably similar difficulty order for various forms and rules of L2 English⁴.

4. The Monitor Hypothesis states that conscious learning can only be used as a Monitor, or an editor, that is, it has the function of checking and making repairs on the output of the acquired system. It is not used to initiate production in a second language.

5. The Affective Filter Hypothesis states that certain affective variables (motivation, self-confidence and anxiety level) are directly related to second language achievement. Affective filter is the “part of the internal processing system that subconsciously screens incoming language based on the learner’s motives, needs, attitudes, and emotional states” (Dulay, Burt, & Krashen 1982, p. 46).

This approach to second language acquisition determines Krashen’s view on the role of grammar and vocabulary teaching, methods, types of classroom activities and curriculum design. In the NA grammar teaching has a very limited role. Only certain rules should be taught; for most learners it will be only the late-acquired simple rules. “Grammar use should be restricted to situations where it will not interfere with communication. We should not expect our students to be concerned with fine points of grammar while they are speaking in free conversation; rather, the time to use the Monitor is in writing and in prepared speech” (Krashen & Terrell 1983, p. 57). Krashen believes that in the long run students will speak with more grammatical accuracy if the initial emphasis is on communication skills.

Vocabulary, on the contrary, is basic for communication, and a lot of attention is paid to vocabulary acquisition. The traditional way of memorizing new lexical items is rejected: “New words are acquired when they are heard in an utterance or in a sentence that is comprehensible. Memorized or drilled vocabulary does not stick; words learnt by rote or drill do not enter permanent memory storage. True vocabulary acquisition with long-term retention occurs only with meaningful exposure in situations in which real communication takes place” (Krashen & Terrell 1983, p. 156). To achieve this, students should listen to input which is so highly contextualized that they can focus on key lexical items and interpret the general meaning of phrases or text.

The curriculum is designed to develop basic personal communication skills, oral and written, such as listening to announcements and requesting information in public places, reading and writing notes to friends, reading signs, including instructions, etc. The goals of the course are based on an assessment of student needs and are expressed in terms of situations, functions and

⁴ The order is the following: 1) -ing (progressive), plural, copula (to be); 2) auxiliary (progressive), article; 3) irregular past; 4) regular past, III singular (-s), possessive ('s).

topics: for example, a request for information (the function of interaction) in a hotel (a situation) to obtain lodging (the topic of communication).

As for methods and classroom activities, at the very beginning Krashen and Terrell (1983) recommend using Total Physical Response along with visual support that will provide students with comprehensible input without forcing them to speak. For example, the instructor can give simple commands like *Stand up!* and then expand them with the target vocabulary (*Touch your right foot with your left hand*). Later, at production stage, a lesson consists of a series of acquisition activities that focus on a particular topic and/or situation. Through these activities the instructor introduces new vocabulary, provides comprehensible input and creates opportunities for student oral production.

Although the authors of the NA claimed that it was based on a coherent theory of SLA and that it had numerous successful outcomes, the approach has been severely criticized. Gregg (1984, p. 94) presented a detailed analysis of Krashen's five hypotheses and showed that each of them is "marked by serious flaws: undefined or ill-defined terms, unmotivated constructs, lack of empirical content and lack of explanatory power." In his opinion, Krashen's theory is based on dogmatic statements without evidential support. Its inconsistency was also proven by numerous studies on immersion programs (see the following Section 1.4.3). Children and adolescents in these programs receive large "doses" of L2 input in a communicative setting over an extended period of time. However, they developed only limited L2 proficiency in the areas of speaking and writing, making numerous errors in their productive use of French (Swain 1985). This supports the position that exposure to comprehensible input is not enough for successful second language acquisition.

One of the key implications of the NA which claims that the learner's language system would automatically develop without language-focused instruction was rejected by Long (1983, 1985, 1988) who has demonstrated that instruction does have an effect, but this effect is indirect and non-immediate. Instructed learners make faster progress than uninstructed learners and reach higher levels of ultimate attainment but they do it in their own way, following their own developmental sequence rather than a sequence imposed by a teacher.

1.4.3 Immersion programs

The term "immersion education" was adopted in the 1960s to describe programs in the context of English-speaking children being taught entirely through the medium of French in

selected schools in Quebec. Very soon the immersion approach was extended and adopted to different learning contexts, purposes and languages. Swain and Johnson (1997) distinguish four contexts when this approach can be applied:

- immersion programs in which the medium of instruction is a foreign language with the purpose of achieving higher levels of target language proficiency (e.g. English immersion in Hungary, French immersion in Australia)
- immersion for majority students in a minority language to enhance second language learning (e.g. for Anglophone students in Quebec)
- immersion for language support and language revival (e.g. to promote Catalan and Basque among Spanish-speaking children in Spain)
- immersion in a language of power to facilitate English-medium education (for example, in Hong Kong, Singapore and South Africa)

The first immersion programs were inspired by the unsatisfactory results obtained in the teaching of French as a second language in Canadian schools. To improve the situation, a group of parents in Quebec proposed a program in which, from the first day of school in kindergarten, their unilingual English-speaking children would be instructed entirely in French. In later grades some subjects were also introduced in English so that by grade 6 about half the curriculum was taught in English and half in French. This program, known as early total French immersion, is the model most typically introduced. However, French immersion programs exist in a variety of forms and differ in the educational level at which the L2 is introduced (early, mid- and late immersion) and the extent to which French is used (total or partial). In early partial immersion programs, approximately half the school day is taught in English and half in French from grade 1. Mid-immersion programs start at grade 4 or 5, and late immersion begins at grade 6 or 7. Prior to entering a mid- or late immersion program, students usually learn French as a second language for at least a year.

In spite of some differences, all immersion programs have core features that distinguish them from other educational programs (Swain & Johnson 1997, pp. 6–8):

1. The L2 is a medium of instruction. This feature differentiates immersion from contexts where L2 is taught formally and only as a subject.
2. The immersion curriculum parallels the local L1 curriculum. It consists of content subjects such as mathematics, science, history taught through the L2 and is defined in terms of L1 speakers' needs, goals and educational norms.
3. Overt support exists for the L1. The students' L1 is taught as a subject at some stage and can be used as a medium of instruction.

4. The program aims for additive bilingualism – the learning of an L2 while developing and maintaining students' L1. By the end of the program, L1 proficiency should be comparable to the proficiency of those who have studied through the L1. In addition, a high, though not native, level of proficiency in the L2 is expected.
5. Exposure to the L2 is largely confined to the classroom with little or no exposure to the L2 outside the classroom.
6. Students enter with similar (and limited) levels of L2 proficiency.
7. The teachers are bilingual in the students' L1 and L2.
8. The classroom culture is that of the local L1 community.

Taking these features into account helps to avoid overextension of the term *immersion* which is sometimes used incorrectly, for example, in relation to language learners who go to live and study in the target community, thus immersing themselves in the target language and culture.

Numerous studies have been done with the participation of immersion students that contributed significantly to SLA research (Hammerly 1987; Harley 1988, 1989, 1992, 1993; Harley & Swain 1985; Kowal & Swain 1997; Lapkin 1984; Lapkin & Swain 1990; Swain 1985, 1991, 1993, 1995, 1998; Swain & Johnson 1997; Swain & Lapkin 1986, 1989). It has been found that although immersion students can reach native-speaker levels in listening and reading, their spoken and written French is, in many ways, non-native-like. Harley and Swain (1985) found that even after years of content-based instruction in French immersion classes, students' language was still plagued by error and, with certain types of verb morphology (e.g. the conditional tense), their performance was less than 50% accurate. Similarly, Swain and Lapkin (1989) found that native-like accuracy was not achieved even after more than 5000 hours of meaning-based instruction.

Different explanations of this low level of accuracy in production were given. One of the reasons is that a focus only on meaning “provides insufficient input of certain forms and no means by which to encourage practice of others” (Leeman *et al.* 1995, p. 218). Swain and Lapkin (1989) call it “functionally restricted input.” They counted, for example, the frequency with which different verb forms were used by the teachers. On average, over three quarters of the verbs used were in the present or imperative. This means that other verbal forms do not occur with sufficient frequency in the classroom settings and learners fail to use them correctly. Moreover, focus on meaning makes correct language use less important (Kowal & Swain 1997). Other reasons for low achievements in accuracy are inconsistent and irregular correction of errors and limited possibilities for learners' output (Swain & Lapkin 1989).

In order to solve the problem of limited output in immersion lessons, Swain and Lapkin (1989, p. 156) claim that learners should be “pushed toward the delivery of a message that is not only conveyed, but that is conveyed accurately, coherently, and appropriately.” The researchers offer possible solutions for the problem of functionally restricted input. To make input linguistically more completed, a teacher should recognize what forms are not present (by tape recording oneself over a period of time in class and analyzing one’s own language use) and develop activities that will involve the use of the missing forms and functions. Finally, Swain and Lapkin (1989, p. 153) propose the need for a carefully planned integration of content and language instruction: “Grammar should not be taught in isolation from content. But then, neither should content be taught without regard to the language involved. A carefully planned integration of language and content, however, holds considerable promise.” Swain (2000) in her review of the French immersion studies claimed that there is value in focusing on form through the use of pre-planned curriculum materials.

1.4.4 Problems with purely meaning-focused syllabi

It has been admitted that a purely analytic approach has a number of advantages if compared with synthetic syllabi. Students are exposed to richer input and more realistic language models. Analytic lessons “can be more interesting, motivation maintained, and attention held, as teachers and students are free to use the L2 to communicate about topics of interest – potentially, topics of relevance to meeting communicative needs beyond the classroom” (Long 2015, p. 25).

The essential claim which underlies the analytic syllabi is that “people of all ages learn languages best, inside or outside a classroom, not by treating the languages as an object of study, but by experiencing them as a medium of communication” (Long & Robinson 1998, p. 18). Re-creating for older learners conditions similar to child language acquisition at first glance seemed to be a reasonable proposal since child L1 acquisition is overwhelmingly successful. However, there is an increasing amount of evidence that adult SLA is maturationally constrained (DeKeyser 2000) and that “older learners no longer have the same capacity as young children to attain native norms in a new language simply from exposure to its use” (Long & Robinson 1998, p. 20). As Cutler (2001) defined it, adults are partially “disabled” language learners because they have learned their L1 so well and because they have a weaker capacity for implicit learning due to age-related declines in the efficiency of instance learning.

Another problem is that there are “fragile” L2 features which may not even be noticed in the input for a long time and, consequently, cannot be learned with positive evidence alone (Long 2015; Long & Robinson 1998). These linguistic features tend to be of low perceptual saliency due to their being infrequent, irregular, semantically empty, and communicatively redundant, and/or because they involve complex forms, meanings, or form-meaning mappings (DeKeyser 2005), for example, inflectional morphology used to mark gender, case, agreement, tense and aspect.

The third problem with purely analytic approaches is “their inefficient treatment of learners’ persistent grammatical errors and their inadequacy for older learners, whose reduced capacity for purely incidental learning makes supplementary opportunities for intentional learning necessary” (Long 2015, p. 12). Adult acquirers may become fluent, but not native-like, speakers, particularly with respect to grammatical competence (Swain 1991). Finally, a purely analytic approach ignores the substantial evidence that L2 instruction works and speeds up acquisition and can improve the level of ultimate L2 attainment in some areas (Long 1983, 1988).

1.5 Analytic task-based syllabi

Analytic syllabi that adopt a task as a unit of analysis share the same basic characteristic: “Rather than focusing on the explicit, discrete knowledge of specific aspects of language in isolation and building up proficiency in an incremental, step-by-step fashion, task-based work aims to feed the learner’s internal syllabus while engaging the learner in holistic communicative use of language” (Van den Branden 2013, p. 629). However, they differ in the ways a needs analysis is conducted to determine syllabus content, in how tasks are selected and sequenced, and in the methodological options, such as group work and focus on form (Long & Crookes 1993).

1.5.1 The Procedural Syllabus

The procedural syllabus is associated with the work of Prabhu on the Bangalore Communicational Teaching Project (Prabhu 1984, 1987) in India from 1979 to 1984. The project

was carried out in eight classrooms with 390 children aged between 8 and 13 in their first and fourth year of learning English, for periods of one to three years.

The innovative idea of the project that made it different from the previous Communicative Language Teaching was replacing “teaching *for* communication” by “teaching *through* communication.” Prabhu (1987) claimed that all language structures can be taught and learned through activities involving language use. Like Krashen (1982, 1985), he argued that language form is acquired subconsciously when the learner’s attention is focused on meaning and, therefore, implicit learning through meaningful practice is more effective than explicit learning.

Prabhu rejected the idea of pre-selection of language items, their isolated practice and sentence production and made a task the basis of each lesson. Prabhu (1987, p. 24) defined task as “an activity which requires learners to arrive at an outcome from given information through some process of thought, and which allowed teachers to control and regulate that process.” Tasks were sequenced according to a commonsense judgment of increasing complexity, the later tasks being either inclusive of the earlier ones or involving a larger amount of information, or an extension of the kind of reasoning done earlier.

Work on tasks usually consisted of two parts. First, the teacher used “pre-task” in a whole-class format in order to present and demonstrate the task, to assess its difficulty for the class, to modify it accordingly if necessary, and to give the learners relevant examples of language use. Second, the task itself was for the pupils to work on, usually individually. The work was followed by the teacher’s feedback on task accomplishment. A lesson was considered successful, if “it had a task-centered pattern and the task set seemed to engage most pupils’ minds, i.e. the task was perceived clearly and attempted seriously, regardless what measure of success was actually achieved” (Prabhu 1987, p. 43).

Prabhu distinguished three types of tasks: opinion-gap, information-gap and reasoning-gap (see Section 2.5). The examples of tasks provided in Prabhu (1987) include calculating distances and planning itineraries using maps and charts, assessing applicants for a job on the basis of biographical sketches, and answering comprehension questions about dialogues. These tasks are typical for CLT but they are distinct from tasks in TBLT as they don’t necessarily involve activities students will ever need to do in English outside the classroom (although they may be useful for language learning). The major difference of the Bangalore project from the previous CLT practice was not in the tasks themselves, but in the accompanying pedagogic focus on task completion instead of on the language used in the process. Long and Crookes (1993) pointed out two innovations of the project concerned the kind of input to which learners were

exposed and the absence of systematic corrective feedback. With respect to input, teacher speech was not preselected or structurally graded, but “roughly tuned” as a natural by-product of the spontaneous adjustments made to communicate with less proficient speakers inside or outside classrooms (Prabhu 1987, p. 57–59). As far as error correction is concerned, ungrammatical learner utterances were accepted for their content, although they may have been reformulated by the teacher.

Despite being an interesting, innovative program, the Bangalore project has been criticized on a variety of grounds, one of the chief complaints being its failure to build an evaluation component into the design (Brumfit 1984; Long & Crookes 1993). In an attempt to remedy this and to somehow evaluate the effectiveness of the project, Beretta and Davies (1985) compared four experimental classes and four control classes in the same schools that had been taught using the traditional structurally-based program. The researchers faced two problems: the problem of validity and designing tests that would be equally fair to both teaching methods. The most serious threat to internal validity was that in one of the schools, one group had maintained its stability over a period of time while the other had not and a third of the students in one group were not available for the tests. As for external validity, three of the four experimental groups were taught by better qualified, more highly motivated teachers. Beretta and Davies (1985) recognized that these problems with validity could influence the test results.

A battery of tests used for evaluation was intended to measure achievement separately for experimental and control groups (by a structure test and a task-based test) and proficiency. The following five tests were used: (1) multiple-choice test; (2) fill in the blanks; (3) dictation; (4) listening/reading comprehension (the learners were required to read, for example, a hotel advertisement and to write answers to spoken questions); (5) task-based test that was a representative sample of the tasks used in the classroom (for example, solving problems related to a timetable and to a calendar).

It was found that the four control groups outperformed experimental groups on a structure test, whereas three out of four experimental groups did statistically significantly better on tests of listening and reading comprehension, and all four experimental groups performed better on a task-based test than control groups as was expected. While admitting the limitations of the evaluation process (particularly the impossibility of full experimental control and, consequently, any generalization, as well as the problems with internal and external validity) Beretta and Davies (1985, p. 126) regarded the results of the Bangalore Project “as being, on the whole, positive.”

Apart from the weak evaluation component of the syllabus design, Long and Crookes (1993) pointed out three more problems with the Bangalore project related to task selection, task sequencing and absence of regular feedback and focus on form. First, they argued that as no needs identification was carried out, no rationale exists for the content of such a syllabus, that is, for task selection. Second, they called the criteria for evaluating task complexity and sequencing tasks unsatisfactory since both appear to be arbitrary processes left partly to real-time impressionistic judgments by the classroom teacher. Finally, there are empirical findings (Carroll & Swain 1993; Lightbown & Spada 1990; Tomasello & Herron 1988) which support the need for corrective feedback and the need for focus on form in language teaching excluded by Prabhu. In spite of all its shortcomings, the Bangalore project is considered an important contribution to TBLT research because it is one of the few examples of a systematic attempt to develop and explore task-based procedures in real classroom contexts (Samuda & Bygate 2008).

1.5.2 The Process Syllabus

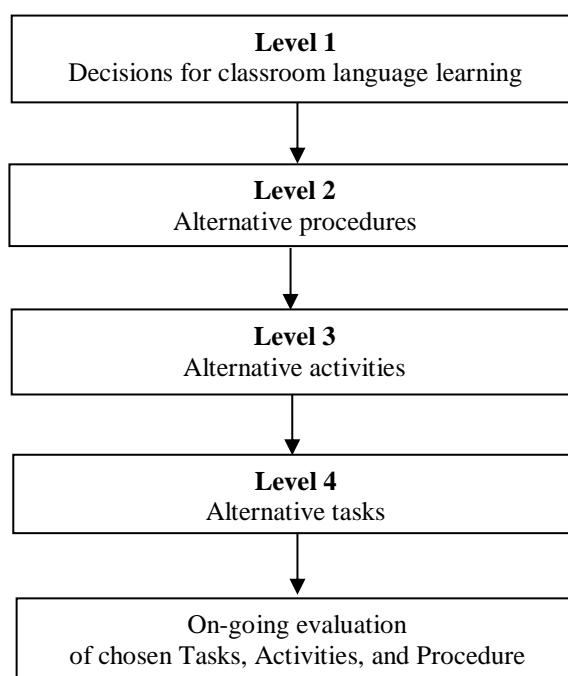
Another task-based approach to course design is the process syllabus (Breen 1984, 1987a, b; Breen & Candlin 1980; Candlin 1984, 1987). Breen and Candlin's focus is the learner and learning processes and preferences, not the language or language learning processes. Breen (1984) argued that every teacher inevitably interprets and reconstructs a predesigned syllabus as well as that learners create individual learning syllabi. As a result, in the lesson-to-lesson reality of language teaching we deal with the real syllabus which is the synthesis of teacher's and learners' interpretation of a predesigned syllabus. Breen (1984, p. 50) concludes that a good syllabus is one "which is positively amenable to alternative interpretations and open to reconstruction through interaction in the classroom."

An important characteristic of the process syllabus is that the syllabus designer no longer pre-selects learning content, but provides "a plan for the gradual creation of the real syllabus of the classroom, jointly and explicitly undertaken by teacher and learners" (Breen 1984, p. 52). Breen (1984, 1987a) and Candlin (1984) suggest a retrospective syllabus that can only be described after the course is over, with process-syllabus designers in general aiming to provide a framework for learning which "deliberately engages reinterpretation and which explicitly addresses teacher and learner capacities to select, subdivide and sequence subject matter for language learning which they perceive as most valuable to them" (Breen 1987a, p. 166). Division of language according to lexis and grammar is not excluded, since learners can choose

to concentrate upon these aspects if they wish, especially as the result of appraisal of a particular learning experience, but this is now the result of social constructs, produced interdependently in classrooms by teachers and learners.

Breen (1984) proposes a hierarchical model of the process syllabus which is made up of four levels, and in which each level entails the level below it (see Figure 2). At the highest level are decisions for classroom language learning relating to participation, procedure and subject-matter (Who does what with whom, in what content, with what resources, when, how and why?). The particular alternatives chosen at the second level would become the agreed procedures for the particular classroom group. The third level indexes alternative activities to be selected on the basis of appropriateness to decisions at Level 1. The fourth, and final level is a bank of pedagogic tasks students may select from to do the activities. The process syllabus is therefore a framework for decisions and alternative procedures, activities, and tasks for the classroom group.

Figure 2. The levels of the process syllabus (Breen 1984, p. 57)



Breen (1987a) proposes a number of justifications for use of the process syllabus in language learning. First, the process syllabus facilitates the synthesis of the teacher's and learners' proposals and provides a means of relating content and methodology. Second, it requires reinterpretation during the learning process, and is therefore flexible, allowing for emerging changes in needs. Finally, classroom decision-making is of utmost priority and it is

seen as an authentic communicative activity in itself. Breen argues that meta-communication and shared decision-making are necessary for language learning.

However, the process syllabus was criticized on various grounds. White (1988) pointed to (1) a lack of formal field evaluation; (2) an assumed unrealistically high level of competence in teachers and learners; (3) an implied redefinition of role relationships and a redistribution of power and authority in the classroom that would be too radical and/or culturally unacceptable in some societies; (4) emphasis on process and procedure rather than on outcome, possibly resulting in an aimless journey; (5) the need for a wide range of materials and learning resources.

Long and Crookes (1993, p. 35) pointed out more problems with the process syllabus. Like the procedural syllabus, it “deals with pedagogic tasks whose availability (in the task ‘bank’) is not based on any prior needs identification, which raises problems for selection.” A question about task sequencing remains unclear although a variety of possible criteria for grading task difficulty were discussed by Candlin (1987). Finally, no explicit provision is made for a focus on language form.

1.5.3 Task-Based Language Teaching

In the last few decades Task-Based Language Teaching (TBLT) has become one of the most important approaches to language teaching and second language acquisition research especially with English as a second and foreign language. The increasing amount of publications on this topic (Bygate, Skehan, & Swain 2001; Ellis 2003, 2006, 2009, 2012; García Mayo 2007; Long 2000, 2015; Nunan 2004; Samuda & Bygate 2008; Shehadeh & Coombe 2012; Swain & Lapkin 2000, 2001; Robinson 2005, 2007, 2011a, b; Skehan 2001; Skehan & Foster 2001; Van den Branden 2006; Van den Branden, Bygate, & Norris 2009; Willis 2007, among many others) attest to the interest in task-based language learning and teaching. Since 2005 the TBLT International Conference has been organized biennially with the purpose of bringing together researchers and educators all over the world with an interest in tasks. The JALT Task-based Special Interest Group founded in 2010 focuses on issues related to TBLT in the Asian EFL context.

Bygate, Skehan and Swain (2001) point out two approaches to task investigation and two groups who have each appropriated the term “task” for their own purposes. These are communicative language teachers and second language acquisition researchers. The pedagogic approach (Lee 2000; Prabhu 1987; Samuda 2001; Willis 1996) presents the problem as one of

understanding how the behavior of the teacher can be made more effective and how learners can interact with tasks more effectively. The research approach (Long 2015; Skehan 1998a; Robinson 2011a) presents the problem as one of how tasks may be used as a device to uncover the effective engagement of acquisitional processes. Task in this sense turned out to be an effective investigation instrument because changing one or more task design factors can significantly influence learners' oral and written performance.

The principles underlying TBLT approach were developed in Crookes (1986), Ellis (2000, 2003), Long (1985, 1989, 1996, 2000, 2015), Long and Crookes (1992, 1993), Long and Norris (2000), Long and Robinson (1998), Nunan (1989, 1993, 2004), Robinson (1994, 1998, 2005, 2007, 2011a), Skehan (1996a, 1998a, b), among others. A theoretical rationale for TBLT was summarized by Long (2015) in a Cognitive-Interactionist theory of instructed second language acquisition. A number of hypotheses about the nature, conditions and mechanisms of second language acquisition offered by different researchers contributed to it. Among them are:

- Input Hypothesis

Krashen (1981, 1982, 1985) claimed that the exposure to comprehensible input is the essential condition for acquisition. Comprehensibility of the input is ensured by its approximate level of difficulty that is slightly higher than the learner's present proficiency level. Such input was called "roughly tuned input." According to Krashen's model, L2 acquisition will automatically occur with time if a learner is exposed to sufficient comprehensible input (for more details see Section 1.4.2).

- Output Hypothesis

Swain (1985, 1993, 1995), after analyzing the performance of French immersion students in Canada, argued that although comprehensible input may be essential to the acquisition of a second language, it is not enough to ensure that the outcome will be native-like. She advanced the comprehensible output hypothesis and insisted that students must be given adequate opportunities to use the target language in the classroom context. Swain argued that output has four functions. First, it provides the opportunity to test hypotheses, to try out means of expression and see if they work. Second, output provides the opportunity both for meaningful practice of one's linguistic resources and for developing automaticity in their use. Third, output may generate responses from speakers of the language that can give learners the information about the comprehensibility of their utterances. The fourth function of output proposed by Swain is that it may force the learner to move from semantic processing to syntactic processing. By syntactic processing she understands instances where learners move beyond processing words as

independently functioning lexemes and come to consider them in their relationship to other words in the sentence (for example, adjectival agreements, verb-subject agreements, word order).

- Interaction Hypothesis

The Interaction Hypothesis (Long 1996) subsumes aspects of the Input Hypothesis and the Output Hypothesis and considers exposure to language (input), production of language (output), and feedback on production (through interaction) as important constructs of second language learning (Gass & Mackey 2006). This line of research was developed by Gass (2003), Gass and Mackey (2006), Long (1996, 2015), Mackey (1995, 1999, 2012), Mackey and Goo (2007), Mackey and Philp (1998), Pica and Doughty (1985, 1988), among others. The emphasis is on the role of negotiated interaction in the development of a second language. It has been claimed that “conversational interaction in a second language forms the basis for the development of language rather than being only a forum for practice of specific language features” (Gass 2003, p. 234). According to Long (2015, p. 53), when learners run into communicative trouble, they are likely to switch their attention from meaning to form long enough to solve the problem and notice the necessary new information. He also pointed out the role of negative feedback in L2 development: “Negative feedback obtained during negotiation work or elsewhere may be facilitative of L2 development, at least for vocabulary, morphology, and language-specific syntax, and essential for learning certain specifiable L1–L2 contrasts” (Long 1996, p. 414). From this perspective tasks can promote language learning by generating productive forms of communication breakdown, when meaning is negotiated through comprehension checks, clarification requests and confirmation checks (Long 1985).

- The Cognition Hypothesis

Robinson’s Cognition Hypothesis (2001b, 2005, 2007, 2011a, b) provides a theoretical basis for making predictions about how task variables affect learners’ production. Robinson (2011a, p. 14) claimed that in task-based syllabi, “pedagogic tasks should be sequenced solely on the basis of increases in their cognitive complexity.” Increasing complexity is argued to promote more accurate production and more complex utterances.

- Noticing Hypothesis

In the recent SLA research, much emphasis has been placed on the concept of attention and noticing. Schmidt (1990, 1993, 1995) argued that attention is essential for learning. The Noticing Hypothesis states that “what learners notice in input is what becomes intake for learning” (Schmidt 1995, p. 20). By noticing, Schmidt (1995, p. 29) understands “conscious registration of the occurrence of some event.”

Long (1996, 2015) and Skehan (1996b, 1998a) are in broad agreement about the SLA motivation for task-based syllabi. They both cite research showing that there is little resemblance between acquisitional sequences and instructional sequences based on linguistic forms (Ellis 1990) and that learning is non-linear and cumulative (Kellerman 1985), rather than linear and additive as synthetic syllabi imply.

Given Long's and Skehan's broad agreement about the SLA motivation for choice of task-based syllabi, there are some differences of scope and focus in their proposals (Robinson 1998). One methodological principle Long (1989, 1991) advocates for is focus on form (for more details see Chapter III), that is, drawing students' attention to linguistic elements as they arise incidentally in a meaning-oriented activity. Another research priority which arises from the Interaction Hypothesis is to establish which tasks are more likely to generate effective negotiation of meaning.

While Long places great importance on the opportunities for focus on form and negotiation of meaning, Skehan (1998a) takes a more cognitive, information processing approach to task-based instruction and focuses on research aimed at identifying the effects of different factors, which determine task difficulty, on the complexity, accuracy and fluency of learner production (e.g. planning time in Foster and Skehan (1996)), and the influence of learner variables such as aptitude on language processing (Skehan & Foster 1997). These are not oppositional perspectives since there is a substantial amount of cognitivist research into task design and performance within Long's framework.

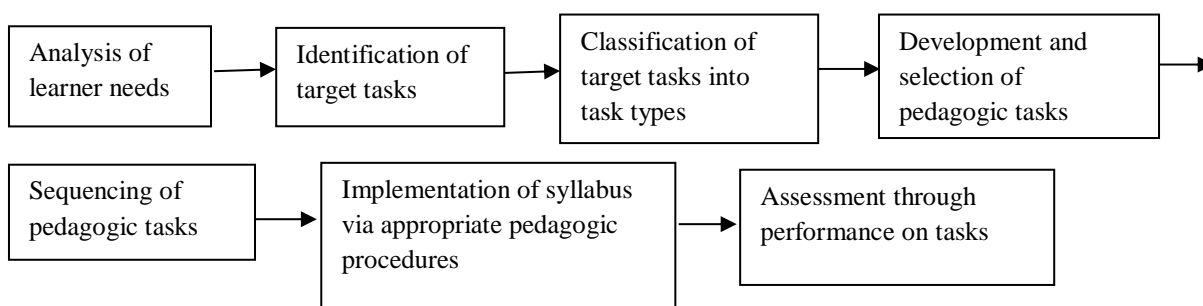
Another approach to TBLT belongs to Ellis (2003, 2009, 2012) who takes a broader view on the theoretical framework of TBLT and questions of task design and implementation. He compared three approaches to TBLT – Long's (1985), Skehan's (1998a) and his own (Ellis 2003) – and showed that the only characteristic that all three approaches share are the emphasis on creating contexts for natural language use and focus on form. They differ in terms of how attention to form is to be achieved, with Long emphasizing corrective feedback, Skehan task-design and pre-task-planning, and Ellis (2012, p. 197) "a variety of ways in the pre-task, main task and post-task phases of a task-based lesson." Major differences in the three approaches involve the role of group work (Ellis does not see it as an essential characteristic), use of focused and unfocused tasks (Skehan and Long favour only unfocused tasks, whereas Ellis also see a role for focused tasks), and view on traditional structural teaching rejected by Long and Skehan and seen by Ellis as complementary to TBLT. The comparison of these approaches is summarized in Table 6.

Table 6. Differences in three versions of TBLT (Ellis 2012, p. 197)

Characteristic	Long (1985)	Skehan (1998)	Ellis (2003)
Natural language use	Yes	Yes	Yes
Learner-centeredness	Yes	Yes	Not necessarily
Focus on form	Yes – through corrective feedback in the main-task phase	Yes – mainly through pre-task	Yes – in all phases of a TBLT lesson
Linguistic support	No	No	Possibly
Types of tasks	Unfocused	Unfocused	Unfocused and focused
Rejection of traditional approaches	Yes	Yes	No

According to Long (1985, 1989, 2015; Long & Crookes 1992, 1993), the process of task-based syllabus design includes a series of steps (see Figure 3). The first step is to conduct a needs analysis (methodology and examples of empirical studies on needs analysis are presented in Long 2005) in order to identify target tasks. Target tasks are classified, at a more general and abstract level, into task types from which are derived pedagogic tasks – the tasks teachers and students actually work on in the classroom. Pedagogic tasks are sequenced according to their intrinsic complexity in order to form the task-based syllabus. This kind of syllabus “would usually, although not exclusively, imply assessment of student learning by way of task-based criterion-referenced tests, whose focus is whether or not students can perform some task to criterion, as established by experts in the field, not their ability to complete discrete-point grammar items” (Long & Crookes 1992, p. 41). Task-based syllabus design as well as other important issues and concepts related to TBLT will be discussed in a more detailed way in Chapter II.

Figure 3. Task-based syllabus design and language teaching (Long & Crookes 1992)



The researchers who advocate the TBLT approach see its potential in its “compatibility with research findings on language learning, its principled approach to content selection, and its attempt to incorporate findings from classroom-centered research when making decisions concerning the design of materials and methodology” (Long & Crookes 1993, p. 41). Robinson

(2011a) argued that tasks provide a context for negotiating and comprehending the meaning of language; they provide opportunities for uptake of corrective feedback, for incorporation of premodified input containing positive evidence and for noticing the gap between a participant's production and the target language. Task demands can focus attention on specific L2 features with consequences for improvements in accuracy of production. Simple task demands can promote access to and automatization of the currently emerged interlanguage that will improve fluency of production. Complex tasks can "prompt learners to attempt more ambitious, complex language to resolve the task demands, with consequences for improved complexity of production" (Robinson 2011a, p. 2). Finally, working with tasks stimulates learners' motivation.

However, Long and Crookes (1993, pp. 42–43) recognized some of the problems with the task-based approach. Among these problems are the difficulty of differentiating tasks, especially tasks and subtasks nested within them (the problem of finiteness); the issue of task difficulty and determination of the relevant sequencing criteria; the lack of a rigorous field evaluation; and a shortage of commercially published task-based materials, especially for languages different from English.

More recent critiques of TBLT have involved issues related to (1) its theoretical and methodological principles, in particular, needs analysis (Long 2015), criteria of task selection (Cook 2000), and the role of negotiation of meaning (Bruton 2002; Foster 1998; Seedhouse 1999, 2005); (2) classroom pedagogy, e.g. putting the task-based syllabus into practice (Littlewood 2007; Samuda & Bygate 2008; Swan 2005); (3) the question of using the task-based approach with beginner learners (Bruton 2002); and (4) its debatable efficiency in introduction of new material (Swan 2005).

It has been claimed that "needs in some classes are too heterogeneous to make needs analysis worthwhile" (Long 2015, p. 94), and that some populations simply have no L2 needs at all, or at least, none that are analyzable or predictable at any useful level, for example, young children and school-age foreign language learners. However, Long (2015, p. 225) argued that in the first case, target task-types are "convenient abstractions to help deal with logistical problems that can arise with short courses and/or heterogeneous groups." In the case of young learners whose needs are difficult or even impossible to predict, other aspects of TBLT, including materials and methodological principles are still relevant and have been shown to be effective.

Cook (2000, p. 156) criticized Long's approach to task selection: "In the real world, people do a whole host of things, both trivial and significant, which one would not want to reproduce in a language class simply for that reason. < ...> In TBLT, however, being "real" is the only one reason for taking target tasks as a starting point". However, the criteria of task

selection offered by Cook himself – to select tasks which are “intrinsically educational, valuable or enjoyable” – seem vague and unclear.

Long’s (1988) claim that negotiation of meaning contributes to noticing and, in some cases, learning of a language form was refuted by some empirical studies. For example, Foster (1998) has demonstrated that learners typically bypass communication problems during tasks with little or no negotiation of form (in contrast to meaning) or language development. Even in the case of a pair-work task with divided information, the oral intervention of some students may be minimal (Bruton 2002), and task-based interaction can constitute a “particularly narrow and restricted variety of communication” (Seedhouse 1999, p. 155).

Bruton (2002) and Swan (2005) argue that TBLT has been imposed on teachers on the basis of limited research findings, generalized from laboratory-based studies in conditions that do not reflect the realities of the average teaching situation worldwide. Similarly, Samuda and Bygate (2008, p. 203) pointed out that practitioners’ concerns include insufficient evidence on how TBLT works in real classroom contexts, and consequently little documented evidence to date of full-scale programme evaluations. In fact, Samuda and Bygate (2008) gave only two examples of large-scale implementation of TBLT. The first one is the Hong Kong Target-Oriented Curriculum, implemented by the Education Department of Hong Kong in 1999 and run in primary and secondary schools. The second example is Flanders task-based programmes for teaching Dutch as a second language introduced by the Flemish government in the 1990s (Van den Branden 2006). In the case of the Hong Kong program, in spite of a long-term investment of time and resources in teacher preparation and support, teachers reported problems in how to respond to the demands of the curriculum and how to handle it on an everyday basis (Candlin 2001; Littlewood 2007). Van den Branden’s collection of reports on the Flanders project “provides a close account of programme development within a particular second language setting, but it has little data about the impact of TBLT on language development over time and there are limited examples of what the finished product looks like” (Samuda & Bygate 2008, p. 211).

Bruton (2002) has questioned task as a unit of syllabus design and argued that there are many instances of communication which cannot be termed tasks (talking to your friend, for example). Another problem he has paid attention to is the question of level. Bruton (2002, p. 287) claims that the implementation of a task-based syllabus at the initial levels can be difficult since “planning of expected language assimilation and significant teacher direction in teacher-fronted activities” are inevitable. Consequently, the application of TBLT approach seems to be

limited to levels where the students have some existing target language oral ability. However, there are examples of task-based courses for absolute beginners (Duran & Ramaut 2006).

Swan (2005, p. 384) has argued that the advocates of TBLT often give the impression of “a paradigm shift in which discredited older methods, which have failed in practice and which are theoretically unsupported, have been replaced by approaches of a diametrically opposed kind based on sound theory and solid research findings,” and this is not the case. However, criticizing the TBLT approach for its unjustified rejection of traditional approaches, Swan (2005, p. 376) himself offers examples of radical and unsupported criticism: “While task-based instruction may successfully develop learners’ command of what is known, it is considerably less effective for the systematic teaching of new language. This is especially so where time is limited and out-of-class exposure unavailable, thus making heavily task-based programs inappropriate for most of the world’s language learners.”

Another limitation of TBLT research that has not been widely discussed in SLA literature is a domination of English as the target language. It seems that it has been taken for granted that results received in empirical studies with L2 English learners can be applied to other languages. However, languages with different and more complex morphological systems such as Russian may impose specific difficulties on syllabus designers and teachers (Gilabert & Castellví, in press).

The perspectives of development of TBLT were outlined by Van den Branden (2013, p. 632): “One of the major challenges TBLT faces today has to do with implementing this promising approach in mainstream classrooms. In addition to the development of task-based curricula and course books, and the development of task-based assessment instruments, what seems to be needed are coherent and intensive teacher training programs.”

1.5.4 Project work

A project is understood as an extended piece of work with a non-linguistic outcome such as making a class magazine or preparing a trip to an unknown planet. Tasks play the role of building blocks for sustaining the project. In this section two approaches to project work (PW) will be reviewed: the one of Legutke and Thomas (1991) and Ribé and Vidal (1993).

Legutke and Thomas (1991, p. 160) define project work as “a theme and task-centered mode of teaching and learning which results from a joint process of negotiation between all participants.” They present the theoretical underpinnings as well as detailed examples of

classroom implementation of PW. Projects which were carried out to teach English in some schools in the UK and German in the 1980s (a Jumble Aid project that consisted of organizing and carrying out an old clothes sale to raise money for a charity and The Catcher in the Rye project to investigate Salinger's novel, among others) are described. The data are based on case studies, informal teacher accounts of project experiences, learner diaries and texts produced as a part of the project work.

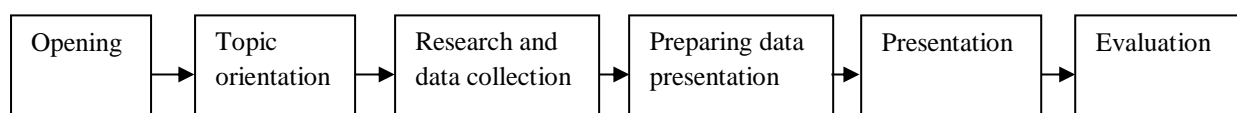
Legutke and Thomas (1991, pp. 158–159) give a list of the fundamental characteristics of project learning. They point out that themes and target tasks for PW derive from “life,” not only from a list of pre-determined curricular items. Learners' involvement and interest in the project are crucial for learning to take place. Project learning is learner-centered; it allows for learner contribution of project ideas and negotiation of topics and tasks and permits learners to discover their specific strengths, interests and talents. The product in PW has the same importance as the learning process. The teacher and the learner take on a multiplicity of roles. The teacher may adopt those of a manager, facilitator, researcher, participant or monitor, and the learner can act as a manager, actor, writer, secretary, researcher, etc.

Legutke and Thomas (1991) view projects as a collection of a large variety of tasks, with different types of task having distinct roles. These roles include developing language skills through controlled practice and focusing attention on problematic language features (language learning tasks), enabling learners to deal with different types of input data and to express their own views and meanings (pre-communicative tasks), acting as a catalyst through which discourse emerges from genuine communicative needs and interaction (communicative tasks), and enhancing learners' managerial and procedural capacities through controlled and guided practice (instrumental and management tasks) (Samuda & Bygate 2008, p. 224–225).

Project work includes six stages, with each stage having its objectives (see Figure 4). For example, the objectives of the opening stage are to develop a positive group dynamic, to introduce learners to a communicative approach, and to introduce textual and multimedia data for research activities. The topic orientation stage aims at sensitizing learners towards the theme and arousing their curiosity, mobilizing existing knowledge and exchanging personal experiences, etc. The evaluation stage is the most questionable. Legutke and Thomas (1991) admit that assessment is clearly an area for further research. Discussing possible criteria of evaluation, the researchers use the model of communicative competence offered by Canale (Canale & Swain 1980) that consists of four categories: grammatical competence, socio-linguistic competence (decisions about appropriacy of language in context), discourse competence (ability to decode and encode coherent written and spoken text) and strategic

competence (ability to use language to communicate intended meaning and to repair breakdowns in communication). All four categories must be assessed, whereas most public examinations are strongly biased towards testing only grammatical competence. Legutke and Thomas propose CCSE (Certificate in Communicative Skills in English) examinations as more appropriate since they are characterized by the clarity of tasks and detailed degrees of skills required. Additional forms of assessment can take place, such as writing course reports or presenting a product (a book, display or a film) and explaining its provenance.

Figure 4. Steps in project work (Legutke & Thomas 1991)



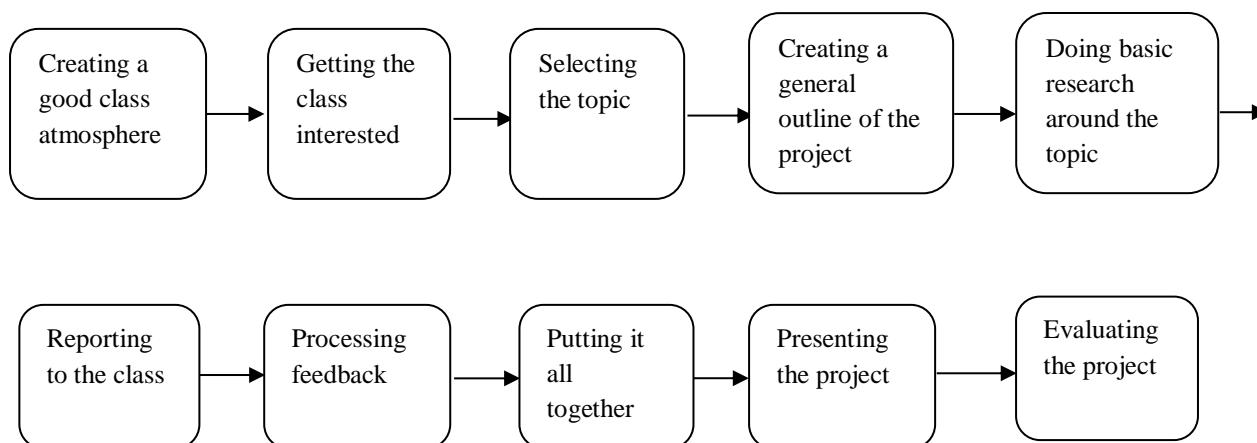
Ribé and Vidal's (1993) proposal of project work is quite similar but has some differences related to task types, functions, and evaluation. They distinguish three generations of tasks:

1. First generation tasks aim primarily to develop communicative ability and are designed for the practice of a specific function/structure or for problem-solving. They depend on the content and order dictated by the syllabus. Strictly speaking, these are not tasks, at least not in Long's (2015) or Skehan's (1998a) view.
2. Second generation tasks focus primarily on content and aim at "developing not only language skills, but also general cognitive strategies of handling and organizing information. Language becomes a vehicle for doing a real piece of work" (Ribé & Vidal 1993, p. 3).
3. Third generation tasks share the characteristics of second generation tasks but have a wider scope. They aim at developing the personality of the student through the experience of learning a foreign language. These tasks have the following characteristics: they are learner-centered, open and flexible; they incorporate the students' previous knowledge and personal experience; the scope and length of the task can be quite extensive (from a few weeks up to a whole year); the thematic content is related to the students' immediate environment and interests; language is approached globally, not sequentially (Ribé & Vidal 1993, p. 4).

Project work involves second and third generation tasks. It is seen as a large task (macro-task) composed of smaller steps (micro-tasks). These steps, as well as time-span and a final

product, are pre-planned in the teacher's mind, but not totally pre-determined. They are also class-generated. The sequence of steps in a typical project is presented in Figure 5.

Figure 5. Steps in project work (Ribé & Vidal 1993)



Ribé and Vidal (1993) provide comments, recommendations and examples for each step. From the moment when the topic is selected to the final presentation of the project, the work is organized in the following way. The teacher and students decide on “an initial structure” of the project, that is, which aspects and areas to include. Then the groups of 3–5 students are formed and areas of work are assigned to them. Each group works on their chosen or allocated sub-theme. Learners, with the teacher's help, should identify needs related both to the content (make a list of sources of information, define the principles of its selection and organization) and the language (a range of relevant vocabulary and structures). Reporting from group to class or from group to group happens throughout the project. This stage is followed by oral and written feedback from the teacher and learners themselves.

The teacher's role in project work is quite different from their role in traditional teacher-centered lessons. Amongst many other things, the teacher is the initiator, planner/co-planner, provider of thematic and language input, provider of language explanations and correction, promoter of group co-operation and an appropriate atmosphere, helper, consultant and evaluator/co-evaluator.

Tasks in project work should provide three basic conditions so that language learning can take place. These are comprehensible input, comprehensible output and content and language processing. Tasks should also create new language needs to avoid using only the language learners already know. In order to provide input relevant to the information and language needs of the project, the teacher is expected to direct the students towards useful resources, such as encyclopedias, magazines, surveys, etc. Content processing requires a mass of information to be

processed and a criterion for analyzing it and extracting the necessary aspects. Language processing involves a series of sub-skills, such as recognizing elements of the language, formulating hypotheses about the content of the message and then verifying them. As for output, it will not necessarily be accurate in the beginning. Many language features may be internalized directly from input materials. However, some language areas will need specific treatment in the classroom. Ribé and Vidal (1993) accept traditional grammar exercises or teacher explanations as a possible way of handling language problems.

The assessment should include not just linguistic accuracy, but all aspects of the project (content, language, presentation, etc.) and also non-linguistic factors such as effort, objectives, attitude, cognitive development, etc. The authors distinguish formative assessment which happens throughout the project and summative assessment which is a final grading step upon the project's completion. Individual work, group work and class work can be the object of assessment. For example, group work can be analyzed by the members of the group, by the class or by the teacher. Ribé and Vidal's proposal for formative assessment includes questionnaires, letters and personal interviews. Summative assessment can take the form of traditional language tests (multiple choice, gap-filling, etc.) if required or can-do questionnaires where learners and the teacher evaluate general language and communicative skills.

Ribé and Vidal (1993, p. 4) argue that PW is appropriate for high school students as they usually don't have clear motivation to learn a foreign language and this work creates it "through involvement with the topic and through the processes that encourage personal creativity." It allows for the possibility to extend language exposure and promotes interaction and negotiation of meaning that can facilitate language acquisition.

However, PW puts a level of responsibility on learners that not all of them are ready to assume. It can cause some difficulties both for teacher and learners. First of all, in order for the whole process to be successful, continuity and stability are needed, which entails students' regular attendance, and this is not always possible. Second, PW supposes a lot of work outside the classroom. Learners usually have to investigate individual questions and prepare work in advance, then during the lesson they share, interact, and receive feedback. If someone comes to the lesson unprepared, he/she will be just a passive listener, and the general progress on the project will slow down. Third, the concept of PW relies heavily on learner creativity. Some learners are shy; some of them will not be interested enough in the topic, which can lead to general disappointment and frustration. For the teacher it is hard to control that everyone participates in the same way and to evaluate individual participation in the group work in the end. Finally, learners should have some previous knowledge of the language (the materials

offered by Ribé and Vidal (1993) are addressed to intermediate learners) because to learn a new language from zero through project work seems extremely challenging. Taking into account all these limitations, it can be assumed that project work is likely to be efficient only in special conditions (regular attendance, learners' interest and readiness to participate, sufficient language level to fulfill all the tasks, among others).

1.5.5 CLIL

Content and Language Integrated Learning (CLIL) is a dual-focused educational approach in which an additional language is used for the learning and teaching of both content and language with the objective of promoting both content and language mastery to pre-defined levels (Marsh & Wolff 2007). In other words, CLIL suggests both learning a curricular subject such as physics or geography through the medium of a non-native language and learning a non-native language by studying a curricular subject (Dalton-Puffer 2011; Mehisto *et al.* 2008).

CLIL has precedents in immersion programmes (North America), education through a minority or a national language (Spain, Wales, France), and many variations on education through a foreign language. It has become the umbrella term for different programs and educational contexts, such as content-based language teaching (Brinton, Snow, & Wesche 1989) and bilingual education (Cummins 1984). Currently, a number of international projects are carried out in order to investigate the results of applying CLIL in secondary schools, and to develop teacher training programs. Among these projects are TIE - CLIL (Translanguage in Europe - Content and Language Integrated Learning: www.tieclil.org), ITALIC (Interacting for Teaching and Learning in CLIL: www.abdn.ac.uk/italic/), and e-CLILT (e-based Content and Language Integrated Learning Training: <http://www.eclilt.net/index.html>).

According to Coyle's (2008) 4Cs curriculum, a CLIL lesson should combine four elements:

- Content

The curricular subjects taught in CLIL include art, economics, history, geography, ICT, literacy, mathematics, music, physical education, science, etc. Some CLIL programs develop cross-curricular links among different subjects. For example, learners might study the history, geography and art of a particular area. In all CLIL contexts there is a need to analyze content for its language demands and to present content in an understandable way.

- Communication

Learners have to produce the subject language in both oral and written forms and, therefore, participate in a meaningful interaction in the classroom. Teachers should encourage increasing students' talking time, self-evaluation and peer and group feedback.

- Cognition

CLIL promotes cognitive skills, such as reasoning, creative thinking and evaluating, which challenge learners.

- Culture

The understanding of other cultures and learners' own culture is an important part of CLIL. This approach gives opportunities to introduce a wide range of cultural contexts and aims to develop learners who have a positive attitude towards a multilingual and multicultural world.

CLIL is not related to one specific methodology. It requires different methods, cooperative classroom management and emphasis on all types of communication. Input has a crucial role and becomes more visual and diverse. Multimedia support is often used to overcome problems caused by special languages. A CLIL lesson can include focus on a subject-matter, pre-activities that permit checking learners' background and previous knowledge, processing the text, work on tasks which are usually subject-orientated and are designed in such a way so that both content and language are recycled, post-reading and follow up tasks that aim to ensure and check comprehension and other activities.

The integration of language and content has a number of benefits for language learning. Firstly, language is acquired most effectively when it is learned for communication in meaningful and significant social situations. Secondly, the integration of content and second language instruction provides substantive basis and exposure for language learning. Thirdly, language acquisition naturally goes hand in hand with cognitive development. Finally, the integration of second language instruction with content instruction respects the specificity of functional language use. Besides improvement of overall and specific language competence, the benefits of CLIL may be seen in terms of cultural awareness, internationalisation, and increased motivation. The advocates of CLIL claim that it helps to introduce a wider cultural context, prepare learners for future studies and / or working life, and develop multilingual interests and attitudes.

However, some important critiques have been addressed to CLIL (Bruton 2011; Long 2015). First, most current CLIL programs are experimental, and there is not enough empirical evidence of how CLIL influences both language learning and content learning. As Long (2015, p. 214) pointed out, "research on CLIL is in its infancy, with most studies only descriptive and often poorly designed." It hasn't been proven that understanding of content is not reduced by

lack of language competence. As for language ability, it seems that it can only be increased by content-based learning after a certain stage. Bruton (2011, pp. 530–531) analyzed a number of studies on CLIL in Spain and concluded that although there are potentially beneficial effects from adopting CLIL initiatives, it cannot be seen in a vacuum, and in Europe “there are serious questions to be asked about the quality of the education state secondary school students receive, with the local version of CLIL and without it. If the actual pedagogy that is typical in most classrooms is producing deficient results, it would seem that adding the burden of using a foreign language for content instruction is perhaps overambitious.” Second, there remains a dearth of CLIL-type materials, and a lack of teacher training programs to prepare both language and subject teachers for CLIL teaching. Third, Long (2015) questioned the efficiency of both pedagogical procedures and CLIL materials. He pointed out that understanding “frozen texts” presented in textbooks on a particular subject becomes the focus of many lessons which is inadequate for language-learning purposes. Finally, there is a risk that learning the content through a language which neither the regular content teacher nor the children typically speak well may result in poor language development.

1.5.6 Comparison of task-based approaches

TBLT, project work and CLIL have a lot in common. They are analytic meaning-oriented approaches aimed at implicit language learning with room for explicit learning and attention to form. Tasks are used as a methodological tool and a unit of organization of the syllabus and learners’ work. However, there are some differences in the criteria of task selection and sequencing, task design, language focus, the teacher’s role, classroom procedures, evaluation, and the area of actual use. TBLT is more frequently used with adult learners, whereas CLIL and project work are usually applied in secondary schools. These three task-based approaches are compared in Table 7.

Table 7. TBLT, CLIL and project work (Gilabert 2013)

	TBLT	Project work	CLIL
Task selection	Determined by needs analysis	Pre-planned by teacher and then negotiated with students	Determined by subject-matter
Task design	By syllabus designer	Negotiated with students	Adapted from subject-matter
Focus on language	Focused/Unfocused Mainly reactive	Negotiated with learners	Reactive
Group arrangements	Pairs or groups	Groups	Unspecified
Sessions organized	Tasks	Steps in the project/tasks	Thematic content

around			units/tasks
Sequencing	Task complexity/difficulty	Follows steps in the project	Determined by subject-matter
Teacher's role	Design and provide input/feedback	Negotiator/facilitator	Subject-matter expert
Evaluation	In terms of task performance	Global evaluation of project by learners, teacher and sometimes external observer	Mastery of subject-matter contents

1.5.7 Problems with task-based syllabi

The task-based approaches have been seen as a better alternative to synthetic approaches with focus on forms and purely analytic approaches with no focus on form at all. However, all task-based syllabi share certain problems such as the difficulty of differentiating tasks, especially tasks and subtasks, determination of the relevant sequencing criteria, and a lack of a rigorous field evaluation. The Procedural syllabus and the Process syllabus deal with pedagogic tasks which are not based on any prior needs identification, and it raises problems for task selection. No explicit provision is made for a focus on language form. Moreover, the Process syllabus assumes unrealistically high level of competence in teachers and learners and a redistribution of power and authority in the classroom that can be unacceptable in some societies.

Problems with TBLT involve issues of needs analysis for some groups of learners (e.g. when there exists heterogeneity of needs within a class or when learners are very young and have no easily predictable needs), the role of negotiation of meaning, and the question of using the task-based approach with beginner learners. Teachers' concerns include a shortage of published task-based materials, a lack of teacher training programs, and insufficient evidence on how TBLT works in real classroom contexts.

1.6 Summary of Chapter I

Chapter I has dealt with the questions of L2 syllabus design and instruction. The chapter has begun with the description of steps which a syllabus designer is expected to follow and a review of options for syllabus design which depend on different approaches to language teaching. In the subsequent sections, three major options in language teaching – synthetic syllabi with focus on forms, analytic syllabi with focus only on meaning and analytic syllabi with focus on form – have been analyzed.

Synthetic syllabi, which are organized around grammatical units, lexical units, skills, notions, functions, situations and topics, have been described. We have discussed some of the problems with these syllabi. Their lack of coherence with SLA findings was highlighted.

Analytic syllabi with an exclusive focus on meaning were then analyzed, including the Natural Approach (Krashen & Terrell 1983) and Canadian immersion programs. Disregard of adult learners' maturational constraints and their reduced capacity for purely incidental learning, as well as the lack of focus of form and inefficient treatment of learners' persistent grammatical errors were emphasized as the major drawbacks of these syllabi.

Finally, analytic syllabi that adopted task as a unit of syllabus design were outlined. This section began with the description of the early versions of task-based syllabi. The Procedural syllabus (Prabhu 1987) was presented as one of the first approximations to a syllabus composed by tasks. The Process syllabus (Breen 1984; Candlin 1987) was defined as syllabus in which each task and other aspects of the program are negotiated with learners. Among more recent task-based proposals, TBLT, project work and CLIL were outlined. They were described as analytic meaning-oriented approaches with a room for explicit learning and attention to form in which tasks are used as a unit of organization of the syllabus and learners' work. Although these proposals seem to overcome the limitations of other analytic approaches and synthetic syllabi, they are not without problems. A lack of a rigorous field evaluation, a shortage of well-designed long-term studies, a lack of teacher training programs, and insufficient evidence on how the task-based syllabus works in real classroom contexts were outlined as the main concerns cited by researchers and teachers.

In the next chapter the key questions of TBLT research which are relevant for this study will be extensively reviewed, including the definition of task, different options in task design and their impact on learner's production, the problem of task sequencing and the models of task complexity.

CHAPTER II

TASK-BASED LANGUAGE TEACHING: KEY CONCEPTS AND QUESTIONS

2.1 Introduction

In the previous chapter different approaches to language teaching have been described, including those that adopted task as a unit of syllabus design. This chapter will focus on Task-Based Language Teaching and the main stages in task-based syllabus design. As shown in the previous chapter, the process of syllabus design involves decisions about what is considered the central unit of the syllabus, unit selection (what to teach), unit sequencing (in what order), methodological implementation (how to teach), assessment (how to measure progress), and program evaluation (whether the syllabus works). Chapter II will deal with different proposals for each stage (except the question of program evaluation which will be discussed in Chapter IV).

First, different definitions of a task will be discussed and the key task characteristics will be outlined. Then we will look at the process of task selection and needs analysis. Special attention will be paid to the controversial issue of task sequencing and Robinson's (2001b) and Skehan's (1998a) approaches to task complexity and task difficulty. Then a number of methodological options in task design and implementation which are available for SLA researchers and classroom teachers will be discussed. A variety of pedagogic task types will be described and their possible effects on learner production considered. A typical three-stage task framework which includes a pre-task, during task phase and a post-task will be reviewed. Finally, the key questions of task-based assessment will be discussed, among which are questions of design of assessment tasks and criteria of task performance evaluation.

2.2 Definitions of task

Since task is a central concept of TBLT, it is crucial to identify what is considered as a task. A number of definitions of a task can be found in SLA literature (for a general review see Bygate, Skehan, & Swain 2001; Ellis 2003; Long 2015; Nunan 2006, among others). In this section only well-accepted definitions have been included.

First attempts to conceptualize a task resulted in all-inclusive definitions of what a task is, such as the one proposed by Long (1985, p. 89) who defined tasks by looking at what people usually do in real life:

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

As Nunan (2006) pointed out, this definition is non-technical and non-linguistic. It implies that in contrast with most classroom language exercises, tasks have a non-linguistic outcome. Non-linguistic outcomes from Long's list above might include a painted fence, a driver's license, a room in a hotel, etc. Another thing to notice is that this definition includes tasks which do not require language use (it is possible to paint a fence without talking).

A narrower definition of a task was offered by Crookes (1986, p. 1): “A task is a piece of work or an activity, usually with a specified objective, undertaken as a part of an educational course, at work, or used to elicit data for research.”

The authors of early task-based syllabi, the Procedural syllabus (Prabhu 1987) and the Process syllabus (Breen 1984, 1987b), in their definitions of a task outlined that tasks have an outcome, and task completion involves some cognitive processes and a specific working procedure:

Task is “an activity which requires learners to arrive at an outcome from given information through some process of thought and which allows teachers to control and regulate that process” (Prabhu 1987, p. 24).

Task is “any structured language learning endeavor which has a particular objective, appropriate content, a specific working procedure, and a range of outcomes. Task is therefore assumed to refer to a range of workplans which have the overall purpose of facilitating language learning – from the simple and brief exercise type, to more complex and lengthy activities such as group problem-solving or simulations and decision-making” (Breen 1987b, p. 23).

These definitions are very broad since they imply that just about anything the learner does in the classroom qualifies as a task. They could, in fact, be used to justify any procedure as ‘task-based’, and, as such, are not particularly helpful (Nunan 2006).

The next important step in defining a task was made by adding communicative characteristics⁵ and a focus-on-meaning rather than focus-on-form goal. Both of these traits were reflected in the definition given by Nunan (2004, p. 4):

“A task is a piece of classroom work that involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is focused on mobilizing their grammatical knowledge in order to express meaning, and in which the intention is to convey meaning rather than to manipulate form. The task should also have a sense of completeness, being able to stand alone as a communicative act in its own right with a beginning, a middle and an end.”

Skehan (1998a, p. 95) identified a series of traits most researchers would agree on when conceptualizing a task. A task is an activity in which:

- meaning is primary
- there is some communication problem to solve
- there is some sort of relationship to comparable real-world activities
- task completion has some priority
- the assessment of the task is in terms of outcome

Three of these characteristics of a task – a primary focus on meaning, relationship to comparable real-world activities and a presence of a clearly defined communicative outcome through which the task can be assessed – are outlined in the definition given by Ellis (2003, p. 16):

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 resources, 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.”

Long (2015, p. 109) criticized definitions given by Ellis (2003), Nunan (2004) and Willis (1996) pointing out that they “have been characterizations of classroom tasks coming from task-supported approaches developed without reference to analyses of learner needs beyond the classroom.” As in his earlier work (Long 1985), he insisted that in TBLT tasks are “the real-

⁵ Willis (1996, p. 23) defined a task as “an activity where the target language is used by the learner for a communicative purpose (goal) in order to achieve an outcome.”

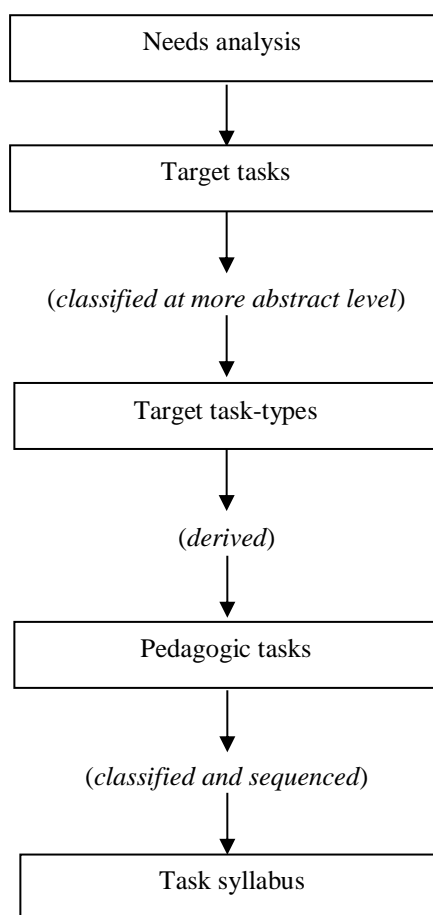
world activities people think of when planning, conducting, or recalling their day” (Long 2015, p. 6).

2.3 Task selection

According to Long’s approach to task-based syllabus design, “selection of syllabus content in TBLT is determined rationally and empirically, according to the results of a learner needs analysis” (Long 2015, p. 223). He argues that a one-size-fits-all approach has long been discredited by research findings, and language courses must be relevant to the needs of specific groups of learners. Long (2005) provided a summary and evaluation of various sources of information for a needs analysis (NA) (published and unpublished literature, the learners, teachers and applied linguists, domain experts, triangulated sources), methods of obtaining that information (expert intuitions, participant and non-participant observation, structured and unstructured interviews, questionnaires, criterion-referenced performance tests, discourse analysis, computer-aided corpus analysis, and triangulated methods, among others), and source and method combinations.

Needs analysis helps to obtain an inventory of target tasks – “those tasks identified as required in order for an individual to function adequately in a particular target domain, be it occupational, vocational or academic” (Long 1985, p. 91). In other words, target tasks for a particular group of learners are what they need to be able to do in the new language in the world beyond the classroom (Long 2015; Nunan 2004). In the situation when the same target tasks are important for all the students in a class and there is time available to teach all of them, pedagogic tasks, that is, tasks that teachers and students work on in the classroom or other instructional environments, are derived directly from each of the target tasks. However, in real classroom practice it is rarely the case. Long (2015) points out that target tasks are not usually themselves the components that make up the syllabus. When a course is of limited duration or when there are several sub-groups of students in a class with some target tasks in common and some unique to each group, target tasks with features in common are grouped together as members of target task-types. Since target task-types are abstract and, for this reason, are unsuitable as units for a task syllabus, pedagogic tasks are derived from them and then sequenced to form the task syllabus (see Figure 6).

Figure 6. Steps and processes in TBLT syllabus design (Long 2015, p. 224)



As an example, Long (2015) presents the process of designing a task syllabus for airline flight attendants. NA identified over 100 target tasks for this group of learners (for details, see Long 2005, pp. 48–62). Several target tasks involved passenger service, including serving of breakfast, lunch, dinner, drinks, snacks, etc. These tasks can be brought together as members of the target task-type “Serve food and beverages.” For this target task-type Long offered a few pedagogic tasks: 1) identify choices between two food items, 2) identify choices among multiple items, 3) respond to taped choices when some items are unavailable, 4) role play, and 5) full simulation with verbal presentation of choices and identification of passenger selections as the exit task.

Pedagogic tasks are simpler versions of the target task-type. They “gradually increase in complexity – the complexity of the pedagogic tasks, not linguistic complexity – until they reach the full complexity level of the target tasks that motivated their inclusion” (Long 2015, p. 225). The last task in a module of TBLT materials, according to Long’s proposal, should be a genuine or a virtual version of the full target task. It serves as the exit task for the unit and it is used to assess students’ ability to perform the target task.

Long outlined several ways of simplifying target tasks. First, tasks may be simplified by breaking a target task down into sub-tasks. As Gilabert (2005) showed in NA conducted for journalists, one of the target tasks “Interviewing a source” involved six sub-tasks: deciding on the decision-making process, contacting the source, documenting the interview, making arrangements for the interview, interviewing, and translating the transcript or interview for publication. Some sub-tasks themselves consisted of two or more parts. For example, the target sub-task “Making arrangements for the interview” consisted of three parts: (1) calling or emailing the source to request an interview, (2) calling or emailing the source to arrange time, place and topics for the interview; and (3) sending an email or fax message to the source to confirm arrangements. Other ways of simplifying target tasks proposed by Long (2015) include use of pre-tasks that build schema to assist with pedagogic tasks that follow, and elaborating the input used in native speakers’ performance of the corresponding target tasks.

In TBLT literature there are other views on the criteria of task selection different from Long’s approach. His proposal is radical to some extent since he excludes linguistic complexity both from the process of task selection and task sequencing. However, in morphologically rich languages such as Russian, linguistic complexity cannot be ignored while designing a language course (Gilabert & Castellví, in press). For this reason, Skehan’s (1998a) proposal may be more suitable for teaching morphologically complex languages because he includes a linguistic factor in the process of task selection. One of the principles of task-based instruction he advocates for is choosing a range of target structures and choosing tasks which meet the utility criterion. Although Skehan (1998a, p. 130), as well as Long (2015), admits that “it is futile to fix on a particular structure and expect it to be learned,” he argues that “there is a need for systematicity in language development, and the only way to resolve this conundrum satisfactorily is to have some method of keeping track of interlanguage development, but not in narrow specific terms”. The utility criterion was offered by Loschky and Bley-Vroman (1993) to describe a degree of involvement of a grammar structure in a task when it is possible to complete a task without a particular structure, but with the structure the task becomes easier (for more details see Section 3.4.4). Skehan (1998a, p. 130) claims that the teacher cannot force students to use any particular structure in task performance, but he/she can create appropriate conditions for its use, although “no guarantees exist as to which structures will actually be used in the ongoing nature of performance.”

Therefore, tasks should be selected and designed having target structures in mind. It corresponds to Ellis’ notion of “focused tasks” (2003), that is, tasks encouraging the processing of specific language features, as opposite to unfocused tasks involving unspecified language use.

Unlike exercises, which are “primarily concerned with practicing a specific form”, focused tasks require that learners “are not informed of the specific linguistic focus” (Ellis 2003, p. 141), and are free to concentrate on meaning and to choose their own resources.

2.4 Task sequencing

2.4.1 Early proposals for task sequencing

Language teaching literature offers a variety of criteria for sequencing items in a syllabus. Communicative value (or usefulness), frequency, learnability (Pienemann 1984, 1985), and difficulty have often been proposed as a way of ordering items in structural and lexical syllabi. In the case of the task-based syllabus, task difficulty, or task complexity, have been considered the main criteria for sequencing tasks. The terms “complexity” and “difficulty” had been used interchangeably before Robinson (2001a, b) drew a distinction between them.

Among the first proposals for grading tasks, a set of criteria offered by Brown *et al.* (1984), Prabhu (1987) and Candlin (1987) can be outlined. One of the first attempts at sequencing tasks from simple to complex was advanced by Brown *et al.* (1984). They distinguished among three different types of tasks which they presented as ranging from easy to difficult: static→dynamic→abstract. In static tasks, all the information is presented to the speaker in the materials for carrying out the task (e.g. a map task in which the speaker has to give directions to the listener). In dynamic tasks, the speaker also has all the information in stimulus materials, but characters, events, and activities change (e.g. a story in a comic strip), and this change forces the speaker to describe the stimulus material in a more precise, detailed way. Abstract tasks are the most difficult since the stimulus material does not contain the content to be communicated. These tasks involve “making reference to abstract concepts, establishing connections between ideas, and providing reasons for certain statements or behaviors (e.g. an opinion task in which learners must choose the most suitable candidate for a scholarship out of a closed list of candidate descriptions)” (Gilbert 2004, p. 98). Beyond this distinction on static, dynamic, and abstract types, Brown *et al.* (1984) pointed out that each task type can be made more difficult by increasing the number of elements and the relationships between them.

Another proposal for sequencing tasks in a syllabus belongs to Prabhu (1987, p. 88–89) who indicated “rough measures of cognitive complexity” based on his experience with the Bangalore project. He suggested the following parameters:

1. Information provided. The amount of information to be handled and a variety in types or sources of information make a task more or less difficult.

2. Reasoning needed. “The ‘distance’ between the information provided and the information to be arrived at as outcome – i.e., the number of steps involved in the deduction, inference, or calculation – is a measure of the relative difficulty of tasks,” (Prabhu 1987, p. 89). For this reason, information-gap tasks are easier and occurred earlier in Prabhu’s procedural syllabus than reasoning-gap tasks.

3. Precision needed. More or less precision may be needed to interpret the information or express an outcome in different tasks. Difficulty level increases with the degree of precision called for.

4. Familiarity. Learners’ knowledge of the world can make tasks more or less difficult for them, depending whether they are more or less familiar with purposes and real-world concepts and situations involved in a task.

5. Degree of abstractness. Working with concepts is more difficult than working with names of objects or actions.

6. Medium of input and outcome. A change “from orally presented tasks to similar ones presented in writing proved to provide a distinct increase in difficulty for project classes” (Prabhu 1987, p. 90).

One more attempt to characterize task difficulty as a set of criteria by which tasks might be graded was made by Candlin (1987). These criteria included the following elements:

- Cognitive load, which concerns the general complexity of the content of the task;
- Communicative stress, which depends on the interlocutor and his/her level of proficiency;
- Particularity and generalizability, which concern the clarity of the goal of the task;
- Linguistic code complexity;
- Process continuity, which derives from the familiarity of the task type.

Although some of these criteria were adopted later in TBLT research, three outlined proposals have been criticized. For example, Skehan (1998a, p. 99) called Candlin’s set of criteria “difficult to relate in any transparent way to actual tasks of the sort that might figure in conventional communicative teaching.” Long and Crookes (1992, 1993) also found early approaches to task sequencing unsatisfactory and pointed out that the issue of task difficulty and determination of the relevant sequencing criteria was one of the most problematic in TBLT.

2.4.2 Two models of task complexity

Task complexity is a notion that has received considerable attention in TBLT studies over the last years (see, for example, Baralt, Gilabert, & Robinson 2014; Gilabert, Barón, & Llanes 2009; Kuiken & Vedder 2004, 2007; Robinson 2001a, b, 2005, 2007, 2011b; Robinson & Gilabert 2007; Skehan 1998a, 2001; Skehan & Foster 2001). As Gilabert (2004, p. 97) points out, the concept of task complexity “was born from the need to establish criteria for sequencing tasks in a syllabus from easy/simple to difficult/complex in a reasoned way that will foster interlanguage development. Rather than looking at the linguistic features of language activities, syllabi that have used tasks as their units have focused on task design in order to find out how tasks impose cognitive demands on learners.” There are two widely recognised approaches to task complexity within the literature on TBLT: Skehan’s (1998a) and Robinson’s (2001a, b, 2005) models.

Skehan (1998a) and Skehan and Foster (2001) adopted Candlin’s (1987) model of task difficulty and grouped some of the factors suggested by Candlin into higher-order categories. Skehan (1998a) proposed a three-way distinction for the analysis of tasks, based on code complexity, cognitive complexity and communicative stress. These dimensions involve the differences “between the language required, the thinking required, and the performance conditions for a task” (Skehan 1998a, p. 99). Code complexity reflects the morphosyntactic and lexical density and variety required. Cognitive complexity is affected by such factors as familiarity, organization, sufficiency, and type of information in a task. Communicative stress, i.e. the conditions under which a task is performed, varies according to such factors as the number of participants, the time available for task completion, the length of texts, and opportunities to control interaction. Skehan’s model of task difficulty is presented in Table 8.

Table 8. Skehan’s model of task difficulty (1998a)

Code complexity	Cognitive complexity	Communicative stress
Linguistic complexity and variety Vocabulary load and variety Redundancy and density	Cognitive familiarity Familiarity of topic and its predictability Familiarity of discourse genre Familiarity of task Cognitive processing Information organization Amount of computation Clarity and sufficiency of information given Information type	Time limits and time pressure Speed of presentation Number of participants Length of texts used Type of response Opportunities to control interaction

Skehan relates task difficulty to the amount of attention needed to perform the task: “Task difficulty has to do with the amount of attention the task demands from the participants. Difficult tasks require more attention than easy tasks” (Skehan & Foster 2001, p. 196). Skehan (1998a, p. 134) points out that it is crucial to make sure that task difficulty is appropriate for learners: “Given that attentional capacities are limited, tasks of appropriate difficulty imply that learners will be able to cope with the demands upon their attentional resources. <...> If the appropriate level of task difficulty is chosen, there is much greater likelihood that noticing will occur, that balanced language performance will result, and that spare attentional capacity can be channeled effectively.”

Skehan’s position is known as the Limited Attention Capacity Model (LACM). He argues that humans have a limited information processing capacity, and L2 learners must therefore prioritize where they allocate their attention during task performance, whereby attention allocated to one dimension of language production will be lost in others, leading to trade-offs (Trade-off Hypothesis). “All other things being equal, there will be pressure on limited attentional resources, so that task complexity is likely to provoke enhanced task performance in some areas at the expense of others” (Skehan & Foster 2012, p. 215). Skehan sees fluency as an aspect of L2 production which competes for attentional resources with accuracy, while accuracy in turn competes with complexity. However, task characteristics can have selective influences which modify the effects of trade-off. Different characteristics of tasks (for example, task structure and information integration) can support different performance areas that would lead to overcoming of trade-off limitations.

A competing model of attention, the Cognition Hypothesis, was proposed by Robinson (2001a, b, 2003, 2011b). He argues that attentional resources are not so limited. Rather, learners draw on multiple attention pools simultaneously. Increasing demands on the cognitive dimension of tasks promotes interlanguage development by causing learners to focus on L2 features required to express the new cognitive distinctions. Robinson predicts that it can result in an increase in accuracy and linguistic complexity.

For sequencing tasks, Robinson (2001a) has drawn a distinction between task complexity and task difficulty, and added the dimension of task condition. His three-dimensional model, known as The Triadic Componential Framework (see Table 9), distinguishes between three types of factors: cognitive, interactive and difficulty factors.

Table 9. Robinson’s Triadic Componential Framework (2001a)

Task complexity (Cognitive factors)	Task condition (Interactive factors)	Task difficulty (Learner factors)
Resource directing +/- few elements +/- here-and-now -/+ spatial reasoning -/+ causal reasoning -/+ intentional reasoning -/+ perspective-taking Resource dispersing +/- planning time +/- single task +/- task structure +/- few steps +/- independency of steps +/- prior knowledge needed	Participation variables +/- one-way flow +/- convergent solution +/- open solution +/- few participants +/- few contributions needed +/- negotiation not needed Participant variables +/- same gender +/- familiarity +/- same proficiency +/- shared content knowledge +/- equal status and role +/- shared cultural knowledge	Ability variables h/l working memory h/l reasoning h/l task-switching h/l aptitude h/l field independence h/l mind/intention-reading Affective variables h/l openness to experience h/l control of emotion h/l task motivation h/l processing anxiety h/l willingness to communicate h/l self-efficacy

Robinson (2001a, p. 28) defines task complexity as “the result of the attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner. These differences in information processing demands, resulting from design characteristics, are relatively fixed and invariant.” Task complexity refers to the intrinsic cognitive demands of the task, and it can be manipulated during task design along resource-directing (+/- number of elements, +/- here and now, +/- reasoning demands) and resource-dispersing dimensions (+/- planning, +/- single task, +/- prior knowledge needed).

The second dimension in Robinson’s model is task condition. It is the dimension that unites interactive factors. They include participation variables (one-way vs. two-way, convergent vs. divergent, open vs. closed tasks) and participant variables (familiarity, gender, shared cultural knowledge, and other characteristics). As Long (2015, p. 234) points out, “teachers and material writers can control and manipulate these and other interactive factors, but they are not factors used to sequence two or more pedagogic tasks. Conditions affect the way students experience a single task.”

By task difficulty Robinson understands what learners bring to the task. Task difficulty may vary for two students due to individual differences between them. Difficulty factors depend on affective variables (such as motivation, anxiety, and willingness to communicate) and ability variables (such as aptitude, working memory, and reasoning capacities). Robinson suggests that differences in ability variables affect learners’ perception of the task with consequences for performance and learning.

Comparing these three dimensions, Long (2015, p. 234) concludes that complexity is “unchanging, a product of a task’s intrinsic properties and the cognitive demands it makes on learners. Difficulty, conversely, is variable, a function of the combination of the complexity of a

task and the abilities individual learners bring to the table, coupled with modifications teachers may make to task condition.” Both Robinson (2007) and Long (2015) claim that pedagogic tasks in a TBLT syllabus have to be sequenced solely according to task complexity, that is, cognitive demands tasks impose on learners.

Robinson’s model has been considered as a significant contribution to the question of task sequencing. However, it has been reported to create some problems for practitioners. As Kuiken and Vedder (2007) pointed out, it is far from clear how all twelve variables of task complexity have to be operationalized and differentiated. Robinson’s (2011a) answer to this problem is that not all L2 dimensions will be relevant to all task performances. He recommends matching the demands of a target task to the dimensions of task complexity, increasing first the complexity of relevant resource-dispersing dimensions, and second, the complexity of resource-directing dimensions.

Another problem with Robinson’s model of task complexity is that he does not consider linguistic difficulty a criterion of task sequencing. As Gilabert and Castellví (in press) point out, in the majority of studies on task complexity, task features affecting task cognitive complexity have been taken as independent factors with linguistic difficulty mostly remaining outside the picture. Additionally, most of the research into tasks has been done with morphologically poor languages such as English. Probably, linguistic difficulty (or code complexity in Skehan’s terms) is less important for English which is characterized by relatively morphologically simple structure, few and clear syntactic marks, and a great tendency to monosyllabic words, which simplifies the lexicalization process and allows memory and attention resources to be oriented to lexical entry processing. However, code complexity is crucial for morphologically rich languages such as Russian. For Russian, especially at the earlier levels, it is very important to control both the grammatical and lexical content needed to complete the task, since excessively high code complexity may impede learning, make achieving an outcome impossible, and, finally, lead to discouragement. Code complexity makes the process of a task-based syllabus design especially difficult since the amount of new morphology needed and the amount of understandable but not acquired morphology can lead the task to collapse. Linguistic difficulty may take so much attention that learners can barely handle other dimensions of the task. For this reason, in the process of task selection and further task design not only must communicative goals and cognitive complexity be taken into account, but the linguistic resources needed to perform the task must also be considered.

Both Robinson’s Triadic Componential Framework and Skehan’s Limited Attention Capacity Model have been the subject of a considerable number of empirical studies (Bygate,

Skehan, & Swain 2001; Foster & Skehan 1996; Gilabert 2004; Kuiken & Vedder 2007; Rahimpour 1997; Robinson 2011b; Robinson & Gilabert 2007; Yuan & Ellis 2003, among others). In most studies attention has been paid to effects of changing one or more dimensions of task complexity (e.g., the amount of planning time, Here-and-Now vs. There-and-Then) on learners' accuracy, complexity, and fluency. D. Ellis (2011) provided a critical review of both models and some studies. He found that comparison of findings in favor or against each model is impossible for several reasons. First, tasks employed and measures applied have varied greatly across studies. Apart from obvious methodological differences, results within and across studies have been mixed, with most findings failing to reach statistical significance. D. Ellis' conclusion (2011), supported later by Long (2015), is that not a single published study provides unambiguous support for either model.

2.5 Research into pedagogic task types

TBLT literature offers a description of a great variety of pedagogic task types. As Ellis (2012, p. 200) points out, tasks can differ in terms of how they are designed and how they are implemented in the classroom. Both design and implementation variables can potentially influence how a task is performed. A number of researchers have tried to classify tasks for classroom and research purposes (Nunan 1989; Pica *et al.* 1993; Prabhu 1987; Willis 1996). Different classifications of tasks have been summarized in Ellis (2003, p. 211–217) who has distinguished a pedagogic classification, a rhetorical classification, a cognitive classification, and a psycholinguistic classification (see Table 10).

A pedagogic classification (Willis 1996) reflects the kind of operations learners are required to carry out while performing tasks. These operations may involve:

1. Listing (brainstorming, in which learners draw on their own knowledge and experience or fact-finding, in which learners find things out by asking each other or other people and referring to books),
2. Ordering and sorting (sequencing, ranking, categorizing, classifying),
3. Comparing (matching, finding similarities and differences),
4. Problem solving (expressing hypotheses, comparing alternatives, predicting the ending, case studies),
5. Sharing personal experience,
6. Creative tasks (projects).

A rhetorical classification (Arnaudet & Barrett 1984) is based on different discourse domains – narrative, instructions, description, reports, etc. Swales (1990) uses the concept of genre to classify tasks (recipes, political speeches, job application letters, etc.).

A cognitive classification (Prabhu 1987) is based on the kind of cognitive operations different types of tasks involve.

1. Information-gap activity, which involves a transfer of given information from one person to another (e.g. pair work in which each member has a part of the total information and attempts to convey it verbally to the other);
2. Reasoning-gap activity, which involves deriving some new information from given information through processes of inference, deduction, practical reasoning, or a perception of relationships or patterns (e.g. working out a teacher’s timetable on the basis of given class timetable);
3. Opinion-gap activity, which involves identifying and articulating a personal preference, feeling, or attitude in response to a given situation (e.g. taking part in the discussion of a social issue).

Finally, a psycholinguistic classification (Pica, Kanagy, & Falodun 1993), which has been widely used in TBLT research, is based on a different configuration of activity and goal.

1. Interactant relationship (i.e., who holds the information and who requests it, one-way and two-way tasks),
2. Interaction requirement (whether obligations to request or supply task-related information are required or optional),
3. Goal orientation (whether the task requires the participants to agree on a single outcome (convergent) or allows them to disagree (divergent)),
4. Outcome options (closed and opened tasks).

Table 10. Classifications of tasks (Ellis 2003, pp. 211–217)

Pedagogic classification (Willis 1996)	Rhetorical classification (Arnaudet & Barrett 1984; Swales 1990)	Cognitive classification (Prabhu 1987)	Psycholinguistic classification (Pica, Kanagy & Falodun 1993)
1. Listing 2. Ordering and sorting 3. Comparing 4. Problem solving 5. Sharing personal experience 6. Creative tasks (projects)	By discourse domains: 1. Narrative 2. Instructions 3. Description 4. Reports, etc. By genre: 1. Recipes 2. Political speeches 3. Job application letters, etc.	1. Information-gap activity 2. Reasoning-gap activity 3. Opinion-gap activity	1. Interactant relationship (one-way/two-way tasks) 2. Interaction requirement (required/optional) 3. Goal orientation (convergent/divergent) 4. Outcome options (closed/opened tasks)

Research in the last few decades has tried to isolate the different variables involved in task design in order to test their effects on production and language development. Gilabert (2004) distinguished two different agendas that have inspired research into task features: interactionist and information-processing perspectives. An interactionist perspective has been concerned “with establishing what modifications can be applied to tasks in order for them to generate specific conversational episodes which, generally, have been regarded as negotiation of meaning” (Gilabert 2004, p. 94). These episodes have been claimed to lead to second language acquisition (Long 1985, 2015). Research tasks have been manipulated along a one-way or two-way flow of information during interaction (Long 1983; Pica & Doughty 1988), their opened-ended or closed outcomes (Long 1989; Loschky & Bley-Vroman 1993; Rahimpour 1997), the convergence or divergence of goals (Duff 1986), optional or required information exchange (Doughty & Pica 1986), pair and group work arrangements (Pica & Doughty 1985), and split versus shared information among participants (Newton & Kennedy 1996; Pica & Doughty 1988). Apart from their interest in negotiation of meaning during production, researchers have also looked into the consequences of task manipulation on the amount of production and the level of participation of learners. Below, a number of design variables will be briefly described, and some research findings on relationships between pedagogic task-types and classroom processes, performance quality, and learning outcomes will be discussed.

- *One-way versus two-way tasks*

In a one-way configuration, “all of the task-essential information is allotted to one learner who must communicate it to the other(s). In a two-way configuration, the task-essential information is distributed between all the learners in a dyad or a larger group, and they must share it for the task to be completed successfully” (Long 1989, p. 13). An example of a one-way task would be a task where a learner describes a picture that his / her classmate, who does not have it, must draw or identify among other pictures. A typical two-way task is “Spot the differences” when learners have the same pictures with some objects in each picture missing or being situated in different places and they have to find out the differences by asking each other about the objects in the picture. Some studies showed that one-way tasks may produce little or no negotiation for meaning (Gass, Mackey, & Ross-Feldman 2005; Long 1983). However, they allow more modification of output than two-way tasks (Iwashita, McNamara, & Elder 2001). Two-way tasks have been shown to promote more negotiation of meaning and more balanced levels of participation (Gass, Mackey, & Ross-Feldman 2005; Pica & Doughty 1988).

Nevertheless, some classroom studies (Foster 1998; Slimani-Rolls 2005) have found little negotiation work and considerable differences at the level of individual learners.

- *Open versus closed tasks*

For an open goal orientation, “learners know that there is no correct solution to the task, whereas for a closed goal orientation they know that there is only one or a small range of possible solutions” (Long 1989, p. 18). Examples of an open task would be a debate about a controversial issue to which each learner in a group can contribute ideas, or choosing objects to take to a desert island. An example of a closed task would be an information-gap task in which learners must spot ten differences in a picture. Open tasks can result in more complex and accurate language (Duff 1986; Skehan 1998a). However, Rahimpour (1997) found that there was a trend for higher accuracy in closed tasks, and no differences in complexity between open and closed tasks. Closed tasks can lead to more negotiation for meaning, more feedback and greater fluency (Rahimpour 1997).

- *Convergent versus divergent tasks*

Convergent tasks require learners to work together to reach a common solution, whereas divergent tasks require them to defend opposing views (Duff 1986). A traditional example of a convergent task is the Desert Island task in which learners must agree on a limited number of objects to take to the island and must therefore work together to find an acceptable solution. Divergent tasks usually take the form of debates in which each group is assigned an opposite position on a controversial issue that they must defend. Some findings have shown that convergent tasks produce more interactional modifications and negotiation for meaning than divergent tasks. Convergent tasks also generate more frequent turn-taking but shorter and less syntactically complex turns than debates (Duff 1986; Skehan & Foster 2001).

Another group of studies have taken an information-processing perspective and have investigated the effects of tasks on production along their degree of familiarity (Bygate 1996, 1999, 2001; Foster & Skehan 1996; Robinson 2001a), number of elements (Kuiken & Vedder 2004; Robinson 2001a), the pre-task (or strategic) and on-line planning time allotted to them (Foster & Skehan 1996; Mehnert 1998; Ortega 1999; Skehan & Foster 1997; Yuan & Ellis 2003), and here-and-now versus there-and-then dimensions (Gilbert 2004; Iwashita, McNamara, & Elder 2001; Rahimpour 1997; Robinson 1995a). These studies “have been concerned with how a balanced performance in the three areas of production can potentially lead to more effective language use and acquisition, as well as with how such information can be used to make sequencing decisions in syllabus design” (Gilbert 2004, p. 96).

- ***Task familiarity***

Task familiarity has been interpreted in a number of ways. For example, by task familiarity Foster and Skehan (1996) meant the task type. Robinson (2001a) interpreted familiarity in terms of content (for example, familiarity with a route marked in a map). Finally, Bygate (1999, 2001) interprets task familiarity as triggering more or less familiar language patterns or discourse genres, and his research deals with the effects of task repetition. Hence, task familiarity can be interpreted as referring to the task type, task topic, or task repetition. Task repetition can have a strong effect on performance especially in the areas of accuracy and fluency. Bygate (1999) found that the same task repeated a few days later triggered more accurate, fluent and lexically complex language. The explanation he provides is that familiarity gives learners “the time and awareness to shift attention from message content to the selection and monitoring of appropriate language” (Bygate 1999, p. 41).

- ***Number of elements***

Higher number of elements increases task complexity, and, according to Robinson’s Cognition Hypothesis (2001a, b), may lead to an increase in accuracy and linguistic complexity. This hypothesis was partially confirmed by his study on task complexity in which Robinson (2001a) combined two variables, familiarity and the number of elements, simultaneously. He found that the more complex version of task, which involved an unmarked map route with many elements and references to an area that was unfamiliar to students, generated significantly higher lexical complexity and lower fluency, with no difference in accuracy or structural complexity.

- ***+/- planning time***

A lot of attention has been paid to task performance under and without time pressure. Providing pre-task planning time (typically five to ten minutes) can result in higher fluency and higher structural complexity (Foster & Skehan 1996; Ortega 1999; Yuan & Ellis 2003). Mixed results have been obtained for lexical complexity. Gilabert (2004) reported significant effects of planning time on lexical complexity, whereas Ortega (1999) found no such effects. Mixed results have also been obtained for accuracy. Pre-task planning time promoted higher accuracy in Foster and Skehan’s study (1996), no differences were found in Yuan and Ellis (2003), and mixed results were reported by Ortega (1999).

- ***Here-and-Now versus There-and-Then***

Tasks in the There-and-Then have been considered to be more cognitively demanding than tasks performed in the Here-and-Now since displaced, past time reference is more complex than present, context-supported reference (Gilabert 2004). Studies that investigated effects of this variable on learner production have shown that There-and-Then narrative elicited more accurate

and less fluent speech than the narrative performed in Here-and-Now (Gilabert 2004; Rahimpour 1997; Robinson 1995a). No significant differences for structural complexity were found. Robinson (1995a) reported an increase of lexical complexity measured by the percentage of lexical words out of the total number of words. However, Rahimpour (1997) and Iwashita *et al.* (2001) did not find significant differences regarding lexical complexity.

This brief review demonstrates that despite the existence of numerous data-based studies in which different design variables have been manipulated, there are few clear and consistent findings. Long (2015) has discussed some of the reasons for this state of affairs. He has pointed out that “many of the varied findings seem due to methodological differences in how studies ostensibly of the same issues are carried out” (Long 2015, p. 244). For example, obvious differences in measures of dependent variables employed in each study may significantly affect the obtained results (Norris & Ortega 2000). Another reason is that this kind of research typically involves the recording of student task performance which is especially problematic in classroom conditions. The result is use of small n-sizes, with all the usual dangers it poses to findings. One more problem is that many tasks simultaneously constitute examples of two or more task-types, so that two or more influences may be at work in producing the observed outcomes. As a possible solution to the problem with clarity and consistency in findings, Long (2015) proposes research programs involving research motivated by the same theory and with use of the same measures.

2.6 Methodological options in a task-based lesson

Work on a task usually consists of three phases: pre-task, during task (or task cycle) and post-task phases. Only the ‘during task’ phase is obligatory in task-based teaching. Options selected from the ‘pre-task’ or ‘post-task’ phases are non-obligatory but can serve a crucial role in ensuring that the task performance is maximally effective for language development (Ellis 2006).

J. Willis (1996) who described task-based approach from a classroom perspective presented task framework which consists of three phases and indicated learners’ and teacher’s roles at each phase (see Table 11). Different methodological options for each phase have been described in Ellis (2006), Skehan (1998a) and Willis (1996), among others.

Table 11. Task framework (Willis 1996, p. 155)

1. PRE-TASK	
The students <ul style="list-style-type: none"> • note down useful words and phrases from the pre-task activities and/or the recording • may spend a few minutes preparing for the task individually 	The teacher <ul style="list-style-type: none"> • introduces and defines the topic • uses activities to help students recall/learn useful words and phrases • ensures students understand task instructions • may play a recording of others doing the same or a similar task
2. TASK CYCLE	
2.1 Task	
The students <ul style="list-style-type: none"> • do the task in pairs/small groups. It may be based on a reading/listening text 	The teacher <ul style="list-style-type: none"> • acts as a monitor and encourages students
2.2 Planning	
The students <ul style="list-style-type: none"> • prepare to report to the class how they did the task and what they discovered/decided • rehearse what they will say or draft a written version for the class to read 	The teacher <ul style="list-style-type: none"> • ensures the purpose of the report is clear • acts as language adviser • helps students rehearse oral reports or organize written ones
2.3 Report	
The students <ul style="list-style-type: none"> • present their spoken reports to the class, or circulate/display their written reports 	The teacher <ul style="list-style-type: none"> • acts as chairperson, selecting who will speak next, or ensuring all students read most of the written reports • may give brief feedback on content and form • may play a recording of others doing the same or a similar task
3. LANGUAGE FOCUS	
3.1 Analysis	
The students <ul style="list-style-type: none"> • do consciousness-raising activities to identify and process specific language features from the task text and/or transcript • may ask about other features they have noticed 	The teacher <ul style="list-style-type: none"> • reviews each analysis activity with the class • brings other useful words, phrases and patterns to students' attention • may pick up on language items from the report stage
3.2 Practice	
The students <ul style="list-style-type: none"> • practice words, phrases and patterns from the analysis activities • practice other features occurring in the task text or report stage • enter useful language items in their language notebooks 	The teacher <ul style="list-style-type: none"> • conducts practice activities after analysis activities where necessary, to build confidence

1. Pre-task

The pre-task phase has a number of functions. First of all, it lets the teacher provide an advance organizer of what the students will be required to do and the nature of the outcome they will arrive at (Ellis 2006). Skehan (1998a) points out that pre-task activities serve to introduce new language that the learners can use while performing the task, to mobilize existing linguistic resources, to ease processing load, and to push learners to interpret tasks in more demanding ways.

Skehan (1996a) outlined two broad alternatives for the pre-task phase: an emphasis on the general cognitive demands of the task, and/or an emphasis on linguistic factors. He claimed that since “attentional capacity is limited, and it is needed to respond to both linguistic and cognitive demands”, then “engaging in activities which reduce cognitive load will release attentional capacity for the learner to concentrate more on linguistic factors” (cited from Ellis 2006, p. 25).

Ellis (2006, p. 21) described four options available during the pre-task phase: (1) supporting learners in performing a task similar to the task they will perform in the during-task phase of the lesson, (2) asking students to observe a model of how to perform the task, (3) engaging learners in non-task activities designed to prepare them to perform the task, and (4) strategic planning of the main task performance.

The first and second options involve presenting learners with an audio or videotape of other learners doing the same task and/or a text (oral or written) to demonstrate an ‘ideal’ performance of the task. Both Skehan (1996a) and Willis (1996) suggest that simply ‘observing’ other learners perform a task can help reduce the cognitive load on the learner.

Engaging learners in non-task activities can focus on reducing the cognitive or the linguistic demands placed on the learner. Examples of pre-task language activities can be found in Willis (1996). They include classifying words and phrases, odd one out, matching phrases to pictures, memory challenge, brainstorming and mind-maps (asking learners to brainstorm a list of words related to the task title or topic), thinking of questions to ask, among others. Additionally, Newton (2001) suggests cooperative dictionary search (i.e. asking learners to look up words in their dictionary), and matching words to their definitions as possible ways in which teachers can target unfamiliar or task-essential vocabulary in the pre-task phase. However, “there is always the danger that pre-teaching vocabulary will result in learners treating the task as an opportunity to practice pre-selected words. In the case of task-supported teaching this can be seen as desirable but in the case of task-based teaching it can threaten the integrity of the task” (Ellis 2006, p. 25).

Finally, learners can be given time to plan how they will perform the task, although in Willis’s model planning is a part of task cycle. Strategic planning may also involve the provision of linguistic forms/strategies for performing the task.

2. During task

During task phase involves a variety of options that determine task complexity (Robinson 2005) and can influence the nature and amount of the language that students produce (e.g. availability of planning time, a number of elements, existence of reasoning demands and other factors). These options have been described in the previous section. Here we will mention task

performance options offered by Ellis (2006) that were not included in Robinson's Triadic Componential Framework (Robinson 2005). One of them is allowing the students access to the input data while they perform a task. For example, in a story retelling task students can be permitted to keep the pictures/text or be asked to put them on one side as they narrate the story. Ellis (2006) has argued that this can influence the complexity of the task, as tasks that are supported by pictures and texts are easier than tasks that are not. Another task option is introducing some surprise element into the task. Skehan and Foster (1997) failed to find that it had any effect on the fluency, complexity or accuracy of the learners' language. However, Ellis (2006) notices that requiring learners to cope with a surprise serves as an obvious way of extending the time learners spend on a task and thus increases the amount of talk, and it may also help to enhance students' intrinsic interest in a task.

3. Post-task

The post-task phase has three major pedagogic goals: (1) to provide an opportunity for task repetition, (2) to encourage reflection on how the task was performed, and (3) to encourage attention to form, in particular to those forms that proved problematic to the learners when they performed the task (Ellis 2006, p. 36). Several studies (Bygate 1996, 2001; Lynch & Maclean 2000) have indicated that when learners repeat a task their production improves in a number of ways, for example, in terms of complexity and fluency. Reflecting on the task can have various forms. Willis (1996) recommends asking students to present a report on how they did the task and on what they decided or discovered. Ellis (1997) has proposed student-based evaluations of tasks that can be carried out quickly and effectively using simple questionnaires.

In a post-task, special attention can be paid to language forms. As Ellis (2006, p. 38) points out, although attention to form, in one way or another, can occur in any or all of the phases of a task-based lesson, it is during the post-task phase when "students can be invited to focus on forms, with no danger that in so doing they will subvert the 'taskness' of the task". Some researchers (Long 2015; VanPatten 1996) insist that focus on forms should be completely excluded from a task-based lesson since it has no direct effect on learners' interlanguage systems. However, Willis (1996) and Ellis (2006) allow traditional practice of selected forms, such as substitution, gapped sentences, jumbled sentences, transformation drills, etc. These production practice activities may help learners to raise language awareness and "automatize forms that they have begun to use on their own accord but have not yet gained full control over" (Ellis 2006, p. 40). Willis (1996) proposes language analysis activities and language practice activities as a part of the post-task phase. They involve learners in a study of the language forms that were actually used or needed during the cycle. These activities "give learners time to

systematize and build on the grammar they know already, to make and test hypotheses about the grammar and to increase their repertoire of useful lexical items” (Willis 1996, pp. 102–103).

In the post-task phase, learners’ attention to form can be also drawn through collective review of learner errors committed during task performance, consciousness-raising tasks and noticing activities. The teacher can address learners’ persistent errors with the whole class. A sentence illustrating the error can be written on the board, students can be invited to correct it, the corrected version is written up, and a brief explanation provided (Ellis 2006). Lynch (2001) offers ‘proof-listening’ as a way of conducting a post-task analysis. It is based on repeated playing of a recording of the task.

Consciousness-raising tasks constitute tasks in their own right and, therefore, can be used as the main task in a lesson (for more details see Section 3.4.8). But they can also be used as follow-up tasks to direct students to attend explicitly to a specific form that they used incorrectly or failed to use at all in the main task.

Finally, a number of suggestions have been made for developing noticing activities as a follow-up to a task performance. Fotos (1994) found that dictation exercises that had been enriched with the target structures that students had tackled initially in consciousness-raising tasks resulted in consistent learner attention to the structures. Lynch (2001) recommends getting students to make transcripts of an extract (90–120 seconds) from their task performance as a method for inducing noticing.

2.7 Task-based assessment

As well as other parts of a task-based syllabus, assessment in TBLT is organized around tasks rather than grammar or vocabulary. Long and Norris (2000, p. 600) state that “genuinely task-based language assessment takes the task itself as the fundamental unit of analysis, motivating item selection, test instrument construction and the rating of task performance.”

Brindley (1994, p. 74) defined task-based language assessment (TBLA) as “the process of evaluating, in relation to a set of explicitly stated criteria, the quality of the communicative performances elicited from learners as part of goal-directed, meaning-focused language use requiring the integration of skills and knowledge.” TBLA has generated interest since the early 1990s, primarily in conjunction with the ongoing development of task-based language teaching and the pursuit of developing appropriate testing models for this approach to pedagogy (Norris 2002, 2009, 2016). The introduction of tasks into assessment design has also been motivated by

the need for tests that better represent examinees' abilities to use the language (Mislevy, Steinberg, & Almond 2002).

Selection of tasks for TBLA is based on needs analysis for specific language program contexts and purposes of the assessment. These purposes can be quite different, for example, making decisions about students (such as a quick estimate of language proficiency for placing students into language courses, end-of-unit achievement testing, etc.); classroom feedback (improving instruction, understanding learning outcomes, motivating students, etc.); and program evaluation (demonstrating student learning, revising curriculum, improving teacher performance, etc.) (Norris 2014).

An example of task-based assessment within a school-based curriculum with a description of design, implementation and evaluation of assessment tasks can be found in the manual addressed to classroom teachers of English in secondary schools in Hong Kong (http://cd1.edb.hkedcity.net/cd/eng/TBA_Eng_Sec/index.html). In this educational context task-based assessment involves the direct assessment of learners' performance; it is criterion-referenced; and it focuses on the attainment of specific objectives and improving learning outcomes. Since task-based assessment is criterion-referenced, it is very important to specify the criteria used to evaluate learner performance. The manual includes a list of criteria for assessing a group oral presentation in secondary schools in Hong Kong (linguistic accuracy, fluency, presentation style, content, and collaboration with others) but without specifications on how each dimension is measured.

A more general proposal of criteria for evaluating task performance which does not refer to any particular educational context has been made by Norris (2014). He has distinguished different characteristics of learner performance that can be assessed:

- Performance products: Does something happen, does something get produced?
- Performance steps: Do critical elements of the task happen?
- Performance qualities: linguistic and paralinguistic features (e.g. accuracy, fluency, gesture, pragmatics), content (e.g. meaning), and other attributes (e.g. visual support, efficiency of task completion).
- Performance levels: How well does the task performance meet expectations, standards, etc.?

Long's (2015, p. 332) position is that "in task-based assessment, the focus remains on task completion, with a measurable behavioral outcome. Regardless of missing articles or a wrong preposition here and there, can the student perform the task or not?" In order to assess

students' ability to perform the target task, the exit task, which is the last task in a module of TBLT materials, can be used.

For Long (2015), one of the crucial questions in the task-based assessment is whether and how to assess the language used in task performance. Views on this question differ. The decision depends on intended use of the assessment, that is, on who will use the results and for what purposes. A teacher may choose to penalize students who completed the task albeit with some parts of their oral or written performance ungrammatical and/or sociolinguistically, pragmatically or culturally inappropriate. Others may opt to focus solely on task completion regardless of all these errors. Long's proposal (2015, p. 333) is "a judgment call, better made case by case." As he points out, sociolinguistic and pragmatic errors might be overlooked in some cases, but could potentially be serious when committed by a hotel receptionist, a diplomat or during a business meeting.

Introducing a linguistic "caboose" to a test of task-based abilities has some problems. First, it can "lead to difficult questions regarding the frequency and/or degree of ungrammaticality or inappropriateness that will be tolerated" (Long 2015, p. 333). Second, there is a danger of reorientation of a task-based course to a course which works on language as the object, with a focus on forms. In spite of these problems, it is the linguistic quality of learners' performance that has usually been evaluated by researchers. As Pallotti (2009) pointed out, most studies have assessed complexity, accuracy and fluency within communicative tasks, but very few of them discuss whether communication was successful in achieving its goals. He offered adequacy as another measure of L2 performance that represents the degree to which a learners' performance is more or less successful in achieving the task's goals.

Long (2015) claimed that successful task completion is easy to measure by observing the outcome; however, this process can also involve difficult decisions. For example, if a learner is asked to write an e-mail and invite his/her friends to a birthday party and he or she forgets to mention when the party takes place, can this task be considered as successfully completed? Probably not, since crucial information is missing. On the other hand, the reason for this insufficiency of information is likely to do with the learner's lack of attention, not with task complexity or a lack of necessary linguistic resources in the learner's repertoire.

In spite of some problems and sometimes controversial decisions, task-based assessment offers a lot of advantages (Norris 2014). First, it may demonstrate things that other item types do not, for example, that a learner can do a specific task under specific conditions which is particularly important for professional domains (e.g. test of aviation English), that he or she has a certain proficiency (e.g. certification tests such as TOEFL) or displays specific language

qualities (assessment as a part of classroom practice or SLA research). Second, task-based assessment offers rich opportunities for eliciting and observing language use with meaningful content in context. Third, task-based performance can reveal multiple aspects of language ability and/or development within a single task (e.g. accuracy, complexity, fluency). Fourth, task-based assessment provides relevant and comprehensive frames of reference for learners and teachers to generate feedback on form-function-meaning relationships. Finally, it generates interpretable data regarding the worthwhile outcomes of language learning programs (summative uses, e.g., for accountability, certification, program evaluation).

2.8 Summary of Chapter II

The chapter focused on the key questions of TBLT and presented the most influential approaches to different aspects of task-based syllabus design (selection, sequencing, design, implementation, and assessment) and results of studies on tasks that were carried out in the last three decades.

A number of definitions of a task were analyzed, from Long's (1985) broad non-linguistic view on tasks as things that people do in everyday life to narrower and more precise definitions given by Skehan (1998b), Nunan (2004), and Ellis (2003). They pointed out that a task is distinguished by a primary focus on meaning, relationship to comparable real-world activities and a presence of a clearly defined communicative outcome.

Then the process of task selection as a starting point in task-based syllabus design was described. Long's proposal (1989, 2015) which has been widely accepted in TBLT research was outlined. He claimed that it is needs analysis that determines which tasks to include in the syllabus regardless of the language knowledge tasks require. In contrast to Long's radical position, Ellis (2003) and Skehan (1998a) saw the need for use of focused tasks which draw the learner's attention to the predetermined target structures but in an unobtrusive way. This approach was described as more suitable for morphologically complex languages.

The problem of task sequencing that has received a lot of attention in the recent TBLT studies was also outlined. We saw that both in early proposals (Brown *et al.* 1984; Candlin 1987; Prabhu 1987) and more recent ones (Robinson 2001a, 2007; Skehan 1998a, 2001) task complexity or difficulty has been seen as the main criterion of sequencing tasks. However, researchers proposed different factors that make a task more or less difficult. Skehan's (1998a) and Robinson's (2001a) models of task complexity and their hypotheses regarding the effects of

cognitive load on the distribution of the learner's attention during task performance were revised. It was shown that none of the hypotheses was completely confirmed by empirical studies.

The chapter summarized different approaches to task classification and the key findings of the research on task features. From an interactionist perspective, tasks have been manipulated along the flow of information (one-way versus two-way), the number of solutions that can be reached, and the convergence or divergence of their goals. We saw that such research has been interested in how task manipulation, as measured by the quantity and quality of interactional moves, can lead to acquisition. From an information-processing perspective, studies on the dimensions of task familiarity, the number of elements, amount of planning time, and present and past time reference were outlined. We saw that the manipulation of the cognitive demands that tasks impose on learners may have specific consequences for learners' production, usually assessed in terms of lexical and structural complexity, fluency and accuracy.

Finally, the question of task-based assessment was discussed. In TBLT, the learner's progress is supposed to be measured by his or her capacity to complete an assessment task, although in practice researchers and classroom teachers also (and often exclusively) evaluate the language produced by a learner.

CHAPTER III

FOCUS ON FORMS VS. FOCUS ON FORM

3.1 Introduction

The issue of whether and how second language forms can be learned and taught has certainly been one of the most debated in second language acquisition research (DeKeyser 1995). For a long time language pedagogy has assumed that the teaching of language necessarily involves focusing on forms (primarily grammatical forms), and the principle debate concerned how they should best be taught (Ellis 2001). In the early 1980s Krashen (1981, 1982) claimed that grammar can only be acquired unconsciously from comprehensible input and that teaching grammar or correcting learners' errors has no effect on their interlanguage. However, most experimental studies did show gains in accuracy for instructed groups where specific language forms were somehow targeted in meaning-oriented activities. The concept of focus on form offered by M. Long (1989, 1991) as opposed to focus on forms has had a great impact on SLA theory and research.

This chapter reviews the investigation that has been carried out on focus on form and different possibilities of integrating it into task design (proactive focus on form) and task performance (reactive focus on form). The following questions related to focus on form will be discussed below:

- What is "focus on form"?
- Why focus on form?
- What forms should be focused on?
- What are the ways to focus on form?

In the last section special attention will be paid to focus-on-form techniques (input flood, input enhancement, input elaboration, negotiation of meaning, implicit and explicit feedback, input processing, consciousness-raising tasks and other techniques) and their possible effects on language learning.

3.2 Approaches to teaching language forms

Among the major issues raised by SLA researchers is the controversial question of whether and how to include grammar in L2 instruction. Various investigations and experiments were done in order to find the best solution (Carroll & Swain 1993; De Graaff 1997; De Graaff & Housen 2009; DeKeyser 1995, 1998; Doughty & Williams 1998; Ellis 1990, 2003, 2012, 2016; Fotos 1993, 1994; Fotos & Ellis 1991; Harley 1989, 1993, 1998; Hulstijn & De Graaff 1994; Krashen 1992, 1993; Leeman *et al.* 1995; Lightbown & Spada 1990; Long & Robinson 1998; Long 1991, 2000; Loschky & Bley-Vroman 1993; Pica 1985; Robinson 1995b, 1996a; Schmidt 1990; Sharwood Smith 1993; Skehan 1996a, 1998a; Spada & Lightbown 1993; Swain 1998; Swain & Lapkin 2001; Tomasello & Herron 1988; VanPatten & Cadierno 1993; VanPatten & Sanz 1995; White 1991; Williams & Evans 1998). Long and Robinson (1998) distinguished three main approaches to language teaching: focus on forms, focus on meaning and focus on form. Hulstijn (1995) referred to these approaches as the strong interface position, non-interface position and weak interface position.

Most foreign language teachers strongly believe in the necessity of explicit grammar instruction. Focus on forms through presentations of the rules, drills, error corrections, and learning lists of exceptions by heart is the common practice in L2 teaching, no matter what language is taught and to what group of learners. However, this type of instruction often results in a low level of learner attainment (Skehan 1996b). SLA findings demonstrated that “learners do not move from ignorance of a form to mastery of it in one step. Typically, when a form first appears in a learner’s interlanguage (IL), it is used in a non-target-like manner, and only gradually improves in accuracy of use. It sometimes shifts in function over time as other new forms enter. It quite often declines in accuracy or even temporarily disappears altogether due to a change elsewhere in the IL” (Long 1991, pp. 44–45), a phenomenon known as U-shaped behavior (Kellerman 1985). Consequently, attempts to teach isolated items one at a time fail unless the structure happens to be one the learner can process and is psycholinguistically ready to acquire.

Obvious lack of the expected effect of grammar teaching led some researchers to think that it should be abandoned altogether. The Direct method, introduced at the end of the 19th century, the Audiolingual method that was popular during 1960s, the Natural Approach of the early eighties (Krashen 1982), and some versions of Communicative Language Teaching play down the role of explicit grammar teaching. Researchers who support this “non-interface position” (Felix 1981; Krashen 1982, 1985, 1992; Krashen & Terrell 1983; Paradis 1994;

Schwartz 1993; Wode 1981) see no role for focus on form in language teaching and hold that linguistic competence remains unaffected by instruction. This position was disproved by numerous experimental studies that found considerable gains in accuracy for instructed groups in comparison with uninstructed groups (for reviews of these studies see Doughty & Williams 1998; Ellis 2003; Long & Robinson 1998; Norris & Ortega 2000, 2001). De Graaff and Housen (2009) claimed that instruction has effects on the basic dimensions of SLA: it can accelerate the rate of acquisition, leads to higher ultimate levels of attainment (especially higher levels of complexity and accuracy), and helps (or may even be necessary) to overcome premature fossilization. Immersion programs in Canada also have shown that focus only on meaning is not sufficient to give learners the necessary competence they need to produce target-like forms in the L2 even after many years of study (Hammerly 1987; Harley 1993; Swain 1985; Swain & Lapkin 1989).

The “weak interface” position (Hulstijn 1995) reflects the achievements of SLA theory over the last thirty years and has been largely supported by various investigations since Long (1989, 1991) distinguished between focus on formS and focus on form. The term “focus on form” was first introduced by Michael Long in 1988 in a conference paper to refer to an approach where learners’ attention is attracted to linguistic forms “as they arise incidentally in lessons whose overriding focus is on meaning or communication” (Long 1991, p. 46). It contrasts with a structure-based approach – ‘focus on forms’– which always “entails isolation or extraction of linguistic features from context or from communicative activity” (Long 1991, p. 46). Focus on form consists of “an occasional shift of attention to linguistic code features – by the teacher and/or one or more students – triggered by perceived problems with comprehension or production” (Long & Robinson 1998, p. 23). It involves “reactive use of a wide variety of pedagogic procedures to draw learners’ attention to linguistic problems in context, as they arise during communication, thereby increasing the likelihood that attention to code features will be synchronized with the learner’s internal syllabus, developmental stage and processing ability” (Long 2015, p. 317). The attentional switch can be initiated by the learner as a response to communicative difficulties or can be initiated by a teacher or an interlocutor outside the classroom, usually in the form of reformulation, as an attempt to draw learner’s attention to an alternative way of encoding what he or she just said.

Ellis (2016) offers a different perspective on focus on form. He notes that it is a malleable construct that has ‘stretched’ over time and expanded well beyond Long’s initial description. Ellis offers an updated definition of the construct based on the view that the term is best used to refer to specific kinds of ‘activities’ or ‘procedures’ rather than to an ‘approach’. He argues that

focus on form can be utilized in instructional contexts where more traditional structure-based instructional approaches (i.e. focus on formS) have been the norm.

Although Long (2015, p. 317) insists that focus on form should be reactive, that it is “a response to a difficulty – a missing vocabulary item, a problematic verb ending, and so on – due to a learner’s current incomplete stage of L2 development,” proactive focus on form based on predicting learning problems and errors and using a variety of techniques (e.g. input flooding, input enhancement, input elaboration, and task-essential language) proved to be efficient as well. Teachers can plan in advance to ensure that a focus on form will occur, or they can wait for a pressing learner need to arise and develop an “on-the spot” focus on form in response (Doughty & Williams 1998).

Long (1991) argued that focus on form has beneficial effects on IL development. It speeds up the rate of learning, affects acquisition processes in ways possible beneficial to long-term accuracy and it appears to raise the ultimate level of attainment. Focus on form is expected to lead to “effortful retrieval of the missing item from long-term memory, to the learning of a new item, or at least to noticing the item in the input” (Long 2015, p. 317).

Drawing L2 learners’ attention to target forms in primarily meaning-oriented situations is expected to help learners to notice the difference between their current interlanguage and the target language. Swain (1998) distinguishes different levels of noticing. Learners may simply notice a form in the target language due to its frequency. They may also notice their own linguistic inadequacies, that is, the difference between their interlanguage and target language (“notice a hole”). Finally, learners may notice the difference between what they want to say and what they are able to say (“notice the gap”). Doughty and Williams (1998) argue that noticing and the order of acquisition depend on salience in the input and communicative function or meaningfulness in the output. If learners notice certain forms, for whatever reason, for example, frequency or unusualness, they are more likely to acquire them than forms that were not noticed (Robinson 1995a; Schmidt 1990; Skehan 1996a). But, even if a learner notices a form, without a communicative need, acquisition may be delayed.

The key findings of research into focus on form were analyzed by Norris and Ortega (2000) and Goo *et al.* (2009). Norris and Ortega (2000) in their statistical meta-analysis reviewed 49 experimental and quasi-experimental studies conducted between 1980 and 1998 in laboratory-type settings (de Graaff 1997; DeKeyser 1995; Robinson 1996a) and in real classrooms (Doughty & Varela 1998). The studies were conducted largely using English as the target language of instruction (46%), with adult learners (79%) who were typically of low proficiency (36%), and in university settings (65%). The analysis showed the following:

1. Focused L2 instruction results in large target-oriented gains.
2. The effectiveness of L2 instruction is durable, although the effectiveness of focused instructional treatment did seem to decrease from immediate post-test to delayed post-test, but not significantly.
3. Explicit types of instruction are more effective than implicit types.
4. Focus on form and focus on forms interventions result in equivalent and large effects. Focus on form treatments produced slightly larger effect sizes than focus on forms treatments.

Speaking about outcome measures Norris and Ortega (2001) distinguish four ways of how gains in accuracy are usually measured: 1) metalinguistic judgments (participants are required to evaluate the grammaticality of L2 target structure, e.g. grammaticality judgment tests), 2) selected response measures (participants are asked to choose the correct response from a range of alternatives, e.g. multiple choice tests or comprehension questions), 3) constrained constructed response (participants are required to produce the target forms under highly regulated circumstances where the use of the appropriate form is essential for grammatical accuracy to occur), 4) free constructed response (participants have to produce language with relatively few constraints and with meaningful communication as the goal for L2 production, e.g. oral interviews, written compositions).

This variety of outcome measures makes the comparison of results across different studies difficult since different measures can produce different results. For example, the ability to judge a sentence as grammatical or ungrammatical does not necessarily correlate with the ability to produce the target feature correctly (Ellis 2001). A lack of standardization in outcome measures “obscures comparisons of treatment effectiveness from study to study” (Norris & Ortega 2001, p. 198). Another problem with these experimental studies was pointed out by Doughty and Willis (1998) who argued that many studies may not have revealed the full effect of instruction, since measures are generally in terms of accuracy. Finally, Long (2015) listed a number of limitations of Norris and Ortega’s study (2001) and of experimental comparative studies on the relative effectiveness of explicit and implicit instruction, in general. These studies tend to be short-term, often lasting just a few hours, and researchers tend to select simple target structures, which favor explicit instruction.

Goo *et al.* (2009) included 34 studies in their meta-analysis, 11 from Norris and Ortega’s research (2000) and 23 new studies published between 1999 and 2011. They used stricter criteria for inclusion of the studies in the analysis and considered only research treatments that had involved comparison of performance in both explicit and implicit conditions with that of a control group. The results were similar to those obtained by Norris and Ortega (2000). There was

a large mean pre-to-post effect size for instructed condition. Implicit instruction was less effective than explicit instruction, although it still yielded a large effect size on immediate post-tests, unlike what Norris and Ortega found.

3.3 Criteria of the choice of linguistic forms

Although most researchers agree that focus on form has a positive influence on learners' accuracy they point out that some forms do not need or may not benefit from instructional focus (De Graaff 1997; DeKeyser 1995, 1998; Doughty & Williams 1998; Ellis 1990, 2003; Fotos 1993, 1994; Fotos & Ellis 1991; Harley 1989, 1993, 1998; Hulstijn 1995; Hulstijn & De Graaff 1994; Long 1991, 1998, 2000a; Long & Robinson 1998; Loschky & Bley-Vroman 1993; Robinson 1995a, b, 1996a; Schmidt 1990; Williams & Evans 1998). One of the reasons is that some fundamental principles of language (such as the principles of binding and structure-dependency) need not be learned by L2 learners because they have already been implicitly acquired during first language acquisition. These principles belong to the core of Universal Grammar (Felix & Weigl 1991). At the same time, it is likely that explicit teaching can foster the acquisition of many grammar rules outside the core of Universal Grammar (Hulstijn 1995).

Very different and sometimes opposite conclusions were made about the criteria of choosing linguistic forms that would benefit from being given explicit or implicit attention. These criteria include, among others, rule difficulty, reliability and scope of the rules, consideration of L1 influence, and analysis of frequent learner errors.

The fact that some rules are easier for learners than others is not only intuitively recognized by teachers of foreign languages but also supported by SLA research (De Graaff & Housen 2009). The work on developmental patterns in L2 acquisition also suggests that some linguistic features are inherently easier to learn than others (Ellis 1997). Rule difficulty doesn't only influence a choice of forms in focus but also determines focus-on-form techniques.

Hulstijn (1995) described different factors which influence rule difficulty from the point of view of learners. Among them are the learner's prior metalinguistic knowledge, contrasts between foreign and first language, duration of acquisition, reliability (the number of exceptions), and purely formal vs. formal-semantic distinctions. When a form distinction corresponds with a meaning distinction (e.g. plural –s), it may be less difficult to acquire than when there is no corresponding semantic distinction (e.g. third person singular –s).

De Graaff and Housen (2009) added the following additional factors which make an L2 feature 'complex' or 'difficult':

- morphological *vs* syntactic *vs* lexical *vs* phonological *vs* socio-pragmatic
- salient *vs* non-salient
- frequent *vs* infrequent
- meaningful *vs* redundant
- semantically/functionally transparent *vs* opaque
- structurally simple *vs* structurally elaborated
- marked *vs* unmarked
- prototypical/core *vs* peripheral
- similar *vs* different from L1 feature
- developmentally early *vs* late
- cognitively demanding *vs* undemanding

The idea of easy and difficult rules was offered by Krashen (1982) who made a distinction between rules that are easy to acquire but hard to learn, and rules that are easy to learn but hard to acquire. The first category of rules has formal complexity and the latter one is considered to be functionally complex because of the multiplicity of form-function relationships (Doughty & Williams 1998; Loschky & Bley-Vroman 1993). Ellis (1990) made a similar pair of distinctions. On the one hand, he distinguished between simple and complex processing operations that correspond to formal complexity and, on the other hand, between transparent and opaque form-function relationships (functional complexity). Different combinations of formal and functional complexity are possible. For example, third person *-s* in English is formally simple but functionally complex because it expresses several semantic concepts at the same time (present tense, singular, third person), and the rule has a number of high-frequency exceptions (modals). Rules can be both formally and functionally complex, for example, the articles in German that contain the information about case, number, gender and definiteness and have special forms for different combinations of these meanings.

Doughty and Williams (1998) also distinguish semantic complexity. It corresponds to the rules for use of forms that are formally simple, but tied to both semantic and discursal concerns that are too numerous for the learner to grasp easily. These are semantically complex rules. English articles or Preterit and Imperfect in Spanish are examples of semantic complexity.

Although rule complexity is a likely criterion for focus on form, complexity can be hard to define. One of the reasons is that it is quite relative, especially, when it depends on learners' first language. For example, the acquisition of English articles will not have the same level of

difficulty for learners who have the article system in their L1 and for native speakers of languages where the articles don't exist.

Once the question of rule complexity is decided, researchers meet another challenge: how to teach each type of rule. Opposing opinions exist in the SLA literature. On the one hand, Hulstijn and de Graaff (1994) and de Graaff (1997) argue that instruction is more effective for difficult L2 features. Hulstijn and De Graaff (1994) suggest that easy rules are relatively straightforward for learners to discover on their own, and perhaps not the best use of instructional time. On the other hand, a number of researchers (DeKeyser 1995; Pica 1985; Robinson 1996a; Williams & Evans 1998) claim that it is the acquisition of easy rules that benefits from instruction. For complex rules, Doughty and Williams (1998) suggest that they must be learned by "feel" and pedagogical intervention is best limited to a few rules of thumb, coupled with repeated exposure to illustrative examples.

Other criteria of choosing linguistic forms offered by Hulstijn and De Graaff (1994) and Hulstijn (1995) are reliability and scope of the rule. Reliability of the rule is the extent to which it holds true in all the cases to which it applies. An example of a reliable rule could be comparative forms of English adjectives and adverbs. Scope of the rule refers to the number of cases in which it can be applied or how many lexical items it covers. An example of a rule with large scope would be plural *-s* in English or gender marks in Spanish (most nouns ending in *-o* are masculine, whereas words ending in *-a* are feminine).

Hulstijn and De Graaff (1994) consider that instruction is useful for rules with both high reliability and wide scope. This statement leaves unclear what to do with rules with low reliability or how to focus on them if not with instruction. Perhaps a traditional drilling as a part of language focus could be efficient.

First language influence should also be taken into consideration when deciding what forms to focus on (Zobl 1980). Doughty and Williams (1998) suggest various ways of how a learner's first language can influence the L2 acquisition process. First of all, the L2 may contain potentially misleading information, so that learners may assume that their L1 forms are directly transferable to the L2. This is a very common phenomenon that affects all levels of language: phonetic, lexical, morphological, and syntactic.

The opposite situation, when forms of L2 don't exist in learners' L1 (e.g. the Russian case system for L1 Spanish learners), causes even more difficulties especially when these forms are semantically complex. The meanings they transfer can be expressed by other items of the learners' first language (e.g. lexically, by word order or by special syntactic structures) or do not have explicit ways of expression at all.

Ellis (1993) argued that learners' errors should be targeted as the forms for instruction, suggesting a need for on-line modification of the syllabus as these errors occur. Loschky and Bley-Vroman (1993) also supported the idea that known areas of learner error/processing difficulty should be the primary targets for focus on form. In addition, Ellis (1993) suggested marked features should receive explicit instruction, since unmarked features may be learned by most learners naturally. This suggestion, as Robinson (1998) points out, begs at least two questions: which definition of markedness is to be adopted, and whether it is true that unmarked features are learnt naturally, without being explicitly attended. Schmidt (1990, 1995) would answer no to the last question, since he argues all learning requires focal attention accompanied by awareness of the form of input.

3.4 Focus-on-form techniques

Focus-on-form techniques have different degrees of explicitness and obtrusiveness that depend on their purpose. Following DeKeyser (1995), an L2 instructional treatment is considered explicit if rule explanation comprised any part of the instruction or if learners were directly asked to attend to particular forms and to try to arrive at metalinguistic generalizations on their own. When neither rule presentation nor directions to attend to particular forms were part of a treatment, that treatment is considered implicit. Explicit teaching aims to "direct learners to reflect upon a particular linguistic form and to exploit pedagogical grammar" (Doughty & Williams 1998, p. 232). The aim of implicit focus on form is to attract learner attention and to avoid metalinguistic discussion. The main differences are summarized in Table 12.

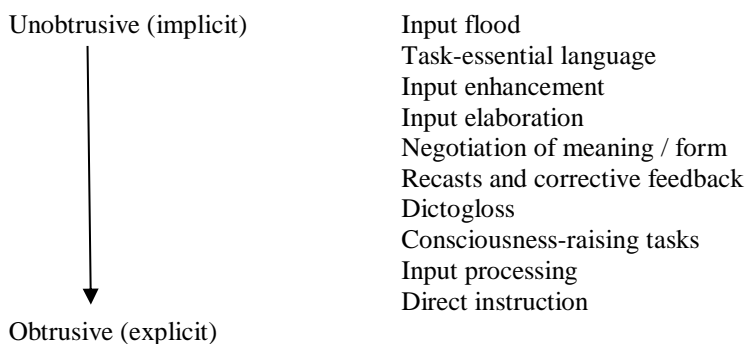
Table 12. Implicit vs. explicit instruction (De Graaff & Housen 2009)

Implicit form-focused instruction	Explicit form-focused instruction
Attracts attention to form	Directs attention to form
Form presented in context	Form presented in isolation
Unobtrusive (minimal interruption of communication of meaning)	Obtrusive (interruption of communication of meaning)
No metalanguage	Use of metalanguage
Free use of target form in spontaneous communication	Controlled practice of target form in focused tasks

Figure 7 shows main focus on form activities arranged according to the degree of obtrusiveness, that is, the degree to which the focus on form interrupts the flow of

communication. Further, a description of these focus-on-form techniques and a brief summary of studies related to their effect on learning different language forms are given.

Figure 7. Focus on form activities according to the degree of obtrusiveness (Doughty & Williams 1998, p. 258)



3.4.1 Input flood

Input flood is a higher presence of target items without rendering the task artificial or contrived. The target form is not emphasized in any way, but it is assumed that its frequency in the input makes it more salient and increases the likelihood of being noticed by students leading to incidental learning of this form without seeming unnatural or disrupting meaning-based activities (Gass 1997; Long & Robinson 1998; Schmidt 1993; Williams & Evans 1998). In the example below the text was flooded with the verbs *ходить* / *ездить* (to go / to walk / to ride) (Chernyshov & Chernyshova 2009, p. 165):

Было время, когда люди только **ходили** пешком, потом **ездили** на лошадях. Сейчас во многих странах модно **ездить** на велосипеде. В последнее время люди больше и больше **ездят** на машинах: на работу, на отдых, в гости и даже в соседний магазин. Сколько же **ходит** пешком современный человек? Вот что говорят об этом английские специалисты: средний европеец **ходит** за свою жизнь пешком 80 500 километров.

There was a time when people only **went** by foot, then **rode** a horse. Nowadays in many countries it is in fashion to **ride** a bike. More and more people **go** by car: to work, to vacations, to see their friends and even to the next shop. How much does a modern person **walk**? The English specialists say an average European person **walks** 80 500 kilometers in his lifetime.

Different studies have been carried out in order to investigate the extent to which input flood is able to foster SLA (Harley 1998; Hernández 2011; Spada & Lightbown 1993; Trahey &

White 1993; White 1998; Williams & Evans 1998, among others). Since input flood is typically combined with other focus on form techniques like textual enhancement (White 1998) or rule presentation (Hernández 2011), the impact of the flood itself is difficult to determine.

Hernández (2011) compared the effect of input flood alone and input flood together with explicit instruction on promoting the use of Spanish discourse markers by English-speaking learners. He found no difference between the two groups. Both of them made significant improvement in their pre-test to post-test scores as compared to the control group where no input flood or explicit instruction was used.

In order to find out which type of enriched input works best, White (1998) compared the effects of (1) a typographically enhanced input flood plus extensive listening and reading, (2) a typographically enhanced input flood, and (3) typographically unenhanced input flood, e.g. input flood alone. She found that all three groups improved equally in the use of the target form, the possessive determiners *his* and *her*, which means that all three types of input were efficient.

However, in other studies (Harley 1998; Williams & Evans 1998) floods of target grammar forms did not give such positive results. In Harley's experiment (1998) with gender assignment in L2 French the author concluded that input flood didn't work because there was little semantic or communicative motivation for gender assignment, which means that failure to assign gender accurately does not result in loss of meaning. Williams and Evans (1998) investigated the effect of input flood and explicit instruction on acquisition of participle adjectives and passive forms. The findings for the two target features differed. ANOVA on the post-test scores revealed that for participle adjectives explicit instruction was more effective than input flood or unflooded texts. For passives, the results were different in different tasks. In the sentence completion task both the instruction group and flood group showed significantly greater increases than the control group but were not significantly different from each other. However, in the narrative task no significant difference was found between the three groups. Qualitative analysis of the dictogloss task showed that learners in all the groups almost didn't use passives. Williams and Evans explained it by the difficulty of the passive form and by the fact that learners were not ready to perceive it. These findings demonstrated that not all forms are equal in terms of effectiveness of focus on form activities. Williams and Evans (1998) considered the interaction of instructional treatment, form type, and learner profile to be an important factor of effectiveness.

Trahey and White (1993) examined whether input flood was sufficient to enable French-speaking learners to learn English adverb placement, in particular, that English allows SAVO word order, but not SVAO which is permitted in French. Learners were provided with positive

evidence showing all the possible word orders in English, but their attention was not drawn to the absence of SVAO. Tests administered after 10 hours of instruction showed that learners succeeded in learning the SAVO position but failed to discover the ungrammaticality of SVAO. Ellis (2003) suggested that input flood may be effective in helping learners acquire new L2 forms but is not very effective in enabling them to eliminate incorrect rules that have entered their interlanguage. It means that both positive and negative evidence are necessary to promote interlanguage development.

3.4.2 Input enhancement

Input enhancement is a method to attract learner attention that consists of flagging target items in different ways (Sharwood Smith 1991, 1993), for example, through manipulation of typography (**larger** type sizes, different typefaces) or use of typographic cues (*italic*, **bold face**, CAPITAL LETTERS, underlining, **color coding**, etc.).

Enhanced texts are used to make a target form perceptually salient to the extent that processing this form is facilitated. According to Schmidt's Noticing Hypothesis (1995, 2001), for a form to be processed for acquisition by L2 learners, it must first be noticed. "Enhancing input using typographical techniques increases the chance that the visually prominent input will be noticed and will thus establish a trace in long-term memory" (Lee & Huang 2008, p. 308).

Although Sharwood Smith (1981) first suggested the potential pedagogical benefits of input enhancement, it was not until 1991 that the first empirical research designed to investigate its actual effects appeared (Doughty 1991; White *et al.* 1991). Since then, a number of studies were carried out in order to compare input enhancement with other focus on form activities (input flood or explicit instruction) and to prove whether it facilitated noticing of the target forms (Gascoigne 2006; Han *et al.* 2008; Izumi 2002; Lee & Huang 2008; Rott 2007, among others). The received results were quite controversial. Some of the studies found comparatively positive effects of input enhancement over input flood for learning grammatical items (Lee 2007; White 1998), whereas others failed to find such helpful effects (Izumi 2002). Explicit grammar instruction has typically been reported to have superior effects than input enhancement alone (Alanen 1995; Doughty 1991).

Some researchers have also shown disagreement on the effects of input enhancement on learners' noticing. Studies of Leeman *et al.* (1995) and Jourdenais *et al.* (1995) provided

evidence to suggest that highlighting forms in the input increases the likelihood of their being noticed and used.

In the investigation of Leeman *et al.* (1995) an attempt was made to draw students' attention to the target usage of the preterit and imperfect tenses in Spanish during the content-based instruction. Two groups participated in the experiment: a focus on form group and a purely communicative group. The difference between the two groups' treatment was that the focus on form group received materials (a reading passage and instructions) where all verbs in the preterit and imperfect were underlined and highlighted with a different color. They were also instructed to pay special attention to the expression of temporal relationships in Spanish and received corrective feedback. The purely communicative group received neither input enhancements nor corrective feedback. The results of the post-test which consisted of 4 tasks (a debate, an essay, a judgment test and fill in the blanks) showed that the focus on form group increased or maintained accuracy in all tasks, while the purely communicative group improved slightly on essays and showed worse results in other tasks.

A similar experiment was described in Jourdenais *et al.* (1995). Participants in the enhancement group received a sample text in Spanish with all preterit and imperfect verb forms highlighted. Participants in the comparison group received the same text with no typographical modification. After that, learners were asked to narrate in the written form the sequence of pictures that they had been given. Analysis of their written production proved that textual enhancement promoted noticing of target forms and had an effect on learners' output. Enhancement participants produced more target features in their written production and used the target forms more frequently in obligatory contexts.

The results of these two studies have provided evidence that input enhancement can be an effective means of drawing learners' attention to target feature of the L2. Similarly, Izumi (2002) claimed that input enhancement was beneficial for promoting learners' noticing, although he did not find any benefit of such noticing in terms of actual learning gains. These findings support Sharwood Smith's claim (1991, p. 122) that "not only is there the distinct possibility that salient input may not be noticed by the learner, the signals may indeed be noticed but with no consequent effect on learning."

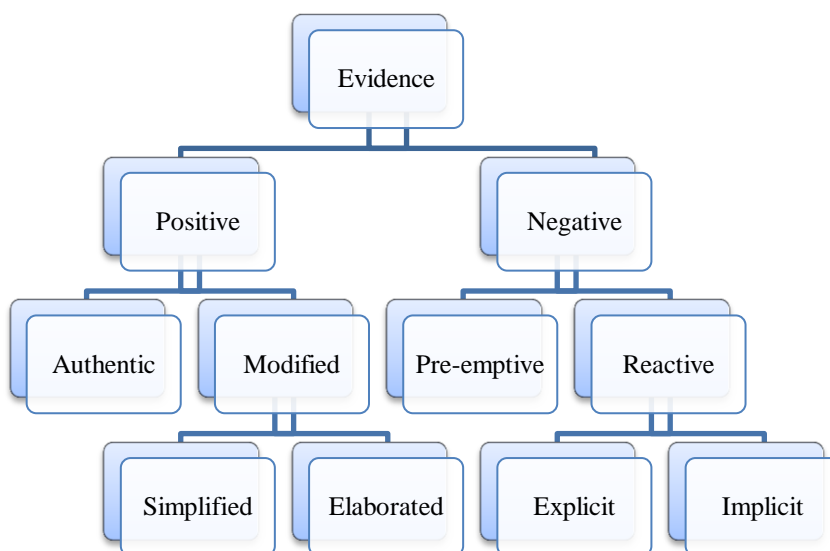
In their meta-analytic review Lee and Huang (2008) presented the results of a systematic synthesis of 16 studies on the effects of visual input enhancement on grammar learning. In order to examine the magnitude of these effects, they calculated Cohen's (1988) effect size *d* values from each individual study. The study demonstrated that students provided with enhanced versions of texts improved from the pretest to the posttest by a medium-sized effect ($d=0.55$). It

means that input enhancement can increase the chances that learners will pay attention to the target forms.

3.4.3 Input elaboration

Before describing input elaboration, which includes modification of positive evidence which learners are exposed to, it seems useful to present taxonomy of evidence types that exists in SLA research (see Figure 8). Evidence can be positive or negative. If positive, it can be either authentic or modified. If modified, it can be simplified or elaborated. Negative evidence can also be of two types: pre-emptive (occurring before an actual error has been made) or reactive. If reactive, it can be explicit or implicit. Explicit negative evidence is an overt correction. Implicit negative evidence can result in a communication breakdown or in a recast. This section will focus on three input types: authentic, simplified and elaborated.

Figure 8. Evidence types (Gass 2003)



Simplification usually involves “relatively short utterances or sentences, a limited range of relatively high frequency vocabulary, a low ratio of dependent to main clauses, and a narrow range of syntactic constructions and verb tenses” (Long 2015, p. 250). As Long (2015, p. 250) notes, simplified materials tend to be “stilted, repetitive, dull and of limited value for language development.” Authentic, or genuine, materials, such as song lyrics, news broadcasts, films, newspaper articles, are originally created largely by and for native speakers. Obviously, this has

the potential to result in excessive linguistic difficulty. Another problem with authentic materials is that although the sources of texts are authentic, the uses to which they are put in classrooms often are not (Widdowson 1978a).

Input elaboration is a technique used for introducing and drawing learner’s attention to specific items that are new and that can be difficult for understanding. These items are not eliminated as in simplified texts, but their comprehension is facilitated comparing with the real, unmodified text. In this way learners get exposure to target vocabulary and structures while still understanding the text. Elaboration involves adding redundancy and regularity to a text. Redundancy is achieved by such devices as repetition, paraphrase, and provision of synonyms of low frequency lexical items, among others. Regularity is attained through parallelism, more frequent use of canonical word order, full noun phrases instead of anaphors, etc. (Long 2015). In Table 13 the examples of three text types are presented.

Table 13. Examples of simplified, authentic and elaborated input (Gilabert 2013)

Simplified text	Real, unmodified text	Elaborated text
The story begins when Norman Bates, who is an <u>employee</u> in a small motel, is <u>looking through</u> a hole in the wall of his motel office.	The story begins when Norman Bates, who runs a small motel, is peeping through a hole in the wall of his motel office.	The story begins when Norman Bates, who runs or <u>is responsible for</u> a small motel, is peeping through or <u>secretly looking through</u> a hole in the wall of his motel office.

Studies of listening and reading comprehension (Kim 2006; Oh 2001; Yano, Long, & Ross 1994) have demonstrated that exposure to elaborated input resulted in improved comprehension, and often comparable (statistically non-significantly different) levels of comprehension to groups exposed to simplified versions of the same texts. In the study carried out by Yano, Long, and Ross (1994), 483 Japanese college students read 13 passages in one of three versions: genuine, simplified, or elaborated. Comprehension, assessed by 30 multiple-choice items, was highest in the simplified group, but not statistically significantly higher than in the elaborated group.

Kim (2006) carried out a complex study in order to investigate whether different types of lexical elaboration and typographical enhancement affect Korean learners’ acquisition of English vocabulary. The typographical enhancement had two levels, enhanced and unenhanced, and the lexical elaboration had three levels, explicit, implicit, and unelaborated. Kim found that explicit and implicit lexical elaboration did not differ in their effect on word recognition: lexical elaboration promoted meaning recognition of vocabulary, but not form recognition. The typographical enhancement of the text in addition to either explicit or implicit lexical elaboration did not seem to affect the acquisition of the previously unknown words.

Oh (2001) investigated the relative effects of two types of input modification – simplification and elaboration – on 180 Korean high school students’ EFL reading comprehension. Learners received six reading passages in one of three forms: (1) baseline, (2) simplified, or (3) elaborated. Comprehension was assessed by a multiple-choice test and by students’ responses on a 6-point unipolar scale. The results supported the suggestion that input should be modified in the direction of elaboration rather than by artificial simplification. Oh (2001) argued that elaboration retains more nativelike qualities than simplification and it is equally or even more successful in improving comprehension.

Comparing three types of positive evidence, Long (2015, pp. 258–259) concludes that “elaboration does its work without removing from the input the very items to which students must be exposed if they are to progress. Genuine texts contain those items, too, of course but are simply too complex for all but ‘advanced’ learners, and so largely unusable as input for acquisition. The undeniable improvement in comprehension that simplification achieves comes at a high cost where language acquisition is concerned – removal of many, usually most, of the learning targets. <...> Elaboration, conversely, achieves nearly as great an improvement in overall comprehension as simplification while retaining almost all unknown material, meaning that new language is available for acquisition.”

3.4.4 Task-essential language

The term “task-essential” belongs to Loschky and Bley-Vroman (1993) who distinguished among three types of involvement of a grammar structure in a task: task-naturalness, task-utility and task-essentialness. In task-naturalness, a grammatical construction may arise naturally during the performance of a particular task, but the task can often be performed perfectly well, even quite easily, without it as happened in experiments with dative alternation in English (Carroll & Swain 1993) or gender in French (Harley 1998).

In the case of task-utility, it is possible to complete a task without the structure, but with the structure the task becomes easier. As an example, Loschky and Bley-Vroman (1993) describe a “spot the difference” task which included locative structures. In Hawaiian language class students completed the task with only a single preposition, but using more locative structures could have helped them to complete the task more quickly.

The last type of grammar involvement in task performance is task-essentialness when the task cannot be successfully performed unless the target structure is used. An example of such a

task is described in Shimura (1990). Participants in his study were asked to identify the reference of reflexive pronouns. Two pictures with two men (“Mr. Thin” and “Mr. Fat”) were shown to them. A learner was presented with a sentence, such as “Mr. Fat believes that Mr. Thin will paint himself.” Then learners were asked to determine whether a given picture matched a given sentence. It was impossible to perform the task successfully without a correct use of the target structure.

There are different views on tasks with task-essential target forms in SLA research. Ellis (2003, 2012) uses the term ‘structure-based tasks’ for this type of task, whereas Long (2015) does not consider them tasks at all. The results of studies on the efficiency of structure-based tasks in eliciting the target forms also differ. For example, Mackey (1999) made use of tasks designed to elicit various question forms in English. The tasks included story completion, picture sequencing, picture drawing, and spot the difference task. The subjects in the study produced a large number of questions of different types during task performance. Mackey found clear developmental gains in questions learners produced, as measured by pre- and post-tests consisting of the same kinds of tasks as during the treatment. However, Samuda (2001) reported that learners in her study, aimed at promoting the use of epistemic modal verbs, failed to use them. Instead, they relied on lexical means to express degrees of probability.

Although task-essentialness is an efficient focus on form technique and it may help to attract learner attention to the target form in an unobtrusive way, to design a task that successfully targets the use of a specific structure can be a real challenge. Firstly, the incorrect use of some forms does not affect task completion, and the task outcome can still be achieved. Secondly, it is not easy to design a task in which the utility of the targeted structure is so clear that learners naturally attend to that structure. Thirdly, the efficiency of a task in eliciting use of the target structure will depend on the learner’s stage of development (Ellis 2003). Learners tend to avoid using the features they find difficult to process (Samuda 2001). Loschky and Bley-Vroman (1993) express doubts that this kind of focus on form will trigger acquisition of new linguistic forms since learners cannot be expected to use a target structure unless they have already internalized it. Consequently, task-essentialness can be achieved only with certain forms and in certain task types, for example, closed comprehension tasks where a specific form can be targeted and used in the input.

3.4.5 Negotiation of meaning / form

According to Long's Interaction Hypothesis (1996), attention to form in meaning-focused communication occurs when learners have the opportunity to negotiate for meaning following a breakdown in understanding. Long (1996, p. 418) defines negotiation for meaning as "the process in which, in an effort to communicate, learners and competent speakers provide and interpret signals of their own and their interlocutor's perceived comprehension, thus provoking adjustments to linguistic form, conversational structure, message content, or all three, until an acceptable level of understanding is achieved." He argues that "negotiation for meaning, and especially negotiation work that triggers interactional adjustments by the native speaker or more competent interlocutor, facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways" (Long 1996, pp. 451–452). Negotiation for meaning serves to highlight linguistic forms that are problematic for learners (Gass & Mackey 2006). There are various ways that meaning or form can be negotiated. Among them clarification request, confirmation check, comprehension check, and language-related episodes can be distinguished.

Clarification request is any expression "designed to elicit clarification of the interlocutor's preceding utterance(s)" (Long 1983, p. 137), for example: "Could you repeat? What?" Confirmation checks are any expressions "immediately following an utterance by the interlocutor which are designed to elicit confirmation that the utterance has been correctly heard or understood by the speaker" (Long 1983, p. 137). Comprehension checks are used to anticipate and prevent a breakdown in communication (e.g. "Do you know what I mean? Do you understand?"). Language-related episodes are defined as "any part of a dialogue in which students talk about the language they are producing, question their language use, or other- or self-correct" (Swain 1998, p. 70).

A number of studies in which the effects of negotiation of meaning on acquisition were examined (Ellis & Takashima 1999; Mackey 1999; Nobuyoshi & Ellis 1993) indicated that it results in acquisitional gains. Mackey (1999) designed different tasks focused on question forms in English. She observed that although learners sometimes had difficulty producing a particular question form during interaction, they negotiated the meaning and were able to reformulate the question making it more target-like or at least comprehensible to their interlocutor. Ellis and Takashima (1999) provided evidence that clarification requests led to more accurate use of past tense forms. However, learners who just listened to other students performing the tasks and reformulating did better in subsequent tasks than those learners who actually spoke during the

treatment. Ellis and Takashima suggested that the “modified input” received by the listeners was more beneficial for learning than “modified output” produced by the speakers.

Nevertheless, the efficiency of negotiation of meaning in drawing learner’s attention to the gap between their production and the target language has not always been supported by empirical research. For example, Foster (1998) reported a classroom observation of the language produced by intermediate EFL students engaged in required and optional information exchange tasks in both dyads and small groups. The results showed that many students in the small groups did not speak at all, many more in both dyads and small groups did not initiate any negotiated interaction, and very few students in either setting produced any modified utterances. The setting of the study within a classroom, as opposed to a laboratory study, was considered as a significant variable. Foster suggested that negotiating for meaning is not a strategy that language learners are predisposed to employ when they encounter gaps in their understanding.

However, statistical meta-analysis performed by Mackey and Goo (2007) showed significant positive effects of interaction on the acquisition of targeted linguistic structures. In particular, Mackey and Goo (2007, p. 405) reported that 28 interaction studies which were qualified for their meta-analysis showed “large mean effect sizes across immediate and delayed post-tests, providing evidence of short-term as well as longer-term effects on language acquisition.”

3.4.6 Task types promoting negotiation of meaning

Focus on form can be integrated into the task type. According to Pica, Kanagy and Falodun (1993, p. 17), a task which promotes the greatest opportunities for learners to experience comprehension of input, feedback on production and interlanguage modification is one which meets the following four conditions:

1. Each interactant holds a different portion of information which must be exchanged and manipulated to reach the task outcome.
2. Both interactants are required to request and supply this information to each other.
3. Interactants have the same or convergent goals.
4. Only one acceptable outcome is possible.

Following these criteria, the most effective task types are jigsaw and information-gap tasks. Jigsaw (Geddes & Sturtridge 1979) is a task in which each participant has part of the necessary information and must exchange it in order to perform the task. Participants work

convergently toward a single task goal. For example, the members of the pairs work to create a text based on a series of pictures. Each student looks only at the pictures that he/she holds. Then they exchange the information to form a coherent whole and write the text in collaboration.

Information-gap tasks are very similar to jigsaws. The idea is that one participant holds information that the other does not know, but needs to know in order to complete a task. Pica, Kanagy and Falodun (1993) insist that it differs from the jigsaw because only one interactant holds crucial, task-relevant information and the other must request it. The gap in the distribution of information results in a one-way flow of information, for example, when students are given a list of questions to conduct an interview with a classmate.

Another task type that generates a great amount of attention to form and encourages students to reflect on their own output is dictogloss (Alegría de la Colina & García Mayo 2007; Kowal & Swain 1997, Swain 1998; Swain & Lapkin 2001; Wajnryb 1990). It consists of reading a short text to the learners at normal speed twice or more times if needed. While it is being read, learners write down some words which can help them recall the text. Then they work in small groups and pool their resources together to reconstruct the text and write a final version. The final versions are then analyzed and compared in a whole class setting. The initial text is intended to provide practice in the use of particular grammatical forms and structures.

Kowal and Swain (1997) transcribed talk of each pair of students doing a dictogloss task. The analysis of transcription proved that the dictogloss created opportunities for metatalk and language-related episodes. Kowal and Swain observed learners noticing things they didn't know ("noticing a gap"), formulating hypotheses and testing them out using different tools (asking each other or their teacher; using their dictionaries and books). However, they noted that the learners did not focus exclusively on the target structure but rather dealt with a wide range of linguistic features.

In the study conducted by Alegría de la Colina and García Mayo (2007), dictogloss and jigsaw tasks generated a great amount of attention to form and reflection on language features. These activities proved to be beneficial for low-proficiency level learners as "they draw attention to form, allow students to solve linguistic problems suited to their needs and make them receptive to feedback" (Alegría de la Colina & García Mayo 2007, p. 110). Swain and Lapkin (2001) investigated the effects of these two task types on accuracy and complexity and came to a conclusion that the dictogloss enhanced accuracy in the production of target forms (pronominal verbs) and led students to notice and reproduce complex syntactic structures. The jigsaw task led to a greater range of vocabulary use and of language-related episodes. The researchers suggested

that perhaps the open-ended nature of this task might inspire greater linguistic creativity but it causes reduction in accuracy.

3.4.7 Corrective feedback

Corrective feedback is another way of drawing learner's attention to a specific form which is widely used in the classroom practice. Researchers usually distinguish explicit and implicit feedback. According to Carroll and Swain (1993), explicit negative feedback would be "any feedback that overtly states that a learner's output was not part of the language-to-be-learned" (1993, p. 361). It involves explicit attention to the targeted feature that can be achieved in various ways, for example, by means of an explicit error correction or a metalinguistic comment or question (Ellis 2003).

Explicit negative feedback was proven to be a necessary measure to draw students' attention to forms that they fail to produce correctly (Carroll & Swain 1993; Lightbown & Spada 1990; Long & Robinson 1998; Roberts 1995; Tomasello & Herron 1988, 1989). Long and Robinson (1998) suggested that if students repeatedly make the same mistake during a task, the teacher can briefly interrupt the group work in order to draw attention to the problem. The interruption is necessary if students' errors are systematic, pervasive and remediable (Long 1991). However, studies on the role of feedback have given contradictory results. Tomasello and Herron (1988) compared two groups of L2 French learners who received different types of input combined with explicit grammar instruction. They found that receiving corrective feedback is more efficient than explicit instruction. Those learners who committed the errors and were given the feedback performed better on a sentence translation task than those who were merely warned about rule exceptions. Tomasello and Herron (1989) argued that students learn best when they produce a hypothesis and receive immediate feedback. The effectiveness of the technique, to the extent that it was demonstrated in these experiments, was supposed to be largely due to the fact that the learners had not simply been told what was incorrect with input from the teachers but led up "the garden path", into error, and produced their own examples. Sharwood Smith (1993) admitting that corrective feedback is vital, interpreted the results of this experiment with more caution. He suggested that learning lexical aspects of second language may be driven by hypothesis formation and negative evidence and other areas of second language may not. The results obtained in Carroll, Roberge and Swain's study (1992) support this suggestion. They investigated the effect of correction on the learning of generalizations regulating suffixation in

French and found that correction had a beneficial effect on word learning but not on the learning of structural generalizations related to word formation rules.

Roberts (1995) didn't find negative feedback as effective as Tomasello and Herron (1989) did. In his experimental study with L2 Japanese he used different error correction techniques such as recasts, clarification requests, repetitions (teacher repeated the question), and confirmation checks. Roberts (1995, p. 180) suggested that "students are only aware of corrective activity in the classroom a fraction of time and even when they are, it is not likely that they understand the nature of the error in many instances." He gave three possibilities of what could happen during error correction. The first one is that students do not notice the corrections. The second suggestion is that they perhaps notice the correction activity on the part of the teacher but do not understand the nature of the error. A third possibility is that they both notice and understand briefly, but only in a brief episode in short term memory that does not result in final intake.

Carroll and Swain (1993) investigated the relative effects of various types of negative feedback, both explicit and implicit, on the acquisition of the English dative alternation by 100 adult Spanish-speaking learners. The subjects in their study were divided into 5 groups (20 learners in each group) according to the type of feedback they received when they made an error. Specifically, upon making an error group A subjects were given explicit metalinguistic information about the generalization (explicit hypothesis rejection). Subjects in group B were simply told that they were wrong whenever they made a mistake (explicit utterance rejection). group C subjects were given a reformulated correct response whenever they made a mistake, giving them a model of the desired response along with implicit negative evidence that their response was incorrect (modeling plus implicit negative feedback). Subjects in group D were asked if they were sure that their response was correct whenever they made a mistake (indirect metalinguistic feedback). Group Z, the comparison group, received no feedback. The subjects were tested twice on the feedback items plus a number of novel items to determine whether they had generalized from the feedback items. Analysis of results revealed that the four treatment groups did significantly better than the comparison group in the first and second testing. Group A which has received explicit metalinguistic feedback outperformed the other groups. On the basis of the obtained results, Carroll and Swain (1993) suggested that adult learners can use feedback to learn specific and abstract linguistic generalizations.

Samuda (2001) found that explicit correction and metalinguistic comments may be helpful and even necessary to ensure that a task achieves its linguistic focus. She designed a task where the target forms were modal verbs to express degrees of possibility and certainty. During

the group discussion, the learners used a number of lexical devices for expressing possibility but not modal verbs. During the report stage when the groups made oral presentations to the whole class, the teacher intervened implicitly by reminding formulaic chunks from the input and explicitly by introducing a modal verb and offering a direct metalingual comment. Only after this explicit focus on form did the learners start to employ the modal verbs the teacher had drawn attention to spontaneously. Commenting on the results of this study, Ellis (2003) underlines that explicit attention to form should not detract from the communicative flow of the task performance since continuous explicit metalinguistic explanation is likely to threaten the primary focus on meaning and destroy the “taskness” of a task.

Implicit negative feedback involves recasts and the strategic use of negotiation of meaning (Ellis 2003) because “learners must infer that the form of their utterance is responsible for the interlocutor’s comprehension problems” (Carroll & Swain 1993, p. 361).

Recast is a type of implicit negative feedback that involves rephrasing an utterance “by changing one or more of its sentence components while still referring to its central meanings” (Long 1996, p. 436). Recasts occur quite frequently and naturally both in meaning-focused classroom activities and communication outside the classroom. In the example below a speaker (the author of this work) used a wrong article with the Spanish word “baguette” (female) while buying bread in a Spanish bakery. Her wrong utterance was followed by a shop assistant’s recast in the form of confirmation check:

Speaker 1: Un baguette, por favor.

Speaker 2: Una baguette?

Speaker 1: Sí, una baguette.

In classroom conditions recasts consist of the repetition of learner’s words corrected by the teacher to draw the learner’s attention to wrong forms. They can be directed at a broad range of linguistic features that have caused some kind of communication problem or can be used to give a focus to a task if they are directed at some pre-determined feature whenever it is used incorrectly by a learner (Ellis 2003).

Recasts have been examined in Doughty and Varela (1998), Long, Inagaki, and Ortega (1998), Mackey and Philp (1998), among others, with results suggesting that they can have a positive effect on acquisition. Doughty and Varela (1998) found that implicit feedback that consisted of recasts and repetitions significantly improved accuracy in the use of English simple past tense. Long, Inagaki, and Ortega (1998) compared the effect of models (sentences that learners listened to and repeated) and recasts on the acquisition of L2 Japanese and Spanish structures. They found no significant gains in the accurate use of the Japanese structure in both groups. However, recasts proved to be more effective than models in the case of Spanish.

Mackey and Philp (1998) investigated the effect of recasts on the acquisition of question forms in English. Their study showed that interaction with recasts was more effective than interaction without recasts in enabling more advanced learners to produce question forms in a series of tasks whereas less advanced learners who were not developmentally ready demonstrated no difference in the two treatments (with and without recasts).

Although in some cases recasts go unnoticed by learners or are misinterpreted as confirmations, research findings have shown that the grammatical, phonological, and especially, lexical, information they provide is noticed in a sufficient number of cases – typically from one- to two-thirds of the time in conversations involving adult learners (Mackey 2012). Long (2015, p. 55) argued that recasts “convey needed information about the target language in context, when interlocutors share a joint attentional focus, and when the learner already has prior comprehension of at least part of the message, thereby facilitating form-function mapping.” It has been suggested that the success of recasts has been a laboratory phenomenon not necessarily found in real classroom lessons (Foster 1998). However, several reviews of the literature, e.g. Long (2007) and Mackey (2012), and three statistical meta-analyses (Russell & Spada 2006, Mackey & Goo 2007, Li 2010) have refuted this charge empirically and have found strong evidence of the efficacy of recasts.

3.4.8 Consciousness-raising tasks

The term “consciousness-raising” was used by Rutherford and Sharwood Smith (Rutherford & Sharwood Smith 1985; Sharwood Smith 1981, 1991; Rutherford 1987) and denoted a deliberate focus on the formal properties of language which could facilitate the development of L2 knowledge.

In SLA theory two types of knowledge are distinguished. The first type has been referred to as explicit knowledge (Bialystok 1981) or declarative knowledge (Anderson 1996). This is “knowledge of the phonological, lexical, grammatical, pragmatic and socio-critical features of an L2 together with the metalanguage for labelling this knowledge” (Ellis 2004, p. 244) and it entails a conscious awareness of how a structural feature works. Explicit knowledge is learnable and verbalizable and “it is typically accessed through controlled processing when learners experience some kind of linguistic difficulty in the use of the L2” (Ellis 2004, p. 245). It is the type of knowledge which is developed through formal instruction. The second type of knowledge has been termed implicit (Bialystok 1981) or procedural (Anderson 1996)

knowledge. This is knowledge of how to do something, for example, how to communicate in a target language (Fotos 1993). Implicit knowledge is held unconsciously and can only be verbalized if it is made explicit (Ellis 2005).

The question about how these two types of knowledge interact remains disputable. Three positions can be identified. According to the non-interface position (Krashen 1981), explicit and implicit knowledge are entirely distinct with the result that explicit knowledge cannot be converted into implicit knowledge. Krashen (1981, 1982) sees implicit knowledge as developing naturally out of meaning-focused communication. This position is supported by research that suggests that explicit and implicit memories are neurologically separate (Paradis 1994). In contrast, DeKeyser (1998) drawing on skill-learning theory argues that explicit knowledge becomes implicit knowledge if learners have the opportunity for plentiful communicative practice. The weak interface position (Ellis 1993) claims that explicit knowledge primes a number of key acquisitional processes, in particular ‘noticing’ and ‘noticing the gap’ (Schmidt 1990). That is, explicit knowledge of a grammatical structure makes it more likely learners will attend to the structure in the input. In this way, explicit knowledge may assist language development by facilitating the development of implicit knowledge. The weak interface position has been used to provide a basis for consciousness-raising (C-R) tasks.

Consciousness-raising tasks (Ellis 1991, 2003; Fotos 1993, 1994; Willis & Willis 1996), or grammar tasks (Fotos & Ellis 1991), are designed to cater primarily to explicit learning and have as a desired outcome awareness of how a specific linguistic feature works (Ellis 2003). Using Schmidt’s (1995) concepts, they are intended to develop awareness at the level of understanding rather than awareness at the level of noticing. C-R tasks make language itself the content and suggest that learners should be deliberately directed to attend to form. In this respect, there can be doubts as to whether C-R tasks are indeed tasks. Ellis (2003, p. 163) argues that they are tasks in the sense that “learners are required to talk meaningfully about a language point using their own linguistic resources.” Any production that occurs during task performance is incidental and does not involve producing sentences containing the target structure as it happens in traditional grammar exercises.

Methodologically speaking, C-R tasks are aimed at learning through problem solving, in accordance with the general principle that what learners can find out for themselves is better remembered than what they are simply told. Ellis (1991) identified the following characteristics of C-R tasks: 1. There is an attempt to isolate a specific linguistic feature for focused attention. 2. Learners are provided with data that illustrate the targeted feature and they may also be provided with an explicit rule describing or explaining the feature. 3. Learners are expected to utilize

intellectual effort to understand the targeted feature. 4. Learners may be optionally required to verbalize a rule describing the grammatical structure.

A C-R task consists of data containing exemplars of the targeted feature and instructions requiring learners to operate on the data in some way. Types of operations include identification, judgment, sorting and verbalization of a rule. In terms of task typology, C-R tasks are convergent, closed tasks with a multiway information gap which requires the exchange of information in order to reach an agreed solution to a problem.

Fotos and Ellis (1991) compared the effects of a C-R task and traditional, teacher-fronted grammar instruction on Japanese learners' ability to judge the grammaticality of sentences involving dative alternation. They found that both methods of consciousness-raising resulted in significant gains in understanding of the target structure. However, a delayed post-test showed that the grammar instruction produced a more durable effect. In other words, the C-R task appeared to have functioned equally well as the grammar lesson in the short term, but was slightly less effective in maintaining proficiency than the grammar instruction after two weeks. Fotos (1994) found no statistically significant difference between these two methods in a follow-up study. These studies proved that C-R tasks result in increased understanding of the targeted feature. Besides, the task completion resulted in quite extensive negotiation of meaning albeit sometimes mechanical in nature.

Fotos (1993) provided evidence that explicit knowledge acquired from completing C-R tasks aids subsequent noticing of the target forms. Three structures problematic for Japanese EFL learners were selected for this study: indirect object placement, adverb placement and relative clause usage. 160 participants were divided into three treatment groups: the grammar task group, who performed three grammar consciousness-raising tasks; the grammar lesson group, who received grammar lessons identical in content to the grammar tasks; and the communicative task group (control group), who performed communicative tasks matched in format, length, and task features, but lacking grammatical content. The first and the second week after the treatment, all three groups were given a dictation exercise with the target structures embedded seven times in 150 words after the first week and four times in 75–80 words after the second week. After the groups had written the dictation, they were given the texts and asked to read the dictation and check what they had written. Then the learners were asked to underline any “special use of English” which they noticed. The grammatical noticing frequencies of the control group were significantly lower than other two treatment groups. Both the grammar task group and the grammar lesson group demonstrated significant levels of noticing of the target structures. Since there were no significant differences between the grammatical noticing

frequencies in these two groups, Fotos (1993) concluded that grammar consciousness-raising task performance is nearly as effective as formal instruction in the promotion of noticing.

Summarizing the value of C-R tasks, Ellis (2003) points out that they are effective in developing explicit knowledge, promoting noticing and providing opportunities for learners to communicate. However, he mentions some limitations of this task type. C-R tasks may not be well-suited to young learners and beginner learners since learners need sufficient proficiency to have a metalinguistic talk about the target features.

3.4.9 Input processing

The theoretical principles and methodology of input processing (IP) were offered by VanPatten (VanPatten 1992, 1996, 2002; VanPatten & Cadierno 1993; VanPatten & Sanz 1995). It involves strategies and mechanisms that promote form-meaning connections during comprehension. The concept of input processing arises from a model of second language acquisition that includes three sets of processes: (1) input processing, (2) accommodation, restructuring, and (3) access, control, monitoring (see Figure 9).

Figure 9. VanPatten's model of SLA (VanPatten & Sanz 1995, p. 170)



IP is concerned with those processes involved in the conversion of input to intake, that is, “the linguistic data actually processed from the input and held in working memory for further processing” (VanPatten 2002, p. 757). The fundamental principles of input processing were formulated by VanPatten (1992) and haven't changed since then. They are the following (VanPatten 2002, p. 758):

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

P1a. Learners process content words in the input before anything else.

P1b. Learners prefer processing lexical items over grammatical items for the same semantic information.

P1c. 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.

P3a. The first-noun strategy may be overridden by lexical semantics and event probabilities.

P3b. 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.

P4a. Learners process elements in final position before elements in medial position.

In order to promote input processing, VanPatten (1992) and VanPatten and Cadierno (1993) offered a particular approach to teaching grammatical forms called processing instruction (PI). It is an explicit focus on form that is input based. PI “seeks to alter the way in which learners perceive and process linguistic data in the input in order to provide the internal learning mechanisms with richer grammatical intake” (VanPatten & Sanz 1995, p. 169).

What is critically different about PI with respect to other treatments that have an input orientation (e.g. textual enhancement, input flood) is that PI first identifies a potentially problematic processing strategy and then “provides activities that push learners away from that strategy. In other words, PI determines not just what is a problem form or structure, but why it is a problem vis-à-vis one of the learning mechanisms involved in SLA” (VanPatten 2002, p. 767). A secondary salient characteristic of PI is that during the instructional phase, learners are never asked to produce the target form. During PI the learner’s job is to process sentences and interpret them correctly while attending to form as well. PI consists of three components which are explicit information about the target form, information about input processing strategies, and structured input activities.

Structured input is purposefully prepared and “manipulated in particular ways so that learners become dependent on form and structure to get meaning and/or to privilege the form or structure in the input so that learners have a better chance of attending to it” (VanPatten 2002, p. 764). Structured input activities can be referential and affective. Referential activities are those for which there is a right or wrong answer and for which the learner must rely on the targeted grammatical form to get meaning. In affective activities learners express an opinion, belief, or some other affective response and are engaged in processing information about the real world.

In order to compare the effect of processing instruction and traditional foreign language instruction on the accuracy in use of the target forms, VanPatten and Cadierno (1993) made an

experimental study. The subjects, 80 English learners of Spanish, were divided into three treatment groups including the control group (no instruction), processing instruction, and traditional instruction. The focus was word order and object pronouns in Spanish. Both instructional groups received explanation about object pronouns and where to place them in the sentence; they both utilized the same vocabulary. The instruction differed in two ways. Firstly, the processing group was told that language learners often misinterpret NVN sequences. Secondly, the traditional group did not practice interpreting sentences, and the processing group did not practice producing sentences. The pretest and the posttest consisted of interpretation tasks (subjects were asked to match each sentence they heard with one of two pictures) and written production tasks. The group that received no instruction showed poor results in all the tests. In the post-test, the processing group made significant gains on the interpretation test, whereas the traditional group did not. On the production test, both the traditional and processing groups made significant gains but were not significantly different from each other. Based on this study, VanPatten and Cadierno argued that altering the way learners process input can alter their developing systems. Moreover, the effects of PI are not limited to processing but also show up on production measures. The general conclusion of the study was that “instruction is apparently more beneficial when it is directed at how learners perceive and process input rather than when it is focused on practice via output” (VanPatten & Cadierno 1993, p. 240).

The limitation of VanPatten and Cadierno’s research acknowledged by the authors themselves is the type of assessment task they used (controlled sentence-level production task). In order to extend this line of investigation, VanPatten and Sanz (1995) studied the effects of PI on different kinds of language output tasks. In this study the PI group was compared to a control group using the same materials focused on object pronouns and word order as in VanPatten and Cadierno (1993). Three output measures were designed to elicit both oral and written production: 1) the same sentence-level task (interpretation test) as used in VanPatten and Cadierno, 2) a structured question-answer interview based on pictures and 3) a video narration task. ANOVA tests showed that the control group that had not received any special instructional treatment did not improve on any of the tests. In contrast, processing instruction significantly affected the grammatical accuracy with which learners interpreted and produced object pronouns. The PI group significantly outperformed the control group in all the tests except the oral video narration test. This was the only case in which a comparison of post-test scores between groups showed nonsignificant results. VanPatten and Sanz concluded that instruction had a positive effect on the subjects’ ability to comprehend and produce the target items.

Various researchers replicated the VanPatten and Cadierno (1993) study using other target structures (e.g. Benati 2001; Cadierno 1995; Farley 2001) and offered evidence that their results were generalizable to other structures. However, some studies reported findings different from those of the VanPatten and Cadierno study. DeKeyser and Solkaski (1996) and Allen (2000) found that there is either no superior effect for PI compared to other instruction or that PI may produce inferior results compared with other instruction and is not generalizable to other structures. Analyzing these studies, VanPatten (2002) argued that the results are different in these studies from VanPatten and Cadierno (1993) not because PI is not efficient, but because the three studies were not replications and had different research questions and methodology.

3.4.10 Direct instruction

Direct instruction, e.g. explicit rule explanation, associated with focus on forms has been excluded from the typical repertoire of teaching strategies in meaning-focused approaches. However, some TBLT theorists suggest that it can be used to draw the learner's attention to a particular form but it must be controlled so that the instruction is a natural part of communicative activity and not just the teacher's metalinguistic talk (Doughty & Williams 1998; Ellis 2003).

A number of studies on the use of pedagogic rules have been carried out (DeKeyser 1995; Robinson 1996a). DeKeyser (1995), following Krashen (1982) and Hulstijn and de Graaff (1994), distinguished two types of rules that are better learnt explicitly or implicitly, suggesting that explicit learning is better for simple categorical grammar rules and implicit learning is equally as good as or even better than explicit learning for linguistic prototypes. N. Ellis (1994, pp. 2–3) defined implicit learning as a “nonconscious and automatic abstraction of the structural nature of the material arrived at from experience of instances” and explicit learning as “conscious searching, building then testing of hypothesis; assimilating a rule following explicit instruction.”

A specially created miniature linguistic system was used in DeKeyser's study (1995). The explicit subjects were presented with the grammar rules of this linguistic system before three sessions out of 20 experimental sessions. The implicit subjects received no explanation of grammar, nor was it ever even mentioned to them that the sentences they saw had underlying grammar rules. After the last learning session all subjects were administered grammaticality judgment and production tests. Their results supported the initial hypotheses. DeKeyser found that explicit teaching of categorical rules (rules that require supplying a morpheme to mark

gender, number, or case) were more efficient than having students induce the rules for themselves. The implicit group's knowledge of these rules was not only much more limited than that of the explicit group's but, in fact, no evidence for implicit learning of abstract patterns was found. This was true for new forms. However, for old forms, the learning condition did not make a difference, "probably because these forms tend to be memorized as unanalyzed wholes rather than constructed on the basis of implicit or explicit rules" (DeKeyser 1995, p. 397). Prototypicality patterns, that is, the prototype structures of the allomorphy, were learnt better by the implicit subjects than by the explicit subjects.

Castellví and Markina (2015) investigated the effects of including a short explicit rule explanation in a pre-task on the acquisition of prepositional constructions in Russian. The results of the study were presented at the TBLT Conference in Leuven, Belgium. The main research question was whether a brief rule explanation before the task performance makes a difference and whether implicit focus-on-form techniques commonly used with English can facilitate noticing in learning Russian and to what extent they are efficient. The following hypotheses were formulated:

1) In the case of implicit instruction, the learner's attention will be focused mostly on meaning (that is, on the prepositions themselves) and, consequently, learners will not pay enough attention to the required morphological cases which will cause inaccurate use of case forms.

2) A short explicit explanation before the task performance will help learners notice necessary forms that will lead to more accurate use of these forms after the exposure to the same amount and type of input as with implicit instruction.

21 low-intermediate Catalan- and Spanish-speaking learners of Russian participated in the study. They were divided into two groups, and during a two hour session they had to perform a task with two different pre-task conditions for each group: in one group (further explicit group) an explicit explanation of the target forms preceded work with the input whereas in the second group (further implicit group) no rule explanation was provided. The explanation consisted of a very brief (approximately 5 minutes) presentation of the prepositions and their associated morphological cases. There was no practice or further exemplification of grammatical forms. It means that the use of an introductory explanation before starting the pre-task activities was the only difference in performance of both groups.

The target structures included 12 prepositions with space meaning: *in, on* + prepositional case; *to the left of, to the right of, near, in the center of, opposite* + genitive case; *in front of, behind, between, under, above* + instrumental case. Most prepositions were new to learners. For

this reason, all students showed very poor results in the pre-test which proved that students in both groups had the same baseline.

The pre-task materials included a scheme with the target items that learners could use as a reference while doing the task, reading and listening. In order to attract learners' attention and make them notice target forms, input flooding and input enhancement were used. Texts for reading and listening were flooded with lexical items needed to describe a room and furniture and prepositions of space. Target items were typed in bold. In the materials received by explicit group, colour enhancement was also used to mark groups of prepositions followed by the same case form.

A find the differences task connected with object placing was used in the study. Learners worked in pairs and each of them had a picture of the same room with some objects missing. They were allowed to look only at their own pictures and not the partner's. They had to ask a partner if he or she had this or that object and if not they had to explain where it was located. The second student had to draw the object and explain what was in his picture and where it was. In the end the learners were asked to report the differences in object location they had found. It should be mentioned that this task supposes a high linguistic complexity for learners of Russian. In contrast to the same task in English, it has an additional grammar load since learners have to deal with case forms which follow each preposition, and have to decline nouns and adjectives which is difficult for low proficiency students as the case forms are still not automatized for production or comprehension.

The task was preceded by a pre-test and followed by a post-test that consisted of a fill in the blanks test (FB) and a multiple-choice test (MC). The fill in the blanks exercise was aimed at checking the knowledge of both prepositions and related morphological cases, and they were evaluated separately. The multiple-choice test was designed as purely grammatical to avoid several possible answers.

In order to check whether there were significant differences between the pre- and post-test in each group, the Wilcoxon signed-rank test for two related samples was performed. The data were analysed in the following different ways: 1) The total results in the pre- and post-test of the implicit and explicit groups, that is, the results of both fill in the blanks (FB) and multiple choice (MC) exercises were analyzed together (FB+MC). 2) The results of the two exercises were analysed separately (Pre-test FB and Post-test FB; Pre-test MC and Post-test MC). 3) The use of prepositions and case forms in the fill in the blanks exercise (FB Prepositions and FB Cases) were analyzed.

In all the cases Wilcoxon signed-rank tests confirmed that there are statistically significant differences between the pre- and post-test both in the implicit and the explicit groups. This means that learning took place under both conditions. An important question is how efficient it was and whether a brief rule presentation made any difference. In order to answer this question, the effect size r was measured using Effect Size Calculator at from the Colorado University website (<http://www.uccs.edu/~lbecker/>). The effect size was proven to be medium or large in both groups according to Cohen’s criteria (1988), but in the explicit group it was higher – more than 0.8 – a result never achieved in the implicit group (see Table 14).

Table 14. Effect size r (Castellví & Markina 2015)

Test	Implicit group	Explicit group
FB+MC	0.69	0.85
FB	0.71	0.89
MC	0.71	0.89
FB Prepositions	0.79	0.93
FB Cases	0.59	0.79

FB – Fill in the Blanks, MC – Multiple Choice

These results mean that our first hypothesis was only partially confirmed. In the group where focus on form was entirely implicit learners showed significant improvements and paid attention not only to meaning but also to the required morphological cases. However, the group who received a short explicit explanation before the task performance showed a larger effect size in all the tests which means that their gains were more significant not only at a grammatical level as was expected but also in the use of prepositions. It proves that this kind of explanation is beneficial as it can help learners notice necessary forms and also facilitate meaning processing.

3.5 Summary of Chapter III

In Chapter III the findings of research into focus on form have been reviewed. The notion of “focus on form” introduced by Long (1989, 1991) was discussed, and its differences from traditional focus on forms outlined. A number of criteria for choosing target forms to be focused on during instruction was listed, including rule complexity, reliability and scope of the rule, influence of learners’ L1, analysis of learners’ errors, and frequency.

It was demonstrated that not all forms are equal in terms of effectiveness of focus on form activities. Explicit focus on form was proved to be better for simple rules than implicit

techniques. The combination of direct instruction plus carefully considered examples of the rules appeared to be the most effective learning condition for complex rules.

A considerable number of studies have proven the efficiency of attention to form within a meaningful context. The chapter included the review of results of two statistical meta-analyses (Goo *et al.* 2009; Norris & Ortega 2000), which confirmed that focused L2 instruction resulted in large target-oriented gains and that its effectiveness was durable.

The chapter presented a variety of ways to draw the learner's attention to particular language forms. A description of the most frequently used focus-on-form techniques and a brief summary of studies related to their effect on learning different L2 forms were given. The analysis of findings reported in most studies showed that groups who received treatment in which the target structures (grammar forms, lexical items, word order) were somehow outlined (through enriched input, corrective feedback, consciousness-raising tasks, input processing, etc.) outperformed control groups who received the same treatment but with no focus on form.

CHAPTER IV

THE EFFECTS OF TASK-SUPPORTED AND TASK-BASED LANGUAGE TEACHING ON L2 LEARNING

4.1 Introduction

In Chapter II, a general theoretical framework of Task-Based Language Teaching has been presented, and steps in task-based syllabus design have been described. Chapter III has dealt with focus on form, one of the fundamental TBLT principles. This chapter consists of three parts. First, two key terms of the current study, “task-supported” and “task-based,” will be distinguished. Second, the research on the evaluation of task-based courses will be reviewed. The chapter will specifically focus on process-product evaluations which have attempted to answer the question as to whether the TBLT approach really works. Study design and the results obtained will be presented for each study. This section will try to answer the following questions:

- i) What are the learners’ and teachers’ perceptions of TBLT and what problems have they encountered with it in real classroom conditions?
- ii) How does each of the compared approaches affect learning of the target items and/or the fluency, complexity, and accuracy of L2 learners’ production?
- iii) Which of the approaches used in reviewed studies proved to be more efficient in terms of the measures used in these studies?

4.2 Task-supported and task-based language teaching

The value of tasks has been recognized by many language teachers, material writers, and course designers. However, they have differed in the way they have employed tasks. Some methodologists “have simply incorporated tasks into traditional language-based approaches to teaching. Others, more radically, have treated tasks as units of teaching in their own right and have designed the whole courses around them” (Ellis 2003, p. 27). These two ways of using tasks are usually referred to respectively as task-supported language teaching (TSLT) and task-

based language teaching (TBLT). Samuda and Bygate (2008, pp. 58–60) also distinguished the task-referenced approach to the role of tasks in instructional design and assessment.

In the task-referenced approach, progress through the syllabus is measured by performance on target achievement tasks. Tasks are principally used as a way of setting achievement targets and assessing the desired outcomes of instruction, as in the case of the Common European Framework of Reference for Languages (2001). However, assumptions about the extent to which tasks may be used in teaching are not made. This approach will not be the subject of this study.

In TSLT, tasks are used as an element in an overall program of instruction within a range of syllabus types, and/or to realize a range of curricular objectives (to develop fluency, to raise awareness of specific linguistic features, to assess progress, to provide practice, to activate prior knowledge, etc.). Tasks are usually used in conjunction with different types of pedagogic activity (exercises, rule explanation, focused practice, etc.). They are “not necessarily used for assessment purposes, and the syllabus itself may be defined by categories other than tasks. <...> Tasks are seen as tools to be exploited by teacher and learners in the service of particular language aims and objectives” (Samuda & Bygate 2008, p. 59).

TSLT usually employs a methodological procedure consisting of presentation-practice-production (PPP). A language item is first presented to the learners with or without an explanation. This item is then practiced in a controlled manner through exercises. Finally, opportunities for using the item in free production are provided. It is in this “production” stage that tasks may be employed (Ellis 2003). However, Ellis (2003) notes that it would be wrong to characterize task-supported language teaching entirely in terms of PPP. It can take other forms. For example, this sequence can be started with the production stage in which tasks would play a diagnostic role.

In TBLT, as we have seen in Chapter II, task is the central unit of analysis in all the stages of syllabus design – from needs analysis to student assessment. “Tasks themselves constitute syllabus content, and lessons are built around them. If they appear in the syllabus, it is for a reason, and there is no covert grammatical syllabus” (Long 2015, p. 305). Samuda and Bygate (2008, p. 196) summarized the characteristics of TBLT in the following way:

- Tasks define the syllabus and drive classroom activity.
- Task performance is a catalyst for focusing attention on form, and not vice versa.
- Assessment is in terms of task performance.
- Task selection is shaped by real-world activities of relevance to learners and their needs.

Tasks themselves can be very similar and quite often the same in TBLT and TSLT. As Samuda and Bygate (2008, p. 215) point out, “there doesn’t appear to be anything specific in the design of the task that would mark it out as either task-supported or task-based”. However, TBLT and TSLT differ in the ways that tasks are selected and used (see Table 15). The completion of the task and achieving an outcome is crucial in TBLT, while in TSLT it is also possible that the aim of the task is to practise structures involved in a communicative situation. It can be an oral drill or a way to facilitate the contextualized use of structures or vocabulary.

For example, at the A1 level in a task-supported course, a “spot the difference” task can be designed to practice the Genitive case in Russian, comparing two pictures, with one of them lacking some objects present in the other. Attention is directed to the vocabulary related to the present or lacking objects and to the forms of the Genitive case. Even though this is good practice, it is a type of oral drill, with the positive and negative aspects that drills entail. In the same task within the TBLT approach the Genitive case may be necessary, but other morphological cases may be as well since a real world situation will usually demand more than one specific grammatical case. For this reason, proper planning of the linguistic content of tasks is needed in order to define the right amount of new items which, on the one hand, will be feasible for learners, and, on the other hand, will ensure progress both in lexical and grammar aspects.

Table 15. Tasks in TSLT and TBLT

	TSLT	TBLT
Role of tasks	<ul style="list-style-type: none"> - means to support delivery of the program - provide possibilities for communication and negotiation of meaning - activate prior knowledge and improve fluency - focus on determined grammar structures and lexical items - create a learning space for the introduction of new language (Samuda & Bygate 2008) 	<ul style="list-style-type: none"> - units of syllabus design - serve as a catalyst for focusing attention on form - achievement of a task goal as criterion of assessment - help learners to see the gap between their language needs and language knowledge
Task selection	-determined by the forms that should be practiced through the tasks	-on the basis of needs analysis
Task sequencing	-depends on the sequence of the main units in the syllabus (structures, lexical items, topics, notions, etc.)	-determined by task complexity
Language focus	<ul style="list-style-type: none"> -focus on forms through explicit explanation and corrective feedback -language focused on both in a meaningful context and in a decontextualized way 	<ul style="list-style-type: none"> -focus on form through task design and feedback -language addressed as it appears in meaning-oriented activity
Classroom activities	-tasks used along with other activities (grammar exercises, drills, etc.) and usually employed in the “production” stage	<ul style="list-style-type: none"> -work on tasks - usually (but not obligatory) includes pre-task activities, task itself and post-task activities
Assessment	-in terms of accuracy	-in terms of successful task performance (whether desired outcome was achieved or not)

The way tasks are selected, sequenced and assessed is not the only difference between TBLT and TSLT. These two approaches are also characterized by different classroom behavior (feedback, negotiation of meaning, turn taking, the learner's and teacher's role, etc.). These differences have been summarized by Ellis (2006) and are presented in Table 16.

Table 16. Classroom behavior in a form-focused (TSLT) and task-based pedagogy (Ellis 2006, p. 88)

	Form-focused pedagogy	Task-based pedagogy
Discourse structure	Rigid discourse structure consisting of IRF (initiate-respond-feedback) exchange	Loose discourse structure consisting of adjacency pairs
Control over topic development	Teacher controls topic development	Students able to control topic development
Turn-taking	Turn-taking regulated by the teacher	Turn-taking regulated by the same rules that govern everyday conversation (i.e. speakers can self-select)
Question types	Display questions (i.e. questions to which the questioner already knows the answer)	Use of referential questions (i.e. questions that the questioner does not know the answer to)
Learner's role	Students placed in a responding role and consequently perform a limited range of language functions	Students function in both initiating and responding roles and thus perform a wide range of language functions (e.g. asking and giving information, agreeing and disagreeing, instructing).
Negotiation of meaning	Little need or opportunity to negotiate meaning	Opportunities to negotiate meaning when communication problems arise
Scaffolding	Scaffolding directed primarily at enabling students to produce correct sentences	Scaffolding directed primarily at enabling students to say what they want to say
Feedback	Form-focused feedback (i.e. the teacher responds implicitly or explicitly to the correctness of students' utterances)	Content-focused feedback (i.e. the teacher responds to the message content of the students' utterances)
Repetition	Echoing (i.e. the teacher repeats what a student has said for the benefit of the whole class)	Repetition (i.e. a student elects to repeat something another student or the teacher has said as private speech or to establish intersubjectivity)

Long (2015) compared synthetic focus-on-forms instruction within the PPP approach and analytic task-based instruction with focus on form (see Table 17). As discussed above, TSLT cannot be associated completely with PPP, but in classroom practice (and in the present study, in particular) it often takes this three-stage form. For this reason, we find the characteristics given by Long relevant for the current work.

Table 17. Distinguishing characteristics of PPP and TBLT (Long 2015, p. 349)

PPP	TBLT
Code-focused course objectives	Task-focused course objectives
No needs analysis	Needs analysis
Generic, commercially published textbooks	Local, program-specific materials
Linguistic syllabus	Task syllabus

Code-focused tests	Task-focused tests
No TBLT-specific training of teachers of students	TBLT-specific training of teachers and students
Focus on code features	Focus on pedagogic tasks
Early forced student output	Substantial, rich, task-related input
Language-like behavior (drills, etc.)	Communicative language use
Display questions	Referential questions
Mostly explicit error correction	Mostly implicit negative feedback (especially recasts)
Meta-linguistic talk	Little or no meta-linguistic talk
Mostly teacher-initiated exchanges	At least some student-initiated exchanges
Low ratio of topic-continuing to topic-initiating moves	High ratio of topic-continuing to topic-initiating moves
Low proportion of student utterances / turns	High proportion of student utterances / turns

One of the problems of TSLT is its breadth of scope and its lack of a firm theoretical base. As Samuda and Bygate (2008) point out, the ways tasks are most widely used have not been widely researched. They also point out an inaccurate tendency to characterize task-supported as a weak (meaning less efficient) form of task-based learning that can lead to devaluation of tasks as a pedagogic construct in its own right. This tendency can be seen, for example, in the work of Long, who has opposed the task-based syllabus to “some sort of bifurcated syllabus, part task, part structural” (Long 2015, p. 333) referring to task-supported language teaching. He admits that task-supported approaches “have merits, including their location within the existing comfort zone of most teachers, state education authorities, and publishers, which can make them more acceptable, and so more likely to be adopted, in the short term” (Long 2015, p. 7), but outlines the superiority of TBLT. Nevertheless, TSLT has been widely used in different classroom settings, whereas full-fledged genuine TBLT programs are still very rare. In the next section, we will focus on the evaluation of task-based courses and comparative method studies in which TBLT has been compared with task-supported language teaching and PPP.

4.3 Evaluation of task-based courses

As Long points out (2015, p. 350), “full-fledged genuine TBLT programs are still few and far between, and satisfactory opportunities to investigate their effectiveness rare.” However, over the past decade, researchers’ and teachers’ interest in TBLT has considerably increased, and task-based programs have been implemented at the national or regional level in a few places, such as Hong Kong, Malaysia, Thailand, Mainland China, Japan, and Flanders in Belgium (Carless 2004, 2007; McDonough & Chaikitmongkol 2007; Mustafa 2008; Van den Branden 2006; Zhang 2007). How well TBLT works in various contexts, and what challenges practitioners might encounter when implementing TBLT in their classrooms have become

pressing research issues. Two types of TBLT program evaluations have been reported: formative evaluations and process-product evaluations.

4.3.1 Formative evaluations

Formative evaluation (Burrows 2008; Li 1998; McDonough & Chaikitmongkol 2007; Ruso 2007; Van den Branden 2006) usually includes qualitative data and tries to address the issue of teachers' and learners' reactions to TBLT, their concerns, levels of satisfaction, correspondence of the course to their learning expectations and communicative needs, etc.

An example of formative evaluation is the study carried out by McDonough and Chaikitmongkol (2007) who investigated teachers' and learners' reactions to a task-based EFL course at Chiang Mai University in Thailand. They collected the teachers' and learners' impressions about the course over a 12-month period during the pilot testing. Thirty-five learner-participants enrolled in the task-based course were first-year students in the English department between the ages of 17 and 19. They were all native speakers of Thai with an intermediate level of English. The EFL teacher-participants (n=13) worked in the English department at Chiang Mai University.

In order to identify the participants' reactions to the task syllabus the researchers used a qualitative analysis of oral and written data elicited through task evaluations, learning notebooks, observations, course evaluations, interviews, and field notes. Learners completed a task evaluation at the end of each task, which consisted of six open-ended questions about the aspects of the task that the learners liked and those that they disliked; the skills, knowledge and strategies they had learned; the real world relevance of the materials; and the teaching approach. The learners were also required to keep a learning notebook and they were encouraged to record any information about the vocabulary, strategies, skills, and tasks that they were learning. Several EFL teacher-participants observed the pilot classes and were asked to comment on the effectiveness of the teaching procedures and materials. The information obtained from task evaluations and observations was used to create the course evaluation, which consisted of 12 open-ended questions about various aspects of the course, including the teaching approach, the materials and activities, the real world applicability of the content, and the tasks.

Two secondary data sources were also used in the study. One of them was informal interviews carried out with learners in the pilot class and teacher-participants who had observed the classes. The other secondary data source was the researcher's (Wanpen Chaikitmongkol)

notes, including her reactions to the teaching materials, which she recorded in a notebook while preparing lessons, during class, and immediately after class.

Analysis of the data revealed that teachers' and learners' impressions about the task-based centered on three themes: (a) increased learner independence, (b) concerns with the course content, and (c) perceptions about real world relevance. Both teachers and learners believed that the task-based course helped learners to become more independent. As for the course content, half of the learners initially expressed concerns about a perceived lack of grammar instruction. Several learners stated that the linguistic forms relevant for Task 1 were structures that they already knew and they did not learn anything from the task. However, by the end of the semester teachers and learners no longer voiced complaints about the amount or type of grammar instruction provided in the task course. Finally, participants recognized the relevance of the task-based course to the learners' real world academic needs but did not comment on its relevance to any specific needs outside an academic context. The participants believed that the skills they had acquired in the task course were applicable to a variety of academic tasks that they carried out in other courses, such as delivering oral presentations, carrying out collaborative projects, and listening to lectures.

Analysis of data also permitted the detection of some learners' and teachers' concerns about the task-based course. First, the participants reported that they needed time to adjust to the task-based teaching approach. Second, the learners found that the amount of teacher support and guidance provided in the course was not sufficient. Finally, the participants had some reservations about the amount of materials used in the task-based course. Some teachers felt that too many activities were assigned for each lesson.

Summarizing the findings of their study, McDonough and Chaikitmongkol (2007, p. 123) pointed out that "Thai EFL teachers and learners generally had positive reactions to the task-based course. They believed that it encouraged learners to become more independent and that it targeted their real world academic needs." In order to improve the course, teachers and learners required activities and information that could help them adjust to task-based teaching, reduction of the amount of materials per lesson, and learners also required more teacher support and guidance to carry out the tasks successfully.

Lai, Zhao and Wang (2011) examined the implementation of TBLT in the context of the online Chinese courses for absolute beginners offered at a public virtual high school in the United States. Their study addressed two research questions: what are learners' and teachers' reactions to TBLT and what issues emerge from the implementation of TBLT in an online context? Data for the study consisted of students' (n=38) weekly self-reflection blogs, students'

course evaluation (three Likert scale questions on learners' enjoyment of the course and four open-ended questions eliciting the aspects of the synchronous sessions that they liked and disliked), researchers' classroom observation notes and recorded synchronous sessions, weekly debriefing meetings with the teachers, and the teachers' end-of-semester interview data.

At the end of the semester, students rated their enjoyment of the course positively (5.64 on a scale of 7) and expressed satisfaction with the amount of learning (5.33 on a scale of 7). The analysis of students' weekly blogs across the 12 weeks revealed that some students went through a shift in mindset. In particular, some students demonstrated a shift from being totally reliant on the instructor for explicit instructions, to taking more and more initiatives and responsibility for learning on their own. However, most learners expressed a preference for explicit instruction.

The analysis of information obtained in weekly meetings with the teachers and the teachers' end-of-semester interview data detected some problems that teachers encountered. Their concerns included: (1) students becoming easily frustrated over the extensive use of the target language, (2) students expecting the language needed for the pedagogical tasks to be pre-taught, (3) some students not being active participants during the group work, (4) some students not actively engaging in guessing and deducing and always waiting teachers to tell them everything they need to learn, and (5) students expecting instructor-led talk rather than that they themselves would play a central role.

This study revealed a series of issues emerging from implementing TBLT in the online course. Lai, Zhao and Wang (2011, p. 94) pointed out that some issues identified have been very much the same as those of TBLT in face-to-face classrooms. For example, like McDonough and Chaikitmongkol (2007), they outlined the need for strategy training to familiarize students with the philosophy and principles of TBLT as well as the potential of TBLT to change students' approaches to learning and facilitating autonomous language learning.

Some other findings have differed from the face-to-face TBLT classroom literature due to the particular nature of the online context. Lai, Zhao and Wang (2011) outlined difficulty in building up students' rapport in the online context since learners did not know each other, whereas Ruso (2007) found that TBLT increased students' rapport during face-to-face lessons. On the other hand, the face-to-face classroom TBLT literature reports that "shy" students and students with low language proficiency may find face-to-face interaction during task performance stressful (Burrows 2008; Li 1998), whereas Lai, Zhao and Wang noticed that text-chatting in the online context helped to decrease learners' stress and anxiety levels.

Other studies exploring the potential of TBLT in various classrooms have reported positive student perceptions of the task-based courses. Ruso (2007) conducted a study on the

implementation of TBLT in two first-year English classes at the Eastern Mediterranean University in Cyprus and reported positive perceptions and increased participation from the students. Lee (2005) experimented with TBLT in a high school in Taiwan over one semester and came to a similar conclusion of positive perceptions and enjoyment. Van den Branden (2006) who summarized the results of the TBLT course of Dutch for immigrants in Belgium also outlined students' enthusiasm for the materials.

However, some studies have raised a note of caution concerning the classroom implementation of TBLT in a few sociocultural contexts (Bruton 2005; Burrows 2008; Littlewood 2007). Burrows (2008) pointed out that the passive learning style of the Japanese students as well as their over-reliance on the teacher collectively weakened the implementation of TBLT in this particular context. Carless (2004, 2007) found that teachers' beliefs, the proficiency levels of the students, and the sociocultural realities of Hong Kong primary schools collectively contributed to teachers' transforming TBLT into task-supported teaching. Other classroom factors that have been identified to challenge classroom-based TBLT were as follows: (a) crowded and cramped classrooms created discipline issues and "uncontrollable" noises (Bruton 2005; Carless 2004; Li 1998); (b) students of different proficiency levels demonstrated unbalanced involvement and contributions – students with higher language proficiency benefited more from doing tasks, whereas students with lower language proficiency and with shy personalities became frustrated (Burrows 2008; Li 1998); (c) in many cases, students avoided the use of the target language in fulfilling the communicative tasks (Carless 2007; Littlewood 2007).

Formative evaluations are important but they do not answer the question as to whether TBLT works better than other approaches. In order to answer these questions, process-product evaluations are required. Their results will be discussed in the next section.

4.3.2 Process-product evaluations

Process-product evaluation has focused on learning outcomes and has usually been a part of comparative method studies. TBLT has been compared with the traditional PPP approach and other communicative approaches such as task-supported language teaching or production-based instruction (De la Fuente 2006; González-Lloret & Nielson 2015; Lai, Zhao, & Wang 2011; Nielson 2014; Shintani 2011).

In a short-term study, De la Fuente (2006) investigated the effects of three vocabulary lessons (one traditional and two task-oriented lessons) on acquisition of basic meanings and

morphological aspects of 15 Spanish words. She compared 30 English-speaking college students in three intact classes under one of three conditions: a PPP lesson, a task-based lesson with focus on form and a task-based lesson with explicit focus-on-forms component which could be viewed as a task-supported lesson. There were ten students in each class, randomly assigned to one of the three conditions. At the time the treatment lessons were carried out, learners had been exposed to approximately 43 hours of communicative L2 instruction in Spanish.

The study included two sessions. On day 1, all learners in the three groups participated in a preliminary 45-minute input-based session on the general topic of food. The learners received a vocabulary list for the lesson which did not contain the target words. The input-based component of the session consisted of listening to a dialogue (provided in writing as well) taking place in a supermarket and answering comprehension questions. On day 2, the 50-minute experimental session, different for three groups, took place.

The PPP lesson started with a presentation stage during which students were given a dialogue with the target words embedded and enhanced in the text. After this, in order to practice the target forms, they carried out three explicit focus-on-forms activities: (1) fill in the blank exercise; (2) an oral question–answer exercise with a focus on gender and number; and (3) a written activity where they had to classify the target words in semantic groups. At a production stage, students had to perform in pairs a role-play situation (at the restaurant).

The task-based group began the lesson with the same dialogue, but with no focus on formal aspects of the target words. Then learners performed a one-way, role-play, information-gap task which required them to use the target lexical forms while keeping attention to meaning, in order to achieve the goal of ordering food from a restaurant menu. One student, acting as the customer, had to order from a menu containing the new target words. The other student, who had the role of the waiter, had to explain any of the words the first student did not understand. At the end they repeated the same restaurant-ordering role play in pairs.

The task-supported group performed the same task as the task-based group. The difference was that instead of a task repetition, a teacher-generated, explicit focus-on-forms lesson component was incorporated at the end of the lesson.

In order to measure the outcome, two vocabulary discrete-point tests were employed. The first test was administered after the task, to measure immediate retrieval of the words' forms. Students were exposed to 15 slides, each containing an image of one target word. For each image they were asked to say the word. The second test was a delayed vocabulary test. This test was administered one week after the treatment, and measured both the retention of target words and the acquisition of formal aspects (gender and article agreement). The same testing procedure as

for the first test was used. In addition, for each image learners had to say the word with its corresponding definite article and provide the plural form.

De la Fuente found that all three groups performed equally in the immediate world retrieval tests after treatment. However, the delayed post-test did show a significant overall difference between the PPP group and both task-oriented groups. Task-based lessons seemed to be more effective than the PPP lesson. In terms of word retrieval, no significant difference was observed between the task-based and the task-supported groups but the task-supported group with an explicit focus-on-forms component was more accurate with gender and number morphology than the task-based group.

Another vocabulary study was carried out by Shintani (2011) who investigated the comparative effects of two types of treatment – one of which emphasized input and the other output – on the vocabulary acquisition of young EFL learners. Thirty-six Japanese children, aged 6–8, participated in her study. They were divided into three groups: input-based, production-based and control group. The two experimental groups and one control group each consisted of two classes respectively- one class with 6–7-year-old students with four months of English learning experience, and one class with 7–8-year-old students with 16 months of experience. The learners received six lessons (two 45-minute lessons per week). The three groups received three different types of treatment:

1. Students in the input-based (IB) group were not required to produce output. Three listen-and-do tasks were designed and conducted in one lesson that was repeated in each of the six lessons. These lessons were task-based.

2. Students in the production-based (PB) group were required to produce output. One set of the five activities was conducted in each lesson and repeated for six lessons with different item groups. The teacher usually provided corrective feedback on the learners' production by way of recasts. These lessons followed PPP approach.

3. The control group received a set of three activities (English songs, Total Physical Response, and alphabet practice) without being exposed to any target words.

Twenty-four nouns were selected as target items which included three familiar categories for Japanese children: animals, home appliances, and fruit and vegetables. The 24 items were introduced differently to the two experimental groups. In the PB group, the three sets of target items were introduced separately in each lesson, whereas in the IB group, all 24 target items were introduced each time in order to avoid the learners becoming aware of the linguistic goal. Therefore, the same lesson consisting of the same activities was conducted six times with all the 24 items for the IB group but with different item groups for the PB group.

The study employed a pre-test –immediate post-test –delayed post-test design. Four vocabulary tests were employed (two comprehension tests and two production tests). The test formats were the following:

1. A multiple-choice listening test which required the participants to listen to a word and choose the appropriate picture from six pictures. The test contained 40 questions that included 24 target items together with 16 distractors.

2. A category task test: A listen-and-do task was employed for this test. The participants were asked to listen to sentences and to decide in which of four given situations (fruit and vegetable shop, kitchen, bathroom, and zoo) each sentence had been said.

3. A discrete-item production test: In this test, the researcher asked the participant to name each flash-card for 24 target vocabulary items, which were displayed in turn. The participants obtained one point for each item correctly provided orally, irrespective of morphological errors.

4. A “same or different” task test: The participant and the researcher had a different sheet showing 24 pictures of objects. The participant was asked to name each object, and the researcher then told the participant which picture she herself had. If the participant’s and researcher’s pictures were the same, they put a tick in the square showing the picture. If they were not the same they put a cross.

In order to find whether each type of instruction (input-based and production-based) could enable learners to acquire new vocabulary, the scores obtained by the three groups in the pre-test, the immediate post-test and the delayed post-test were analyzed using a series of repeated-measures ANOVAs. The results of both listening tests showed that the IB and the PB groups significantly improved on the immediate post-test from the pre-test and on the delayed post-test from the pre-test, but there was no significant difference between the immediate and delayed post-tests. There was no statistically significant difference between the control group’s three repeated tests. As for the two production tests, it was found that both the IB group and the PB group improved significantly on the immediate post-test from the pre-test and on the delayed post-test from the pre-test. Furthermore, the IB group performed significantly better on the delayed post-test than on the immediate post-test, while there was no significant difference between immediate and delayed post-tests for the PB group. There was no significant difference between the control group’s three tests.

Another research question of the study was which of the two instructional approaches (input-based tasks vs. production-based activities) results in more newly acquired words. The statistical analysis of the scores obtained by the three groups showed that both the IB and the PB

groups significantly outperformed the control group in the immediate and delayed listening post-tests, and also the IB group significantly outperformed the PB group in the immediate and delayed task-based post-tests. As for the production tests, both the IB and the PB groups significantly outperformed the control group in the two post-tests, but there was no significant difference between the two experimental groups.

Shintani (2011) concludes that the findings provide evidence that both input-based and production-based instruction lead to both receptive and productive vocabulary knowledge. In general, the results show similar levels of effects for input-based and production-based instruction on vocabulary acquisition. However, the input-based group demonstrated better results in the task-based comprehension test and the same levels of achievement in the production tests despite relatively fewer opportunities for production.

An evaluation of a full-fledged TBLT program for adults was reported by González-Lloret and Nielson (2015) who evaluated a task-based Spanish course for the students of the US Border Patrol Academy (BPA). This course was designed as an alternative to the previous grammar-based course which involved working through a different grammatical topic for each class, was taught in English, and provided students very little time to practice communicating in Spanish. Many of the agents who completed the grammar-based course reported not being capable of doing their jobs in Spanish at Border Patrol stations. Most of the student evaluations of the old course criticized the grammar-translation approach and requested more practical Spanish training geared toward the language that trainees would use on the job. For this reason, the administration at the BPA sought a new approach to language teaching and sought guidance from experts in TBLT.

In order to design the task-based program, the course design team conducted a needs analysis and identified seven target tasks that agents were likely to need to complete on the job in Spanish (conducting a high-risk vehicle stop, extracting a suspect from hiding, offering and providing first aid, inspecting a bus, inspecting vehicles at a checkpoint, conducting a routine vehicle stop, and interviewing a person). The target tasks were sequenced into modules in order of task complexity. Each of these target tasks established the goal of a week-long module of instruction, incorporating increasingly complex pedagogic tasks which students completed alone, in pairs, and in groups. The course contained no planned explicit grammar instruction. Grammatical questions were answered when they arose, and the focus of the course was only shifted briefly to form when it was necessary for communication. The end-of-course assessment was performance-based and consisted of a series of tasks which encompassed all of the critical components of the overall course tasks. González-Lloret and Nielson outlined that the passing

rate for the TBLT course was impressive. In the over four years since the new TBLT program had been introduced in the Border Patrol Academy, there were almost zero failures in the final assessment.

In order to evaluate the new TBLT program, three exploratory empirical studies were undertaken: (1) a pilot study comparing the old grammar-based course and the TBLT course, (2) a study to establish whether the students' overall proficiency in Spanish had improved as a result of the TBLT program, and (3) a qualitative study of students' perceptions about the TBLT program.

The first study was designed to compare the oral proficiency of students after completing the grammar-based course and the TBLT course. 39 students participated in the study (19 from the grammar-based course and 20 from the task-based course). Students in both courses completed a full-time, 8-week program. After completing the course, they performed an oral picture-guided narration task. Their performance was assessed in terms of fluency, lexical and syntactical complexity, and grammatical accuracy.

Fluency was operationalized as the total number of syllables per minute. Three different measurements were considered for lexical complexity: (1) the number of lexical words per the total number of words produced, (2) lexical variety (first time use of lexical words), and (3) ratio of lexical to functional words. Mean length of utterance (total number of words divided by number of utterances) was employed as a measure of syntactic complexity. Finally, grammatical accuracy was operationalized as target-like use of noun-modifier agreement and noun-verb agreement.

González-Lloret and Nielson found that students in the task-based group performed significantly better than those in the grammar-based group on measures of fluency and structural complexity, and performed very similarly in terms of lexical complexity. Students in the grammar-based group outperformed students in the TBLT group in terms of grammatical accuracy, but not statistically significantly. Although the results of the study give clear evidence of efficiency of the task-based program, González-Lloret and Nielson point out that they must be interpreted with caution, given the small sample of students and the lack of a pre-test to compare both groups.

The second study undertaken as a part of the TBLT program evaluation was implemented to establish whether the students' overall proficiency had improved during their time in the TBLT program. The Versant Spanish test, which is a computerized oral proficiency assessment test, was administered to 256 students prior to their enrollment in the TBLT course. A post-test

was administered after course completion. The test provides scores for overall proficiency, sentence mastery, vocabulary, fluency, and pronunciation.

González-Lloret and Nielson reported that student performance on the Versant test improved statistically significantly after the task-based course. The mean overall proficiency score on the pre-test was 30.21. The mean overall proficiency score on the post-test was 37.68. Therefore, the overall proficiency scores improved an average of 7.47 points after the TBLT course. There were smaller effects for fluency and pronunciation (with change scores of 5.86 and 3.53) than for sentence mastery and vocabulary (with change scores of 7.55 and 7.16). A series of paired samples t-tests indicated that all of these change scores were statistically significant.

Finally, a qualitative study of students' perceptions of the Spanish TBLT program was carried out. Two electronic surveys were created, one for the students enrolled in the task-based BPA Spanish course and the other for students who had completed the program and were employed as active Border Patrol Agents in the field. 21 students and 16 agents completed the survey.

The surveys included a four-point Likert scale asking participants to rate their agreement with several statements about the course, from Strongly Agree to Strongly Disagree, as well as two open-ended questions- one asking them what they had enjoyed most about the course and the second one asking them what could have been improved. The results of the student survey revealed that students strongly agreed that their classes had focused on how to do their jobs (3.29). The students did not believe the program had focused mainly on grammar (1.67) or writing (1.71). In addition, although students felt they could use Spanish to do their jobs (2.62), many did not yet feel prepared to talk to native speakers (1.95). This result was also confirmed by their open-ended questions in which they suggested the learning of more practical, everyday Spanish to "talk to Spanish speakers outside of the job."

A similar positive feedback was obtained from Border Patrol Agents who had graduated from the task-based Spanish program. Their answers suggested high levels of satisfaction with the program. They found the program interesting (3.6) and useful in terms of topics (3.4) and vocabulary (3.4), and they strongly agreed that what they learned in the program was applicable to their jobs as new Border Patrol Agents (3.2).

The results of González-Lloret and Nielson's (2015) study are in favor of TBLT since they indicate that the task-based course prepares trainees to complete critical job tasks in L2, improves participants' overall L2 proficiency, and students seem to find the task-based program useful and relevant to their on-the-job needs.

A number of recent TBLT studies have focused on online task-based courses. As Nielson (2014, p. 296) notes, “with the rapidly increasing popularity of online language courses, there is an urgent need for research to confirm that the principles for course design that work so well in other contexts are, in fact, a good fit for distance language learning.” Below two studies on online, task-based Chinese courses for beginner (Lai, Zhao, & Wang 2011) and intermediate (Nielson 2014) levels will be reviewed.

The study on the online Chinese course for beginners with a task-based component conducted by Lai, Zhao, and Wang (2011) which has been presented above, apart from qualitative data, includes the analysis of learners’ oral production. The online course described in the study consists of asynchronous and synchronous components. The asynchronous components involve student self-study of the e-textbook, additional online learning resources (such as Chinese podcasts, Chinese character learning software, and online Chinese dictionary), and weekly individual language and culture assignments (e.g. recording oral responses to complete a dialogue; writing a short essay; discussions on given cultural topics). Additionally, students are required to attend one 1-hour small group (3–5 students) synchronous session with their instructor each week. Prior to the study, the synchronous sessions were usually run in the fashion of didactic teaching and structured practice of linguistic items via the Initiation–Response–Evaluation (IRE) classroom discourse pattern. In 2007, the researchers introduced a TBLT syllabus for these synchronous sessions. The TBLT syllabus was implemented in half of the online Chinese classes, while the other half followed the syllabus used in the past.

The tasks in the TBLT syllabus were designed to expand the topic of each unit in the e-textbook. Two TBLT sessions were designed to go with each unit and altogether 12 one-hour TBLT sessions were designed and implemented. The synchronous sessions followed a pre-task, during-task and post-task cycle. The pre-task consisted of mainly input-based tasks or activities aimed at familiarizing the students with the language needed for the main task. The task phase consisted of one or two output-based tasks that were designed to engage learners in working together and using resources available to achieve an outcome. In the post-task phase, learners were usually asked to repeat task performance in order to improve their complexity and fluency.

Thirty-eight students aged 13 to 18 who followed the TBLT syllabus during their synchronous sessions volunteered to participate in this study. The participants were all monolingual Anglo-American high school students. The control group consisted of 36 students with similar profiles.

After completing the course, two groups performed an oral picture description task as a part of their final exam. The recordings of students’ performance were transcribed and coded on

their fluency, complexity and accuracy. Fluency was measured in terms of meaningful words per minute. Accuracy was measured in terms of error-free clauses. Mean length of T-units was calculated in order to measure syntactic complexity. Lai, Zhao and Wang (2011) found that students in the TBLT classrooms demonstrated significantly higher fluency in language production than their counterparts in the control classrooms. There were no significant differences in syntactic complexity and accuracy between the two groups.

The researchers drew attention to the fact that there was a great variation in learners' fluency in the TBLT group: "Students in the TBLT classes seemed to be more divergent in the fluency of oral production than the students in the control classes" (Lai, Zhao, & Wang 2011, p. 89). The analysis of the self-reflection blogs of two learners with extreme results in fluency (the blog of a learner who demonstrated extremely high fluency and a learner who demonstrated extremely low fluency) showed that both learners were taught by the same teacher and had similar prior foreign language learning experience. The important difference stood out in their autonomous learning skills. Lai, Zhao and Wang (2011, p. 90) concluded that "for those who had great initiative and knew how to motivate themselves and how to learn strategically to start with, TBLT seemed to give them opportunities to achieve much. However, those who did not have such resources at their disposal gradually lagged behind and lost ground."

Another program evaluation was reported by Nielson (2014) who conducted the study on the effectiveness of an intermediate-level online task-based Chinese course designed by the specialists at the Center for Advanced Study of Language at the University of Maryland. Using a task-based approach, they developed, implemented and piloted a year-long online Chinese course for high school students.

Following the methodological principles of TBLT, the design process began with a needs analysis, during which 100 high school students enrolled in elementary and intermediate Chinese courses across the country were surveyed. The research team analyzed the results of the survey and identified five target tasks: (1) following street directions, (2) ordering food, (3) providing street directions, (4) negotiating for goods and services, and (5) arranging travel.

From these target tasks, 100 individual pedagogic tasks were derived, all of which were computer-based, and within each module, they were organized by complexity. Students began with input-focused activities which were followed by output-based activities (learners were asked to produce the language by recording themselves responding to a voicemail or another prompt) and communication with fluent interlocutors during weekly 30-minute online conversation sessions.

Two groups of students participated in Nielson's research project- an online group (n=35) that completed the year-long task-based Chinese course and a control group (n=12) that did not take part in the online course, but completed the pre-course/post-course proficiency testing. All participants in both the online and control groups were high school students between the ages of 13 and 17.

Students in the online group participated in performance-based role-play assessments (PBAs) during the course. PBAs consisted of a scenario during which the student assumed the role of a person completing one of the course's target tasks (e.g. follow street directions, order food from a vendor, order food at a restaurant, give street directions, etc.) while a fluent Chinese speaker took on the role of an interlocutor. In addition, students in the online course and in the control group completed the STAMP test of Chinese reading and speaking proficiency.

In order to evaluate PBAs, a Rasch analysis for each task was used. It is based on estimations of the language learners' abilities to perform the subtasks as well as the difficulty of each subtask. The Rasch analyses of the PBAs estimates a statistic that indicates how well a given task is able to spread out the learners in terms of ability on the task. The PBAs "ranged from the absolute lowest (ability to reliably spread out learners) at 0, for the "order food from a vendor" task, suggesting that learners all performed the task the same way (no difference in the learners' performance scores), to quite high at .91, for the "buy a cell phone" task, suggesting that this task reliably separated learners into different levels of performance ability" (Nielson 2014, p. 306).

Nielson (2014) reported that almost all of the students who attempted the PBAs were able to perform the target tasks very well, as indicated by the number of subtasks successfully accomplished out of the total number of subtasks attempted. Mean percentage of subtasks accomplished correctly was from 82% in "Buy a cell phone" task to 98% in "Order food from a vendor" task.

Learners' performance in the task-based group was also assessed in terms of accuracy, fluency and pronunciation in order to find out whether language proficiency and the task-based assessments were related. For each dimension, success criteria were established. A learner demonstrated Chinese language accuracy if his/her questions, comments, and responses were appropriate to the situation. Success criteria for fluency required that student's speech was clear and in Chinese (with the exception of proper names in English), and that pauses and mis-starts did not detract from comprehensibility. Finally, the success criterion for pronunciation was whether a conversation partner was able to understand a student's pronunciation and responded to questions and comments appropriately. Instructors were asked first to determine whether or

not the success criteria had been satisfied and then, if they had, to rate from 1 to 5 how well the learners had performed. Nielson (2014) found that the learners' mean scores on the PBAs (calculated as a percentage of the subtasks accomplished) had a positive, medium-sized to large correlation with the accuracy, fluency, and pronunciation ratings assigned for each performance, although there was much variation in the correlational patterns observed between PBA scores and the available language success evidence.

In order to answer the question as to whether learner proficiency improved as a result of the course, the pre- and post- STAMP scores were compared. The STAMP speaking scores offered a positive answer in the results of the paired samples t-test, which showed that learners improved in their speaking at the end of the course. Furthermore, the results of the speaking test were in favor of the task-based online course: 52% of the online group improved between the pre-test and the post-test, while only 22% of the control group demonstrated improvement. The reading tests showed an improvement in reading proficiency for 24% of the online course participants and just 16% of the control group participants, but Nielson (2014) suggested that this test is not the best tool to measure reading proficiency for the population in her study since 40% of the online course test takers and 75% of the control group test takers were at a reading proficiency level too low to be captured by this test.

Table 18 summarizes the findings of the studies on the evaluation of TBLT programs and courses discussed above.

The overview of task-based program evaluations has showed that TBLT has been implemented for teaching different languages (English, Spanish, Dutch, Chinese, and others) in a variety of foreign language contexts with very young learners, high school learners, and adults, absolute beginners and intermediate level students, in face-to-face classrooms and online contexts. All formative studies reported positive learner reactions to task-based courses although some problems have been detected by classroom teachers. Process-product evaluations demonstrated that after task-based treatment learners improved statistically significantly from a pre-test to a post-test, were capable to perform most of the target tasks successfully and that TB groups outperformed other groups (with PPP approach or control groups) on some measures (fluency, listening comprehension, and speaking proficiency).

Table 18. Evaluation of TBLT programs

Study	Methods compared	Duration	Number of participants	L1 / L2	Procedure	Tests and measures	Findings
De la Fuente (2006)	Group 1: PPP Group 2: task-based (TB) Group 3: task-supported (TS)	Two 50-minute sessions	30 adult college students	English / Spanish	On day 1, input-based lesson (the same for all three groups). On day 2, the treatment session: PPP group: presentation (a dialogue with the target words); practice (explicit focus-on-forms activities); production (a role-play situation at the restaurant). TB group: exposure to the input; role-play, information-gap task (order food from a restaurant's menu); task repetition. TS group: the same task as the TB group with a teacher-generated, explicit focus-on-forms lesson component. An immediate post-test and a delayed post-test (one week later).	Two discrete-point tests (students had to produce the lexical items orally when shown pictures).	1. There were no between-group differences on the immediate word retrieval post-test. 2. Both task-oriented groups outperformed the PPP group on the delayed post-test. 3. There was no difference between the performance of TB group and TS group on word retrieval. 4. TS group was more accurate with gender and number morphology than TB group.
Shintani (2011)	Group 1: input-based (IB) Group 2: production-based (PB) Control group: no exposure to the target vocabulary	12-week course, six 45-minute lessons	36 children aged 6-8	Japanese / English	A pre-test – 6 treatment sessions – an immediate post-test – a delayed post-test (4 weeks later). PB group: drill-like games which required production. IB group: game-like listen-and-do pedagogic tasks with flash cards. Control group: TPR, English songs, practice with writing the alphabet.	1. Discrete-point listening test (listen and choose the correct picture out of six choices provided). 2. Task-based listening test (listen to the utterances and decide in which setting it had been said). 3. Discrete-point production test (name the objects depicted in 24 flash cards). 4. Task-based production	1. Within-group comparisons: 1.1 On both listening tests, both treatment groups improved statistically significantly from pre- to immediate and delayed post-test. 1.2 On both production tests, both treatment groups improved statistically significantly from pre- to immediate and delayed post-test, and IB group also from immediate to delayed post-test. 1.3 The control group's scores showed no change on any measure. 2. Between-group comparisons: 2.1 Both treatment groups

						test (a two-way “Same or different?” test).	outperformed the control group on all four measures. 2.2 IB group performed as well as PB group on both discrete-point measures and the task-based speaking measure, and better than PB group on the task-based listening measure.
Gonzalez-Lloret & Nielson (2015)	Group 1: grammar-based (GB) Group 2: TBLT	8-week course, 8 hours of instruction per day	39 students of US Border Patrol Academy	English / Spanish	The grammar-based course: a traditional structural syllabus; lessons focused on grammatical topics; use of learners’ L1 to explain grammar. The TBLT program: task-based syllabus; no explicit grammar instruction; reactive focus on form; use of L2 (Spanish) only. Students from the two groups were audio-recorded during their fourth week of instruction completing the same task.	Picture-guided narrative task. Grammatical accuracy: target-like use of noun-modifier and noun-verb-agreement. Syntactic complexity: mean length of utterance. Lexical complexity: the number of lexical words per the total number of words produced; lexical variety (first time use of lexical words); and ratio of lexical to functional words. Fluency: number of syllables per minute.	The TBLT group performed statistically significantly better than students from the GB group on syntactic complexity and fluency. Accuracy was better for the GB group but not statistically significantly. There was no difference in lexical complexity and variety between the two groups.
Lai, Zhao, & Wang (2011)	Group 1: task-based online course Group 2 (Control group): language-focused online course	12 weeks (12 one-hour task-based modules)	74 high school students aged from 13 to 18	English / Chinese	TBLT group: The tasks in the TBLT syllabus were designed to expand the topic of each unit in the e-textbook used in the course. Online sessions followed a pre-task, during-task and post-task cycle. Control group: didactic teaching and structured practice of linguistic items via Initiation–Response–Evaluation classroom discourse pattern. At the end of the course, students	A monologic picture-description task. Accuracy: error-free clauses. Syntactic complexity: mean length of T-units. Fluency: meaningful words per minute.	The TBLT group demonstrated statistically significantly higher fluency in language production than their counterparts in the control group. There were no significant differences in the syntactic complexity and accuracy between the two groups.

					from the two groups were audio-recorded performing the same task.		
Nielson (2014)	TB group: completed the online task-based course. Control group: did not take part in the online course.	A year-long online Chinese course	47 high school students between the ages of 13 and 17.	English / Chinese	TB group: input-focused activities, output-based activities, and communication with fluent interlocutors during weekly 30-minute online conversation sessions. Performance-based role-play assessments (PBAs) during the online course. Students in both groups completed the STAMP test as a pre-test and post-test.	PBAs: 1) the number of subtasks successfully accomplished out of the total number of subtasks attempted; 2) accuracy, fluency and pronunciation rated by instructors from 0 to 5. STAMP test of Chinese reading and speaking proficiency.	<ol style="list-style-type: none"> 1. Almost all of the students in the TB group were able to perform the target tasks very well. Mean percentage of subtasks accomplished correctly was from 82% in the most complex task to 98% in the easiest task. 2. Mean learner scores on the PBAs had a positive, medium-sized to large correlation with the accuracy, fluency, and pronunciation ratings assigned for each performance. 3. Speaking proficiency: 52% of the TB group improved between the pre-test and the post-test, while only 22% of the control group demonstrated improvement. 4. Reading proficiency: tests showed an improvement in reading proficiency for 24% of the online TB course participants and 16% of the control group participants.

4.4 Research into task-supported and task-based approaches to teaching L2 Russian

There are a number of studies on Russian as a second or foreign language which focus on the acquisition of specific target forms or general accuracy within one language teaching approach, usually TSLT (Gor & Jackson 2013; Henry 1996; Kempe & MacWhinney 1998; Quero Gervilla 2005; Rubinstein 1995, among others). This line of research can be explained by the fact that task-supported language teaching with presentation – practice – production sequence prevails in Russian lessons both in Russia and in a foreign language context. As Gilabert and Castellví (in press) point out, since the 80s, Russian L2 teaching has wavered from traditional explanations in grammar syllabi, incursions into CLT, all the way to the conscious-practical method (see pp. 22–23 of the present study). For most teachers it is still a common practice to focus heavily on forms and use communicative activities as support practice of the recently explained grammar features. This approach is also reflected in many Russian textbooks which include activities that are clearly communicative but “they are not pedagogical tasks, inasmuch as they lack an achievable outcome whose propositional contents are verifiable” (Comer 2007, p. 184).

Quero Gervilla (2005) analyzed errors in the acquisition of Russian case forms by Spanish-speaking learners. The participants in her study (n=56) were students of different Spanish universities and language academies who were enrolled in different courses of Russian and were at their first (Level 1), second (Level 2), third and fourth (Level 3) year of studying.

Data were obtained by means of grammatical tests and written production. Grammatical tests consisted of 86 questions and included different exercises such as filling in the blanks (put the words in brackets into the correct form), answering questions using the words in brackets and sentence transformation according to a given model. In the second task participants were asked to write a composition about what they like to do in their free time.

Quero Gervilla (2005) reported the results for the use of each case and found that the most errors were made in the dative case (in the grammar tests) and in the accusative case (in learners’ written production). She also compared the percentage of errors in the case forms for three levels of proficiency. The reported results indicate that this percentage of errors decreased from Level 1 to Level 3 in the grammatical test. In other words, students with higher levels of proficiency committed fewer errors in the use of case forms than learners with lower level of proficiency. However, the analysis of written data showed different dynamics for the forms of the dative and accusative cases. Participants of Level 3 were less accurate with the dative case than participants of Level 1 and 2. Forms of the accusative case were produced more correctly by

learners of Level 1 than by the groups of Level 2. This may mean that the acquisition of the case forms (at least, some of them) in Russian is not linear and develops in a zigzag fashion, with backslidings, and occasional U-shaped behavior (Long 2015).

One of the most recent studies was carried out by Denissenko (2016) who investigated the impact of 150 hours of formal classroom instruction on the use of case forms and on accuracy, fluency and complexity of learners' written production. Participants in her study were adult bilingual Spanish and Catalan-speaking learners (n=83) who studied Russian at the Escola Oficial d'Idiomes in Barcelona. Participants were divided into four groups (G1, G2, G3, and G4) according to their level of proficiency (A1, A2, B1.1 and B1.2) following Common European Framework of Reference (CEFR).

Denissenko conducted a longitudinal study with a pre-test/post-test design and two times of data collection for each level group. She also applied a cross-sectional design to the data and examined development throughout the four levels of proficiency. Data in her study were obtained by means of a comic strip task in which participants were asked to write a story based on a set of images provided by the researcher. They were allowed 40 minutes to complete the task.

The researcher performed a number of statistical analyses which showed that 150 hours of formal instruction have a significant impact on G1, but not on G2, G3 or G4. Specifically, the G1 showed a significant improvement in three dimensions: (i) in fluency which was measured by calculating the number of words per minute; (ii) in syntactic complexity measured by total number of independent clauses and total number of dependent clauses; (iii) and in lexical richness, measured by means of Guiraud's Index. However, none of the groups had improved in terms of the amount and quality of errors produced when using case marking in Russian. The lack of any significant improvement in G2, G3 and G4 seems surprising and even discouraging. Denissenko suggests that such stagnation throughout levels 2, 3 and 4 may be due to insufficient amounts of hours of formal instruction and insufficient input.

To our knowledge, there is no published research that reported results of the implementation of a task-based Russian language course. However, there are several calls for L2 Russian programs to embrace TBLT (for example, in the United States) and there have been attempts to apply task-based methodology for classroom procedure (Comer 2007) and for evaluation purposes (Long, Gor & Jackson 2012).

An example of a task-based Russian lesson with a duration of fifty minutes is described by Comer (2007). Participants in his study were college-level second-semester learners of Russian in the United States. At the time when the sample TBLT lesson was conducted, the

students were working on *Nachalo* (2000), a standard textbook grounded in the communicative approach. The sample TBLT lesson used the textbook section's theme (Russian foods) and made the overall objective for the lesson students' oral production of several sentences comparing Russian and American eating habits, specifically in relation to breakfast. During the lesson, students worked on four pedagogical tasks which involved an oral comprehension task and information-exchange tasks.

The study did not follow pre-test – post-test design, and all the author's conclusions are based on analysis of the lesson's recording and his speculations. In terms of student engagement with the language, Comer (2007) evaluates the class as “moderately successful.” Students achieved a desired outcome in the performance of all the tasks. However, the quality of their oral production was lower than expected: “Despite the tasks' structure to encourage students to use sentence-level discourse, the recording reveals that much of their talk was still at the level of words and phrases” (Comer 2007, p. 192). Besides, participants had difficulties with matching subjects with correct verb forms. For this reason, toward the end of session, the instructor made an unplanned decision to treat the conjugation of the verbs *есть/пить* (to eat/to drink) explicitly.

In relation with this metalinguistic episode, Comer (2007, p. 190) highlights “an especially thorny issue for Russian in terms of implementing TBLT within a model of implicit or non-obtrusive grammar instruction.” He suggests that “the students did not have enough attentional resources to forge strong enough form-meaning connections from the rich language input to make the verb forms intake for their developing linguistic systems.” In this situation, a brief explicit grammar explanation conducted in an interactive manner (teacher-student question-and-answer), with opportunities for students to reflect on the patterns of this new morphology seems justified, despite the fact that it is certainly not recommended in the TBLT literature.

Comer concludes that although TBLT has great potential in getting students actively engaged in using the language in the classroom, there are at least two problems with designing a task-based Russian course. The first problem involves a lack of needs analysis for diverse groups of Russian learners. Another problem for implementing TBLT as the structuring principle for Russian is the lack of knowledge of the actual developmental stages of learners' interlanguage. As Comer (2007, p. 193) points out, “without this knowledge, it will be hard to predict what linguistic forms in the input that students receive to complete tasks should receive instructional attention and treatment.”

Another TBLT study which involved L2 Russian was conducted by Long, Gor, and Jackson (2012). In this study tasks were used for evaluation procedure. The aim of the study was

to determine whether there is a relationship between the control of the linguistic features and the Interagency Linguistic Roundtable (ILR) levels, and to identify which linguistic features of Russian correlate with ILR proficiency levels 2, 2+ and 3 on the ILR scale. A computer-delivered battery of 33 perception and production tasks was administered to 68 participants (57 learners between levels 2 and 3 on the ILR scale, and 11 native speaker controls). The tasks sampled subjects' control of Russian phonology, morphology, syntax, lexis, and collocations. Statistical analyses showed that 32 of the 33 tasks in the data-collection battery (18 in perception and 14 in production) significantly differentiated ILR proficiency levels 2 and 3. On the basis of the results of the study, checklists of linguistic features for L2 learners that match typical levels of control of those features, measured as percentages accurate, to proficiency levels on the ILR scale were made. The results of this study reveal differential potential of tasks. However, the question about the efficiency of the task-based approach for learning Russian has not been discussed in any of the reviewed studies.

4.5 Summary of Chapter IV

In the first part of this chapter, two approaches that adopted tasks for different purposes – task-supported and task-based approaches – have been compared. Their distinguishing characteristics have been outlined, and different strategies of classroom behavior have been described. We have seen that differences between TSLT and TBLT involve the way tasks are selected, sequenced and assessed; type of feedback and questions, negotiation of meaning, turn taking, the learner's and the teacher's role, language focus, learning materials, and other factors.

Then the findings of studies that evaluated the efficiency of both approaches have been reviewed. One of the most important outcomes of task-based treatment reported in these studies was learners' ability to perform the target tasks successfully. In some studies, L2 production was measured in terms of CAF. TBLT was found to have beneficial effects on fluency. Accuracy, on the contrary, was better for the grammar-based or task-supported groups than for the task-based groups but not statistically significantly. Lexical complexity was measured only in one of the reviewed studies, and no difference in lexical complexity and variety was found between task-based and grammar-based groups. Syntactic complexity has presented mixed results. Some studies have shown significant gains in syntactic complexity after task-based treatment, whereas others have shown no significant gains in this dimension of production. Finally, the results of some studies on Russian L2 teaching were discussed.

CHAPTER V

STUDY DESIGN AND METHODOLOGY

5.1 Introduction

In the previous chapter a number of studies on evaluation of task-based programs and courses were reviewed. In this chapter the motivation for this study, the research goals, and the specific questions and hypotheses addressed by this study will be advanced. The experiment designed to answer those questions will be described. First, the justification for the target items in this study will be provided. Second, the procedure and the results of the pilot study conducted prior to the experiment will be described. Then the chapter will focus on the experiment itself by providing information about the participants in the study, design and procedures, treatment sessions and data collection, the transcription and coding of learners' oral production, and, finally, the measures applied in the study in order to evaluate the results of two types of treatment, task-supported and task-based.

5.2 Motivation for this study

A review of the literature on TBLT program evaluations showed that, firstly, there is the need for process-product evaluations of task-based courses, especially long-term studies with various "program-fair" outcome measures (Long 2015). Secondly, the potential of tasks beyond the TBLT approach, for example as units of classroom activities or assessment within task-supported language teaching, has not been fully investigated. Thirdly, there is not enough data about the implementation of TBLT with morphologically complex languages, such as Russian. As Larsen-Freeman (2010, p. 223) pointed out, the study of target languages, which have a more complicated morphology than English, the most studied language in SLA, is welcome, "especially because languages with more complicated morphology call into question what have appeared to be straightforward distinctions such as regular and irregular morphological processing." Hence, there is the need to study how the morphological complexity of Russian can condition the application of meaning-focused and learner-centered methodologies such as TBLT,

and to determine whether this type of approach may actually provide effective teaching practice in this language.

5.3 Research goals and questions

The general goal of the experiment is to compare two approaches to language teaching in order to determine whether one of them was more efficient for learning morphologically complex target forms (Russian case forms and verbs of motion) and for improving learners' oral and written production in terms of its accuracy, fluency, syntactic complexity and lexical diversity.

This study extends previous research in at least three ways. Firstly, it reconsiders the question of the efficiency of the task-supported approach that has been doubted by some advocates of TBLT. Secondly, it provides new evidence about the effects of the task-based approach on learning a morphologically rich language (Russian) which is different from the L2 usually involved in TBLT studies (English, Spanish, French, etc.). Thirdly, it challenges some of the widely accepted proposals for task-based syllabus design by reviewing the problem of task design and sequencing in the light of high code complexity.

The study consists of two separate parts. The first part (group Rus I) is focused on the Russian case system. The second part (group Rus II) is designed around verbs of motion with prefixes.

The study will try to answer the following questions:

RQ 1: Do the task-supported approach and the task-based approach help to promote the accurate use of case forms and prepositions?

RQ 2: Which of the two approaches is more efficient for learning new lexical items (Russian verbs of motion with prefixes)?

RQ 3: How does each approach affect learners' oral and written production in terms of general accuracy, syntactic and lexical complexity, and fluency?

5.4 Hypotheses

On the basis of the reviewed literature and personal teaching experience the following hypotheses have been advanced:

Hypothesis 1. Effect of task-supported and task-based treatment on the use of prepositions and case forms

Learners in both groups will improve their use of prepositions after the treatment, without any significant difference between the two groups. Learners in both groups will show improvement in the use of case forms as a result of treatment. After the task-supported treatment learners will use the case forms more accurately than after the task-based treatment since all activities in TSLT (drills, explicit instruction, controlled practice, and others) aim at promoting accurate use of the target forms.

Hypothesis 2. Effect of task-supported and task-based treatment on learning new lexical items (verbs of motion)

Since previous studies have found TBLT to be beneficial for learning new lexical items and their morphological characteristics, we hypothesize that learners in the task-based (TB) group will produce more target forms (verbs of motion) and will be more accurate in their use than learners in the task-supported (TS) group while performing oral and written tasks. However, the TS group will demonstrate higher scores in grammar tests.

Hypothesis 3. Effect of task-supported and task-based treatment on learners' oral and written production

Hypothesis 3 is divided into four sub-hypotheses which correspond to each of the CAF dimensions.

Hypothesis 3.1 Effect of task-supported and task-based treatment on learners' general accuracy

We hypothesize that learners will improve their accuracy both after task-supported and task-based treatment and that learners in the TS group will demonstrate higher general accuracy than learners in the TB group.

Hypothesis 3.2 Effect of task-supported and task-based treatment on learners' syntactic complexity

No significant improvement in syntactic complexity is expected for either the TS group or the TB group because of learners' low level of proficiency and a lack of exposure to complex syntactic structures in the input. We also hypothesize that the two groups will not differ in their syntactic complexity scores.

Hypothesis 3.3 Effect of task-supported and task-based treatment on learners' lexical complexity

We suggest that learners will improve their lexical diversity (e.g. they will use higher variety of vocabulary) as a result of experimental sessions. The task-based treatment will result in higher lexical complexity than the task-supported treatment thanks to its focus on meaning and communicative orientation.

Hypothesis 3.4 Effect of task-supported and task-based treatment on learners' oral fluency

We hypothesize that learners in the TB group will improve their fluency, whereas learners in the TS group will show no significant differences in their fluency scores. The task-based treatment will result in higher oral fluency than task-supported treatment because it will give learners more possibilities for communication and interaction.

The research questions have determined the current study design and methodology which will be described in the next sections.

5.5 Justification for target items

The current study was integrated into the university curriculum; for this reason, the target items were predetermined by the university course programs “Llengua Russa I” (further Rus I) and “Llengua Russa II” (further Rus II). These courses are based on structural syllabi, that is, on the lists of grammar forms, structures, and vocabulary topics. The syllabus for the first year focuses on the nominal morphology, that is, forms of nouns, pronouns and adjectives. During the second year, learners study verbal morphology (verbal aspect and conjugation, verbs of motion, and imperative). In order to determine target vocabulary and grammar for the present study, the syllabi for the first and second year of Russian were analyzed with the aim of eliciting forms and meanings that learners had seen at the time the experimental treatment started and the new forms and meanings that they had to learn during the treatment. Following the Russian language syllabi, the first part of the experiment was focused on the Russian case system, and the second part included the verbs of motion and corresponding prepositions and case forms. Due to the peculiarity and complexity of the Russian morphological system, it seems appropriate to give a brief grammatical comment on the target forms.

Case is a grammatical category whose value reflects the grammatical function performed by a declinable part of speech (e.g. noun, pronoun, adjective, and numeral) in a phrase, clause, or

sentence. Russian has six cases: nominative, genitive, dative, accusative, instrumental, and prepositional case. For example:

Я живу **в** маленьком городе.

Ya zhivu **v malenkom gorode** (prepositional case, locative meaning).

I live in a small town.

Я еду **в** маленький город.

Ya edu **v malenkiy gorod** (accusative case, meaning 'direction').

I'm going to a small town.

There are three factors which complicate form-function mapping and make the acquisition of Russian cases a complex task for L2 learners (Magnani & Artoni 2015). First, both cases and their respective inflectional endings are numerous and contain information about other relevant features, such as number (singular or plural), gender (masculine, feminine, neuter), animacy (animate or inanimate). Moreover, some forms are affected by a set of phonological constraints, such as stress shift and vowel removal or insertion. For example: zamok (castle) nom.sing. – zamka gen.sing.; devochka (girl) nom.sing. – devochek gen.plur. Secondly, Russian case endings hardly ever match their functions unambiguously. That is, one ending can perform different functions (e.g. the ending –e marks neuter singular nominative, as well as feminine singular dative and singular prepositional), and, conversely, one function can be performed by different forms (e.g. feminine singular instrumental can be marked by –ju, –oj or –ej). Thirdly, each case can be used in a variety of syntactic structures (e.g. genitive case for comparison; dative case for indirect object; accusative case for direct object, etc.).

Errors in case forms can be perceived merely as grammatical errors in form (1), or lead to an odd interpretation of the sentence (2), misunderstanding or ambiguity (3).

(1) Я живу в *Барселона.

Ya zhivu v *Barcelona.

I live in Barcelona.

(2) Это письмо я послал *тебя.

Eto pismo ya poslal *teba.

This letter I sent you.

(3) Я купил *сын машину.

Ya kupil *syn mashinu.

I bought son car. (I bought my son's car or I bought a car to my son?)

The target forms for Rus I groups (case forms with their meanings and prepositions) and some examples of their use are presented in Table 19.

Table 19. Target forms for Rus I: Case forms and prepositions

New prepositions and meanings of case forms	Target forms seen before the treatment
1. Nominative case	
	1.1 Subject: <i>Барселона красивая</i> (Barcelona is nice); 1.2 Copular noun: <i>Барселона – красивый город</i> (Barcelona is a nice city).
2. Accusative case	
	2.1 Direct object: <i>читать книгу</i> (to read a book); 2.2 Direction: В/на (to) + Accusative case: <i>идти в школу</i> (to go to school), <i>идти на работу</i> (to go to work); 2.3 В (on) + days of week: <i>в субботу</i> (on Saturday).
3. Prepositional case	
1.1 Expression of time: В/на (in) + Prepositional case: <i>на прошлой неделе</i> (last week), <i>в октябре</i> (in October), <i>в 2012 году</i> (in 2012); 1.2 О (about) + Prepositional case: <i>читать о Москве</i> (to read about Moscow).	3.1 Expression of place: В/на (in/on/at) + Prepositional case: <i>работать в библиотеке</i> (to work at the library), <i>быть на работе</i> (to be at work); 3.2 На (by) + means of transport: <i>ехать на машине</i> (to go by car).
4. Genitive case	
Без (without), для (for), из (from), около (near), слева от (to the left of), справа от (to the right of), напротив (in front of), в центре (in the center of) + Genitive case.	4.1 Possession: <i>книга учителя</i> (a book of the teacher); 4.2 Absence: <i>У меня нет брата.</i> (I don't have a brother.) <i>В городе нет парка.</i> (There is no park in the city); 4.3 Expression of quantity (after numerals and adverbs): <i>сколько книг</i> (how many books), <i>много книг</i> (many books), <i>две книги</i> (two books).
5. Dative case	
К (to + person) + Dative case: <i>ехать к другу домой</i> (to go to a friend's house).	5.1 Indirect object: <i>дать (книгу) другу</i> (to give (a book) to a friend); 5.2 Age: <i>Мне 20 лет</i> (I'm 20 years old); 5.3 In constructions <i>Мне нравится</i> (I like), <i>Мне нужно</i> (I need); 5.4 По (by) + Dative case: <i>говорить по телефону</i> (speak by phone).
6. Instrumental case	
Перед (in front of), за (behind), под (under), над (above), между (between), рядом с (next to) + Instrumental case.	6.1 С (with) + Instrumental case: <i>гулять с другом</i> (to walk with a friend), <i>кофе с молоком</i> (coffee with milk); 6.2 After verbs <i>быть</i> (to be), <i>работать</i> (to work), <i>стать</i> (to become), <i>интересоваться</i> (to be interested in), <i>заниматься</i> (to do); 6.3 Expression of time (seasons and parts of the day): <i>зимой</i> (in winter), <i>утром</i> (in the morning).

Verbs of motion (VM) is the name for a lexical-grammatical group of verbs that consists of 28 verbs and their derivatives with prefixes with a general meaning of “motion, displacement”: to go, to run, to fly, to swim, to bring, to drive, to climb, etc. These verbs can cause difficulties for learners both at formal (irregular verbs) and semantical levels. Verbs of motion are divided into unidirectional and multidirectional verbs. Some of them are

differentiated according to the manner of motion (Talmy 1985), whether some means of transport are used or not. For example, the English verb **to go** corresponds to four Russian verbs:

Идти (idti) – to go by foot in one direction (one-way trip).

Ходить (hodit') – to go by foot in more than one direction (a round trip or walking without any direction); general capacity to walk.

Ехать (ehat') – to go by transport in one direction.

Ездить (ezdit') – to go by transport in more than one direction.

Adding different prefixes to the verbs of motion modifies their meaning and also influences the verbal aspect (imperfective/perfective). For example:

Приходить (prihodit') – to come/to arrive by foot (imperfective).

Прийти (priyti) – to come/to arrive by foot (perfective).

Приезжать (priezzhat') – to come/to arrive by transport (imperfective).

Приехать (priehat') – to come/to arrive by transport (perfective).

The number and complexity of semantical and formal nuances that must be taken into account when choosing an appropriate verb and the obligatory use of specific prepositions and corresponding case forms after some of the verbs of motion turn the acquisition of these verbs into a real challenge. Table 20 presents the verbs of motion and corresponding prepositions which were included in this study.

Table 20. Target forms for Rus II: Verbs of motion

New target items	Target forms seen before the treatment
<p>I. Verbs of motion with the prefixes: приходить – прийти (to come by foot) приезжать – приехать (to come / to arrive in a transport) уходить – уйти (to leave by foot) уезжать – уехать (to leave in a transport) доходить – дойти (to reach some place by foot) доезжать – доехать (to reach some place in a transport) проходить – пройти (to pass by some place by foot) проезжать – проехать (to pass by some place in a transport) входить – войти (to enter) выходить – выйти (to go out) заходить – зайти (to drop in) переходить – перейти (to cross)</p> <p>II. Other verbs поворачивать – повернуть (to turn)</p>	<p>Verb to go: идти-ходить, ехать-ездить, пойти, поехать</p>

5.6 Task elaboration: pilot

5.6.1 Objectives

The objective of the pilot was to test the whole experiment including treatment tasks and control tasks. In particular, it had the following purposes:

- to test the treatment tasks and their sequence; to make sure that the level of task complexity was feasible for learners;
- to determine to what extent focus-on-form techniques used in the tasks could help to draw learners' attention to the target forms;
- to test control tasks and check whether their performance would promote the use of the target items.

5.6.2 Participants

The participants of the pilot, a total of 20 people (17 females and 3 males), were bilingual Catalan and Spanish-speaking learners of Russian between 18 and 20 years old. They were enrolled in the program “Modern Languages and Literature” at the University of Barcelona and were doing the courses “Llengua Russa I” and “Llengua Russa II.” At the moment the pilot started, the learners had been exposed to approximately 80 hours of instruction.

5.6.3 Procedure

The pilot was carried out in classroom conditions. First, a pre-test was carried out with the Rus I group. Then a task-supported treatment started. It included 8 sessions (14 hours in total). The lessons were given by professor Joan Castellví and the author of this thesis. An immediate post-test followed these sessions. The same procedure (a pre-test – 8 sessions – an immediate post-test) but with the addition of control oral tasks was repeated with the Rus II group which consisted of the same students (see Table 21).

Table 21. Pilot procedure

Spring semester (Rus I)	Autumn semester (Rus II)
Pre-test: grammar and written production	
	Pre-test: oral production
Time interval: 7 days	
8 sessions	
Time interval: 5 days	
Immediate post-test: grammar and written production	
	Immediate post-test: oral production

In this way, the treatment and control tasks were implemented under similar conditions as the planned conditions for the experimental study, that is, within the same time frame and with the learners who had the same characteristics as learners in the two experimental groups (their age, Spanish or/and Catalan as L1, learners' level of Russian at the moment of the experiment, and their educational background).

5.6.4 Treatment tasks

5.6.4.1 Task selection and design

As we have seen in Chapter II, in the task-based approach the starting point in task selection is a needs analysis which allows establishing the target tasks for a specific group of learners (Long 1985). In our study, the descriptors presented in the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (further CEFR) (2001) were used as a reference for determining learners' needs. Since at the end of the course "Llengua Russa II" students were expected to have an A2 level of language proficiency, the descriptors for this level were chosen in order to design pedagogical tasks (see Table 22).

Table 22. CEFR descriptors for A2 level (2001, p. 224)

Describing & Narrating	Information Exchange	Settings
<ul style="list-style-type: none"> • people, appearance • objects, pets, possessions • background, job • places & living conditions • events & activities • likes/dislikes • plans/arrangements • habits/routines • personal experience 	<ul style="list-style-type: none"> • work & free time • simple directions & instructions • pastimes, habits, routines • past activities 	<ul style="list-style-type: none"> • basic common needs: personal details, daily routines, info requests • routine everyday transactions • familiar situations & topics • everyday situations with predictable content

Apart from learner needs described in CEFR (2001), criteria for task selection included (1) learners' previous linguistic knowledge, (2) the target forms described in Section 5.2, and (3) a potential level of code complexity (that is, the complexity and variety of linguistic resources needed to carry out a task) of each task. As the pilot was done with task-supported treatment, task sequence followed the order of presentation of the target items in the syllabi established for the groups Rus I and Rus II. The pilot tasks are described in Table 23.

Table 23. Pilot tasks

Tasks	Communicative goals
Rus I	
<p>What is the film about? Write synopses of three films so that other students in the group can guess what films you are talking about. You may use the key words for help.</p>	Give a brief description of the content of a film.
<p>Menu for the day Prepare a detailed menu for six people (a child, a sportsman, a vegetarian, a pregnant woman, a person on a diet, a millionaire) for one day. Write what they are going to eat and drink for breakfast, lunch and supper. Take into account that they have special needs and preferences.</p>	Speak about needs and likes/dislikes of different people.
<p>Agenda Make a plan for the day. Say when and where you have to go and who you have to see. Take into account the following information: you have a toothache, and the dentist starts to work at 14 o'clock; your friend invited you to a restaurant at 21 o'clock, etc.</p>	Make a plan for the day including meetings with different people in different places.
<p>Object placing You have a picture of the same room as your partner but some objects are missing. You are allowed to look only at your picture and not at your partner's. Ask your partner if he/she has the objects that you have. If he/she does not have them, explain where they are located. Then listen to your partner and draw the missing objects in your picture.</p>	Describe a room by explaining the location of the objects; locate objects on the basis of oral or written information.
Rus II	
<p>Library schedule Six students worked in the university library last week. Each student worked 10 hours. Fill in the library attendance table on the basis of information that you and your partners have. Say when (day and time) each of the students worked in the library.</p>	Ask for information in order to know another person's weekly schedule and arrange a meeting; explain your daily routine.
<p>Street directions 1. You and your partner have a map of the city center in St. Petersburg. Explain to him/her the route that is marked in your map. Your partner has to follow your instructions and draw the route on his/her map. 2. Help Russian tourists in Barcelona. You are in plaza Catalunya. Explain to Russian tourists how to get to Sagrada Familia, Ciutadella park, Palau de la Musica Catalana, and the port.</p>	Request and follow street directions using a map; explain how to get to different places.

Different focus-on-form techniques described in Chapter III were used in task design, in particular: task-essential language, promotion of negotiation of meaning, input flooding, input enhancement, corrective feedback, and direct explanation. Most of the tasks (except the menu of the day task and the agenda task) were information gap tasks. This task type is believed to promote negotiation of meaning (Doughty & Williams 1998; Pica *et al.* 1993) which may be beneficial for language acquisition according to the Interaction Hypothesis. Tasks were designed

in such a way that most target forms were task-essential, that is, a successful performance of the task was not possible without using these forms (Loschky & Bley-Vroman 1993). Texts in the pre-task were flooded with the target items (input flooding) which were graphically enhanced (typed in bold and in red).

5.6.4.2 Problems in task performance

During task performance some problems in task design related to cognitive and code complexity were detected. From the six pilot tasks described in the previous section learners struggled considerably while performing the film task and failed to perform the library schedule task. Observation of learners doing these tasks and the further analysis of their content helped to detect the following problems:

- Learners didn't know well or could not use properly lexical items they were expected to know (for example, numerals or days of week). Combined with the difficulty of using new grammatical forms it increased code complexity of both tasks to such a degree that most students couldn't perform them successfully. Some students managed to achieve a desired outcome, but with the use of their L1.
- A high number of elements in the library schedule task (agreement on 5 days per week for 6 people with different timetables) and reasoning demands (simple logical or mathematical operations) increased cognitive complexity of the task to such an extent that it was not feasible for learners. It was obvious that the code complexity of the task required so much attention that learners could not handle other variables of the task.
- In the film task students focused mostly on meaning and did not notice the target grammar forms. They used the preposition **o** (about) correctly, which was essential for task completion, but did not pay attention to the forms of prepositional case which are obligatory after this preposition. For this reason, their performance contained a significant number of grammar errors.
- Task sequence in the pilot didn't correspond to the degree of task complexity, especially code complexity. Learners had to perform more complex tasks ("What is the film about?" and "Library schedule") before easier tasks.

These conclusions led us to rearrange task sequencing (introduce more linguistically and cognitively easier tasks before more complex tasks) and modify the design of individual tasks. Changes in task design will be described in the following section.

5.6.4.3 Modification of task design

In order to improve the efficiency of the film task, more pre-task activities were elaborated. They included listening and reading exercises based on the input enriched with the target forms (**о** + prepositional case). Input flooding and input enhancement were used to draw learners' attention to the endings of nouns and adjectives after the preposition **о** (about).

The library schedule task required more changes. The fact that all the learners failed to perform this task could be explained by excessively high code complexity and additional cognitive load (for more discussion on linguistic complexity and task design in Russian see Gilabert & Castellví, in press). In order to make this task feasible for learners, it was necessary to decrease the amount of computation and to ensure that learners could use task-essential language forms properly. After some changes were made in task design (see Table 24), participants of the pilot performed a modified version of the task. The results were satisfactory: all the groups achieved an outcome (filled in the table of library attendance correctly), although one group needed more time to finish the task.

Table 24. Modification of the library schedule task

	Task 1 (the original version)	Task 2 (the modified version)
Objective	To agree the use of a computer with a Russian keyboard in the library.	
Task conditions	<u>Six students</u> need to use the computer. Each student has a different class schedule and can work before or after his/her lessons. The computer is available <u>from 8 to 20</u> five days per week.	<u>Four students</u> need to use the computer. Each student has a different class schedule and can work before or after his/her lessons. The computer is available <u>from 10 to 18</u> five days per week.
Group organization and information distribution	20 learners divided into five groups (four learners in each group). Each learner had a different piece of <u>information about all six students in the task</u> (e.g., Lena came to the library at 12 on Monday. Victor came two hours later than Lena.)	20 learners divided into five groups (four learners in each group). Each learner had <u>information about one student in the task</u> (e.g., Lena came to the library at 12 on Monday. On Tuesday and Wednesday she came at 14.)
Cognitive complexity	High cognitive complexity: high number of elements (5 days, 12 hours, 6 students); reasoning demands that entail logic operations and simple calculation (e.g., Student A came to the library at 10:00. Student B came two hours later than Student A).	Lower cognitive complexity: lower number of elements (5 days, 8 hours, 4 students); no reasoning demands (necessary information, e.g. day and time of visiting the library, is presented explicitly and clearly).
Code complexity	High code complexity: verbs of motion (приходить/прийти (to come) and уходить/уйти (to leave)); present and past tense; imperfective and perfective verbal forms; numbers (to express time) and days of week.	Lower code complexity: verbs of motion only in the past tense. Less vocabulary demands as a consequence of pre-task activities focused on numbers and days of week.
Results	The task was unattainable for all the groups.	The task was successfully accomplished by four groups; one group could not finish the task on time (25 minutes).

5.6.4.4 Conclusions for further task design and sequencing

In general, task performance during the pilot could be considered successful. However, some errors were committed in the design of two tasks and task sequencing. First, it was not taken into account that code complexity of Russian might require so much attention that learners could barely handle the cognitive load (high number of elements and reasoning demands). It led to a performance failure of one of the tasks. Second, learners' current knowledge was sometimes overestimated, and code complexity of some tasks was higher than expected.

Based on the pilot and the analysis of detected problems, a series of decisions were made regarding further task design and sequence:

1. Focus on form should happen throughout the whole task. As demonstrated in Chapter III, focus on form may happen proactively during task design, reactively during task performance, or post-actively after the task has been performed. Reactive focus-on-form is incidental and is mostly unpredictable as it depends on learners' internal syllabuses; however, proactive and post-active focus on form can be addressed during task design. The pre-task phase must provide greater exposure to input, due to the greater variability of forms and numerous morphosyntactic conditions that affect form-meaning connections which learners have to notice. In order to encourage noticing, different focus-on-form techniques should be used, such as input flooding, visual input enhancement, input simplification, input elaboration, etc. Post-task work can be addressed towards processes such as hypothesis testing, rule-formation, and form-function mapping. This work should give learners a possibility to internalize, modify or consolidate what they know about the language.

2. Linguistic difficulty needs to be computed during task design in order to avoid cognitive overload. As Gilabert and Castellví (in press) point out, if not gauged appropriately, the whole computation by L2 learners during task performance may collapse and be transformed into negative perceptions of difficulty and anxiety that may cause insufficient processing and learning.

3. An adequate balance of cognitive demands and linguistic difficulty is crucial. In order to balance them out, task designers should monitor the amount of 'minimal code' demanded by each task.

4. Tasks can be organized by increasing complexity according to students' needs. A framework for task sequencing, at least for morphologically rich languages, should include both cognitive complexity and linguistic difficulty.

As there was no time for the second pilot, these modifications were applied directly to the treatment tasks of the current study which will be described in Section 5.4.4.1.

5.6.5 Control tasks

5.6.5.1 Task design

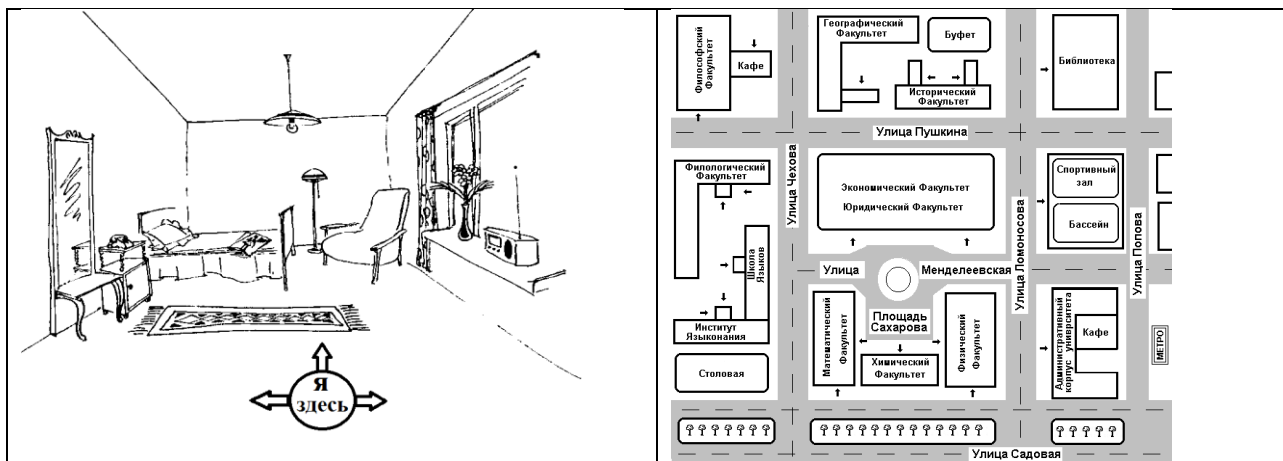
In order to compare two different types of treatment, task-based and grammar-based, two types of control tasks were designed: grammar tests and written and oral tasks. Data collection was carried out with the learners of Rus I (grammar tests and written production) and the learners of Rus II (grammar tests, written and oral production) following a pre-test – immediate post-test design.

Three types of grammar tests were provided: (1) fill in the blanks; (2) multiple choice test; and (3) grammaticality judgment test. The written task for learners of the group Rus I was to answer a friend's letter. The questions that learners had to answer to do the task were formulated in such a way that learners had to use a variety of target forms – prepositions and case forms in different meanings. Students of Rus II were asked to write an email inviting their Russian friends to a birthday party and explaining the road from the metro station to their home.

Students of Rus II also performed three oral tasks individually. They were asked (1) to describe a room, (2) to give directions looking at the map, and (3) to tell a story on the basis of a comic strip (see Table 25).

Table 25. Pilot oral tasks (Rus II)

Task 1 You've rented a room. Your parents haven't seen it and ask you how it is and what you have there. Give a detailed description of your room.	Task 2 You are studying at the Moscow University. You have a new student in your group. Explain to him how to get to the Philological Faculty from the metro station.
--	---



Task 3

Look at the pictures and explain what happened.



5.6.5.2 Results

Since the objective of the pilot was to test control tasks and check whether their performance would promote use of the target items, learner’s written and oral performance was analyzed following two criteria: (1) their general ability to perform a task (in order to avoid excessively high task complexity as happened with some treatment tasks), and (2) the number of target forms produced.

We found that learners both in Rus I and Rus II successfully performed the written tasks and used the target forms and vocabulary as expected. However, the analysis of learner’s oral performance, which had been recorded and transcribed, detected some problems in task design.

The Room description and the Map task were completed successfully, but the production of some learners was too short (30–40 words) and did not permit an evaluation of progress in the

use of the target forms. The narration task based on the comic strip caused more problems. First, four students out of twenty seemed to be blocked or frustrated and just narrated the story in Spanish. One of the reasons could be the fact that they were not familiar with this task type. Another reason was the lack of knowledge of some basic vocabulary necessary to tell the story (e.g. a man, a woman, to call by phone, to give, etc.). Learners who managed to describe the comic strip avoided using the target items (verbs of motion with prefixes). They used a number of strategies to replace these verbs: (i) used only the verb “to go” (идти/ходить) without prefixes; (ii) found other ways to express the necessary meaning (e.g. “he says Good bye” instead of “he leaves”); (iii) just ignored the pictures which required the use of a verb of motion (e.g. to cross the street) and did not describe them at all, or (iiii) used Spanish or Catalan words.

5.6.5.3 Modification of control tasks

The oral control tasks were modified based on information obtained from the pilot. The comic strip task was excluded since it did not promote the use of the target forms and was found to be unattainable for some learners. The design of the room description and the map task remained the same but more subtasks were elaborated in order to increase the amount of production. An image of the second room was added to the room description task, and the task explanation was slightly changed. Instead of renting a room, learners were offered a situation of renting a flat with a living room and a sleeping room. They had to describe both rooms. In the map task more destinations were marked on the map, and, consequently, more directions were required. The task consisted of explaining the way from the metro to the bank, from the bank to the library, and from the library to the Philological Faculty.

5.7 The current study: design and methodology

5.7.1 Participants

Originally, three groups of participants, 68 learners in total, took part in the current study. They were adult learners of Russian as a foreign language who were enrolled in the programs “Modern Languages and Literature” and “Linguistics” at the University of Barcelona. The study was carried out under real classroom conditions as a part of the Russian language courses

“Llengua Russa I” (further Rus I) and “Llengua Russa II” (Rus II). At the moment the experimental treatment began, learners in the group Rus I had received approximately 80 hours of instruction and the learners in the group Rus II about 120 hours.

From the original pool, a number of participants were eliminated from the study for the following reasons: 1) having a Slavic language (Polish, Ukrainian or Russian) as L1; 2) missing more than 50% of treatment sessions; 3) demonstrating a very poor level of attainment which caused their inability for oral and/or written production in the pre- and post-tests. In the end, the total number of participants in the study was 54. We haven't included a control group of native speakers of Russian in the current study because the proficiency level of participants was low and could not be compared with a native-like level.

In order to prove group comparability, first, background information about the participants was obtained, and, second, the results of pre-tests were statistically analyzed and compared. Group comparability on the pre-tests will be discussed in Section 6.3 of the next chapter. Group background information is summarized in Table 26. More detailed information on the participants is provided in Appendix A.

First, we compared task-supported and task-based groups of Rus I. Learners' ages ranged between 18 and 20 years old in the TS group and between 18 and 25 in the TB group. In the TB group, 63% of participants were bilingual Catalan and Spanish speakers. Four learners marked only Spanish as their L1, and three learners had the second mother language different from Spanish or Catalan (Basque, German and French). In the TS group, 90% of participants were bilingual Catalan and Spanish speakers, and the rest of the group marked only Spanish as their L1.

Both groups of Rus II had similar characteristics: a prevalence of female participants, an age range between 18 and 25, and the majority of learners speaking Catalan and Spanish as their L1. In the TS group, 66% of participants were bilingual Catalan and Spanish speakers and in the TB group the percentage of bilingual Catalan and Spanish-speaking participants was 63%. One learner in the TS group was a native Italian-speaker and another learner spoke French as L1. Both reported that they had a high level of Spanish as a foreign language. Students in the TB group of Rus II were the same as in the TB group of Rus I. Thereby, the results of the background questionnaire showed that from the point of view of age, gender characteristics and participants' first language all the groups were comparable.

Table 26. Background information per group

Group	N	Mean Age	Gender	L1
RUS I				
TS	20	18,95	17 females, 3 males	18 – Catalan and Spanish, 2 – Spanish
TB	19	20,37	15 females, 4 males	12 – Catalan and Spanish, 4 – Spanish, 1 – Spanish and Basque, 1 – Catalan and German, 1 – Spanish and French
RUS II				
TS	15	20,33	13 females, 2 males	10 – Catalan and Spanish, 3 – Spanish, 1 – Italian, 1 – French
TB	19	20,37	15 females, 4 males	12 – Catalan and Spanish, 4 – Spanish, 1 – Spanish and Basque, 1 – Catalan and German, 1 – Spanish and French

5.7.2 Design

The study was based on a pre-test – immediate post-test – delayed post-test design. It included two phases – task-based phase and task-supported phase. The group classification was organized according to two criteria: (1) approach (whether the treatment was task-based or task-supported); and (2) the target forms – the case forms for Rus I and verbs of motion for Rus II. On the basis of these criteria, we distinguished four groups:

TS_Rus I

TB_Rus I

TS_Rus II

TB_Rus II

The between-subject factor included a type of approach (task-based vs. task-supported). That is, the groups TS_Rus I and TB_Rus I were compared, on the one hand, and the groups TS_Rus II and TB_Rus II, on the other hand. The within-subject factor was the results obtained at each point in time (pre-test, immediate post-test and delayed post-test), that is, the learners' gains achieved through each type of treatment.

5.7.3 Procedure

The procedure was the same for task-supported and task-based groups (see Table 27). Since the current study was conducted in real classroom conditions, it was adapted to learners' academic schedules in terms of its organization. It posed some limitations on data collection. The immediate post-test of Rus I and the delayed post-test of Rus II were a part of the final exam at

the end of the corresponding course. This fact could have influenced the results obtained by the learners in these tests. On the one hand, they could have made more effort to perform successfully in their final exam. On the other hand, they could have been more nervous, especially during their oral performance. However, the impact of anxiety level and motivation on learner performance is beyond the scope of this study. Since the TS group and the TB group performed the control tasks under the same conditions, the affective variable would not have discredited the comparative analysis of results achieved by learners in the two groups.

Table 27. Experimental procedure

Task-supported approach		Task-based approach	
Spring semester (Rus I)	Autumn semester (Rus II)	Spring semester (Rus I)	Autumn semester (Rus II)
Pre-test: grammar and written production	Pre-test: grammar tests, written and oral production	Pre-test: grammar and written production	Pre-test: grammar tests, written and oral production
Time interval: 7 days			
8 task-supported sessions		8 task-based sessions	
Time interval: 5–7 days			
Immediate post-test: grammar and written production (= final exam of the course)	Immediate post-test: grammar tests, written and oral production	Immediate post-test: grammar and written production (= final exam of the course)	Immediate post-test: grammar tests, written and oral production
3 months	1 month	3 months	1 month
Delayed post-test: grammar and written production	Delayed post-test: grammar tests, written and oral production (= final exam of the course)	Delayed post-test: grammar and written production	Delayed post-test: grammar tests, written and oral production (= final exam of the course)

The organization of treatment sessions will be described in the following section. Written data collection was conducted with the whole group during a Russian lesson, and, in two occasions, during the Russian exam at the end of the term. The participants had one hour and a half to do the grammar tests and the written task. They were not allowed to use dictionaries or any learning materials.

Oral data collection took place in professor Castellví's office when participants didn't have a Russian lesson, and it was conducted by professor Castellví and by the researcher herself. Learners performed the tasks individually. They were given instructions in Russian or in Spanish/Catalan if needed. Then the participants were given as much time as they needed to prepare for a task. It took them about 2–4 minutes to prepare for each task. The performance of two oral tasks lasted 10–15 minutes in total.

5.7.4 Treatment sessions

The study included sixteen task-based sessions (eight lessons with the group Rus I and eight lessons with the group Rus II) and sixteen task-supported sessions. The lessons were given by Dr. Joan Castellví and the author of this work. It is important to keep in mind that although the teacher is an experienced specialist in teaching Russian as a foreign language, the TBLT methodology was new for him at the moment when the experiment started. The general length of sessions for each group was 14 hours. Since TSLT is currently used for Russian lessons at the University of Barcelona, it was used as a reference point for both types of treatment. It's important to keep in mind that language exams which learners have to pass are not task-based, and linguistic criteria of assessment (related primarily to accuracy) still prevail in language teaching. Hence, one of the teacher's goals is to prepare learners for these exams and help them to develop the required linguistic competence, no matter what teaching approach is used. Therefore, the TS and TB treatment included the same language material which was sequenced and presented in a different way according to the approach applied in each case. Two types of treatment will be described in the next sections.

5.7.4.1 Task-based phase

5.7.4.1.1 Task design

In the task-based treatment, eleven tasks were implemented, seven tasks for Rus I and four tasks for Rus II. Apart from six tasks from the pilot, new tasks were designed taking into account the conclusions made in the pilot and CEFR descriptors of target tasks for A2 level. All tasks are described in Table 28.

Tasks designed for the treatment followed a three-part framework offered by Willis (1996) and Ellis (2006) and described in Chapter II. Tasks consisted of pre-task, task cycle and post-task. Some pre-task activities were done individually, and others were performed in pairs or small groups (3–4 learners). Compared to the pilot, the pre-task phase was extended by adding more reading and listening activities and work with vocabulary necessary for task performance. The pre-task activities had two main objectives: 1) to give students an advance idea of what they would be required to do and the nature of the outcome they would be expected to arrive at (Ellis

2006); and 2) to present new language that the learners could use while performing the task and mobilization of existing linguistic resources (Skehan 1998b).

The post-task was usually a part of students' homework. Learners were asked to write compositions, emails, announcements, descriptions or upload oral messages in the Virtual Campus. Then their works were checked and corrected by the teacher and given back to the students with the necessary comments. It increased the amount of individual written and oral production outside the classroom and provided every learner with the opportunity to receive individual corrective feedback.

Most tasks used in the study were information gap (Prahbu 1987) or jigsaw tasks (Pica *et al.* 1993). They were closed, two-way, convergent tasks in which each participant had a part of the necessary information and had to exchange it in order to perform the task. The choice of this task type was motivated by the findings of various researchers who claimed that this task promotes negotiation of meaning and enhances the probability that language learners will pay attention to particular aspects of the language in the context of a meaningful activity (Alegría de la Colina & García Mayo 2007; Doughty & Williams 1998; Pica, Kanagy, & Falodun 1993; Swain & Lapkin 2001). Pica, Kanagy, and Falodun (1993) argued that jigsaw tasks promote the greatest opportunities for learners to experience comprehension of input, feedback on production and interlanguage modification. However, other task types were also used, for example decision making tasks (make a menu, plan an excursion, rent a flat).

Table 28. Tasks implemented in the current study and their correlation with CEFR descriptors

Tasks	Communicative goals of tasks	CEFR descriptors	Target items
Rus I			
<p>Menu for a day Prepare a detailed menu for six persons (a child, a sportsman, a vegetarian, a pregnant woman, a person on a diet, a millionaire) for one day. Write what they are going to eat and drink for breakfast, lunch and supper. Take into account that they have special needs and preferences.</p>	To speak about preferences in food, to explain what a dish is made of	Narrate about likes/dislikes	<ul style="list-style-type: none"> - Без (without), для (for), из (from) + Genitive case - С (with) + Instrumental case - Accusative case (direct object)
<p>Dictogloss Listen to a text three times. In a group try to reconstruct the text as close to the original as possible.</p>	To understand oral information and reproduce it accurately	Oral comprehension of simple texts	В/на (in/on/at) + Prepositional case (singular and plural forms, locative meaning)
<p>Agenda Make a plan for the day. Say when and where you have to go and who you have to see. Take into account the following information: you have a toothache, and the dentist starts to work at 14 o'clock; your friend invited you to a restaurant at 21 o'clock, etc.</p>	To make a plan for the day including meetings with different people in different places	Make plans and arrangements	<ul style="list-style-type: none"> - С (with) + Instrumental case - В/на (in/on/at) + Prepositional case - К (to) + Dative case - В/на (to) + Accusative case
<p>Excursion plan Write an excursion program for one day for different groups of tourists who came to Barcelona (sportsmen, Erasmus students, children, businessmen, etc.).</p>	To talk about interests and likes/dislikes of different people, to plan a day when visiting a city	Narrate about likes/dislikes	Dative case in the structures: <ul style="list-style-type: none"> - Кому нравится что (Who likes what) - Кому интересно что (Who is interested in what) - Кому нужно что (Who needs what)
<p>Object placing You have a picture of the same room as your partner but some objects are missing. You are allowed to look only at your picture and not at your partner's. Ask your partner if he/she has the objects that you have. If he/she does not have them, explain where they are located. Then listen to your partner and draw the missing objects in your picture.</p>	To talk about one's surroundings, to describe the location of things in different environments (in a room; in a city), to localize objects on the basis of oral or written information	Describe objects and possessions	<ul style="list-style-type: none"> - Около (near), слева от (to the left of), справа от (to the right of), напротив (in front of), в центре (in the center of) + Genitive case - В (in), на (on) + Prepositional case - Перед (in front of), за (behind), под (under), над (above), между (between), рядом с (next to) + Instrumental case
<p>Biography Ask two learners about some facts of their biography (when they were born; when they started studying at university, etc.) and about their plans for the future. Compare the answers and say whether they have something in common.</p>	To talk about oneself, saying what happened, when and where it happened; to ask another person about his/her life	Narrate about one's background, work and free time	<ul style="list-style-type: none"> - В/на (in/on) + Prepositional case (expression of time) - Genitive case without a preposition (expression of date) - Dative case without a preposition (expression of age)
<p>What is the film about?</p>	To explain what a film is	Describe a plot of	- О (about) + Prepositional case

Write synopses of three films so that other students in the group can guess what films you are talking about. You may use the key words for help.	about	book/film	- Который (which)
Rus II			
Rent a flat 1. You have to rent a flat in Moscow for three months. Read the advertisements and choose a flat. Explain the reasons for your choice. 2. You are looking for a room to share with your friend. Each of you has found a good option. Describe the room and try to convince your partner to take it.	To understand and give the information in order to rent a flat/ a room; To make a choice based on information received and justify it	Describe places and living conditions	- Около (near), слева от (to the left of), справа от (to the right of), напротив (in front of), в центре (in the center of) + Genitive case - В (in), на (on) + Prepositional case - Перед (in front of), за (behind), под (under), над (above), между (between), рядом с (next to) + Instrumental case
City. Street 1. Look at the picture and explain to your partner what you see in the street (buildings, persons, animals) and where they are located. You should give as many details as possible so he/she can draw it. 2. Look at the map and explain to your partner what buildings (shops, restaurants, etc.) are situated around the University of Barcelona. You should give as many details as possible so he/she can draw it.	To describe a street/an area by explaining the location of the objects/buildings	Describe places and living conditions	- Около (near), слева от (to the left of), справа от (to the right of), напротив (in front of), в центре (in the center of) + Genitive case - В (in), на (on) + Prepositional case - Перед (in front of), за (behind), под (under), над (above), между (between), рядом с (next to) + Instrumental case
Street directions 1. You and your partner have a map of the city center of Saint Petersburg. Explain to him/her the route which is marked on your map. Your partner has to follow your instructions and draw the route on his/her map. 2. Help Russian tourists in Barcelona. You are in Plaza Catalunya. Explain to Russian tourists how to get to Sagrada Familia, Ciutadella park, Palau de la Musica Catalana, and the port.	To request and follow street directions using a map; to explain how to get to different places	Exchange simple directions and instructions	- Verb to go without prefixes: идти-ходить; ехать-ездить - Verbs of motion with prefixes: пойти, поехать (to go); доходить – дойти, доезжать – доехать (to reach); проходить – пройти, проезжать – проехать (to pass by); входить – войти (to enter); выходить – выйти (to go out); заходить – зайти (to drop in); переходить – перейти (to cross) - Поворачивать – повернуть (to turn); налево (to the left), направо (to the right), прямо (straight)
Library schedule Four students at the University of Barcelona worked in the library last week. Each student had his/her own schedule. Some days they worked together, and other days alone. Fill in the library attendance table (from Monday to Friday) based on information that you have. Then exchange this information with your partners. Find out when the students worked together in the library.	To ask for information in order to know other person's weekly schedule and arrange a meeting; to explain a daily routine	Make plans and arrangements Describe pastimes, habits, routines	- Verbs of motion: приходить – прийти, приезжать – приехать (to come); уходить – уйти, уезжать – уехать (to leave) - Из/с (from) + Genitive case - В/на (to) + Accusative case

5.7.4.1.2 Task sequencing

The tasks were sequenced according to task complexity and code complexity. Although in the TBLT approach task complexity is the main criterion of task sequencing, the pilot showed that for Russian, at least for learners with a low level of proficiency, code complexity plays a crucial role. Moreover, Skehan's model of attention (1998a) has been taken into account.

Thus, task sequencing was determined, first of all, by language demands for each task. In order to avoid exceedingly high code complexity, a list of task-useful and task-essential vocabulary and grammar forms was made for each task. The amount of new vocabulary in each task varied from 4 to 10 items. Tasks, which required more new linguistic items, were considered more difficult and were performed after tasks with fewer linguistic requirements.

Task complexity (cognitive load) was also taken into account. In general, it was decreased as much as possible (no time pressure, no reasoning demands, familiar situations and topics) in order to have learners' attention directed to the language forms and to avoid excessive cognitive pressure. However, some tasks imposed more cognitive load on the learner than others due to the following cognitive factors: +/- Here-and-Now, +/- few steps, and a number of elements. The first task for Rus II, rent a flat, supposed Here-and-Now situation, contained three elements (three flats to choose from) and shared information (that is, all the participants had the same piece of information and there was no need to share it), and required from learners few steps to take (to make a choice on the basis of three flat descriptions and justify it). The most complex task, Library schedule, was the last one in the treatment. It imposed more cognitive load on the learners. It implicated a There-and-Then condition, various steps (to fill in the table on the basis of the given information, to share this information with other participants and to complete the table on the basis of information received from them), and a higher number of elements as compared to other tasks.

5.7.4.1.3 Focus on form

In order to draw learners' attention to the target forms, different focus-on-form techniques were used. Tasks were designed in such a way that some target forms were task-useful or task-essential, that is, task performance was faster and more efficient if learners used these forms or even was not possible without using these forms. For example, verbs of motion

are useful, and some of them are essential for giving street directions; the preposition “about” must be used in order to describe a content of a film (What is the film about?), etc.

Focus-on-form techniques were also used in the presentation of input. Texts in the pre-task were flooded with the target items (input flooding) which were visually enhanced (input enhancement). In the example below a text from the street directions task is flooded with verbs of motion and corresponding prepositions which are typed in bold:

Вам надо **перейти** улицу, **пройти через** площадь Университет, **перейти** улицу ещё раз, и вы увидите улицу Pelai. Вы должны **идти прямо по** улице Pelai **до** торгового центра Triangle. Вам надо **пройти мимо** кафе Цюрих и **вернуть направо на** улицу Ramblas. Вы должны **идти прямо по** улице Ramblas и **дойти до** улицы Canuda.

You should **cross** the street, **go through** the University Square, **cross** the street again, and you will see Pelai Street. You should **go straight by** Pelai Street until you **reach** the commercial center Triangle. You have to **pass by** the café Zurich and **turn to the right on** Ramblas. You need to **go straight by** Ramblas and **reach** Canuda Street.

In some cases, input elaboration was used in order to introduce and explain new vocabulary and draw learners’ attention to task-useful or task-essential lexical items as in the following example (menu of the day):

Русские едят три раза в день: утром – завтрак, днём – обед, вечером – ужин. На завтрак они обычно едят **бутерброды (хлеб с сыром, маслом или колбасой)**, яйца, йогурт и пьют кофе или чай.

Russians eat three times per day: breakfast in the morning, lunch in the afternoon, and supper in the evening. For breakfast they usually eat **sandwiches (bread with cheese, butter or ham)**, eggs, or yogurt and drink coffee or tea.

Feedback is an important element in the interactionist model of SLA (Gass 1997; Long 1996) and an instrument to draw learners’ attention to problematic forms. Although some TBLT researchers advocate for implicit forms of feedback (recasts) since they do not interrupt a natural flow of communication, for most Russian teachers explicit error correction seems necessary due to complexity and variety of form-meaning connections in Russian. In the current study learners received implicit feedback at the pre-task and during task performance and both explicit and implicit corrective feedback at the post-task phase.

Direct instruction (for example, a short rule explanation) was given when learners asked for it in the process of communication. It was necessary in the tasks which required the use of various case forms (e.g. object placing task) or verbal forms (street directions).

5.7.4.2 Task-supported phase

Lessons during the task-supported treatment usually followed the PPP model. Unlike the task-based treatment, the task-supported instruction involved what Long (1991) called focus-on-formS. New language material was first explicitly presented and explained by the teacher in Catalan. Then learners practiced the target forms by means of various exercises (fill in the blanks, transformations, asking and answering questions following a model, answering questions after reading a text or a dialog, etc.). The exercises were taken from the course book “Curs del lengua russa. Nivel inicial” (Ruiz-Zorrilla Cruzate, Kornakov, & Castellví Vives 2001). A number of exercises were prepared by professor Castellví specifically for the treatment. Some of them were done by the whole group under the teacher’s control in the classroom. The rest of the exercises were done by learners individually at home. Finally, the participants were expected to produce these forms while performing oral tasks (production stage) in small groups or pairs and written tasks done individually as a part of their homework. Tasks were used to provide learners with the opportunity to speak in Russian in a meaningful context and practice a “form of the lesson.”

Learners’ participation and the teacher’s role were similar to those described by Ellis (2006) and Long (2015) for PPP lessons and discussed in the previous chapter. The lessons were highly teacher-controlled. Discourse structure consisted mostly of IRF (initiate-respond-feedback) exchanges. Turn-taking was regulated by the teacher. There was little need or opportunity to negotiate meaning. The teacher’s discourse was characterized by display questions and form-focused feedback.

Grammar forms were the main units of the classroom activities and they were sequenced according to their presentation in the course book. The content of task-supported sessions is described in Table 29.

Table 29. Task-supported treatment

RUS I			
	Structures and vocabulary	Exercises	Tasks
1	1. Ordinal numbers and Genitive case to express the date: <i>первого февраля</i> (on the first of February) 2. Prepositional case to express time: <i>в этом/ следующем/ прошлом году</i> (this/next/last year); <i>на этой/ следующей/ прошлой неделе</i> (this/next/last week) <i>в феврале</i> (in February) 3. Plural forms of Prepositional case (locative meaning): <i>в больших</i>	Exercises of Unit 7 ⁶ . Read and translate texts and dialogs.	

⁶ Units 7–9 from the course book “Curs del lengua russa. Nivel inicial” (Ruiz-Zorrilla Cruzate, Kornakov, & Castellví Vives 2001) were used.

	<i>городах</i> (in big cities)		
2	О (about) + Prepositional case: <i>говорить о музыке</i> (speak about music)	Unit 7	What is the film about?
3	1. Repeat the structures for expression of time and the preposition о (about) + Prepositional case 2. Который (which/that)	Unit 7	
4	1. Plural forms of Genitive case after numerals and transitive verbs 2. Structures for expression of time (year, century): <i>в 1992 году</i> (in 1992); <i>в девятнадцатом веке</i> (in the nineteenth century)	Exercises and texts of Unit 8. Match the year and the event.	
5	1. Prepositions без (without), для (for), из (from) + Genitive case; с (with) + Instrumental case 2. Vocabulary Food	Unit 8	Menu for a day
6	1. Prepositions of space: около (near), слева от (to the left of), справа от (to the right of), напротив (in front of), в центре (in the center of) + Genitive case; в (in), на (on) + Prepositional case 2. Vocabulary Room, Furniture	Exercises of Unit 8. Practice room description.	
7	Prepositions of space: перед (in front of), за (behind), под (under), над (above), между (between), рядом с (next to) + Instrumental case	Exercises of Unit 8	Object placing
8	Practice the structures: <i>идти куда/к кому</i> (to go where/to see who); <i>жить где/у кого/с кем</i> (to live where/with who); <i>встретиться где</i> (to meet where)	Exercises of Unit 9	Agenda
RUS II			
1	Prepositions of space: около (near), слева от (to the left of), справа от (to the right of), напротив (in front of), в центре (in the center of) + Genitive case; в (in), на (on) + Prepositional case; перед (in front of), за (behind), под (under), над (above), между (between), рядом с (next to) + Instrumental case	Fill in the blanks exercises	Object placing
2	1. Verbs of motion without prefixes: <i>идти – ходить</i> (to go by foot), <i>ехать – ездить</i> (to go by transport) 2. В/на (to) + Accusative case (direction) 3. В/на + Prepositional case (location)	1. Grammar exercises (fill in the blanks, put the word in a correct form) 2. Read a text and answer the questions	
3	Vocabulary City	True/false exercise, make a description of different cities	
4	1. Vocabulary City 2. Presentation of verbs of motion with prefixes	Read a text, translate new vocabulary	
5	1. Verbs of motion with prefixes: пойти, поехать (to go); доходить – дойти, доезжать – доехать (to reach); проходить – пройти, проезжать – проехать (to pass by); входить – войти (to enter); выходить – выйти (to go out); заходить – зайти (to drop in); переходить – перейти (to cross) 2. Vocabulary to give directions: поворачивать – повернуть (to turn); налево (to the left), направо (to the right), прямо (straight)	1. Exercises (fill in the blanks) 2. Read and translate a text. Find verbs of motion in this text and explain their use.	Street directions
6	1. Verbs of motion 2. Prepositions из/с (from), мимо (by), до (to) + Genitive case; в/на (to), через (through) + Accusative case	Read a text and fill in the gaps with the verbs of motion.	Street directions
7	Presentation and practice of verbs of motion: приходить – прийти, приезжать – приехать (to come); уходить – уйти, уезжать – уехать (to leave)	1. Grammar exercises 2. Make short	

		dialogs in a pair following a model.	
8	Verbs of motion: приходить – прийти, приезжать – приехать (to come); уходить – уйти, уезжать – уехать (to leave)	Drill exercises	Library schedule

5.7.5 Data collection

5.7.5.1 Grammar tests

Participants were provided with three types of grammar tests:

1. Fill in the blanks. Learners had to put a word in brackets in the correct form and write a corresponding preposition if necessary. In some cases, several options were possible, for example:

Я живу _____ (парк). Я вижу его из окна.

I live _____ (park). I can see it from the window.

Possible prepositions could be: *напротив* (in front of), *около*, *рядом с* (near), *недалеко от* (not far from).

2. Multiple choice test. Learners were asked to choose one correct answer from three options.

3. Grammaticality judgment test contained 20 sentences for Rus I and 15 sentences for Rus II, some of which were correct and others of which contained a grammar error. Learners were asked to detect sentences with an error and correct them.

In the test for Rus I, every item (a case form in the specific meaning) appeared 6–12 times. This disproportionate distribution of forms was motivated by the fact that some forms have double difficulty, for example, prepositional and accusative case in the meaning of place and direction when learners have to choose both the correct noun form and the correct preposition (*в/на*). Another example is expression of direct object (accusative case) when the form depends both on the gender of the word and animate/inanimate characteristic. These forms appeared in the tests more times. In the test for Rus II every target item (verbs of motion with corresponding prepositions) appeared 6–8 times. Grammar tests which were used in the current study are presented in Appendix B.

5.7.5.2 Written tasks

Groups of Rus I and Rus II performed one written task (each group had a different task). Learners of the group Rus I were asked to answer a friend's letter. The questions that learners had to answer doing the task were formulated in such a way that in order to answer them learners had to use a variety of target forms – prepositions and case forms in different meanings. The task for students of Rus II consisted of writing an email in order to invite their Russian friends to a birthday party. This task required use of the target forms – verbs of motion with and without prefixes and corresponding prepositions. The two written tasks can be found in Appendix C.

5.7.5.3 Oral tasks

Oral performance included two tasks – the room description and the map task (see Appendix C). The first task was expected to encourage learners to use a variety of prepositions of space and case forms required by these prepositions. The second task involved giving street directions looking at the map. The map was designed in such a way that some of the target forms were task-essential or task-useful (e.g. turn to the left/to the right, cross the street, go straight, etc.).

5.8 Transcription and coding

The letters and emails written by participants as a part of the written tasks were transcribed and digitally formatted by the researcher prior to analysis, working directly with the handwritten original versions. All works were typed in MSWord, with auto-correction disabled. The transcriptions were kept as close as possible to the original handwritten texts; each writer's spelling and punctuation choices were transcribed verbatim and learners' paragraphing and spacing were replicated. There were a few cases of words that were lined through by learners in the handwritten versions. These words were not included in the transcribed text.

The transcription of learners' oral performance was also carried out by the researcher. Two versions of transcribed texts were created. The first version was as close as possible to the original text produced by the participants. It included words in Spanish or Catalan and disfluency

phenomena, e.g. false starts, repetitions and self-corrections. False starts are utterances that are begun and then either abandoned altogether or reformulated in some way (Foster *et al.* 2000, p. 368). Repetitions are words, phrases or clauses that are repeated without any modification (Skehan & Foster 1999). Self-corrections (or self-repairs) “occur when the speaker identifies an error either during or immediately following production and stops and reformulates the speech. Self-corrections will therefore include an element of structural change” (Foster *et al.* 2000, p. 368). This set of “raw” transcriptions was saved as Version 1, which was used to measure oral fluency (see Section 5.9.2.4).

In the second version, false starts, repetitions and self-corrections as well as asides in the L1 were excluded. This version was used in order to measure accuracy and complexity of oral data (see Sections 5.9.2.2 and 5.9.2.3).

In cases of unclear pronunciation, colleagues were asked to give a second opinion, and if the meaning of a word still was not clear, it was replaced with a question mark ‘?’ in the transcribed text. Usually these incomprehensible words were a result of language interference (e.g. “brod” instead of *khleb* “bread”). They were classified as lexical errors.

5.9 Evaluation procedures

5.9.1 Grammar tests

For each of the three grammar tests, fill in the blanks, multiple choice and grammaticality judgment, a number of correct answers was counted. One point was given to a correct answer, and zero points were assigned to incorrectly answered or unanswered items. Then the percentage of the correct answers was calculated for each test by dividing the number of correct answers by the total number of items in the test and multiplied by 100. Grammatical tests for Rus II were shorter than for Rus I because this part of the experiment included less target forms than the first part. The total scores for each test are presented in Table 30.

Table 30. Total scores for grammar tests

Rus I		Rus II	
Test	Total score	Test	Total score
Fill in the blanks	103	Fill in the blanks	45
Multiple choice	20	Multiple choice	15
Grammaticality judgment	20	Grammaticality judgment	15

For example, in order to evaluate the results of multiple choice test in Rus I, the following formula was applied:

$$\frac{\text{Number of correct answers}}{20} \times 100$$

5.9.2 Oral and written tasks

5.9.2.1 Research into CAF

Complexity, accuracy and fluency (CAF) have provided the standard of measurement in L2 research for nearly two decades (Ellis & Barkhuizen 2005; Housen *et al.* 2012; Lambert & Kormos 2014; Larsen-Freeman 2006, 2009; Norris & Ortega 2009; Pallotti 2009; Skehan 2003, 2009; Skehan & Foster 2012; Wolfe-Quintero *et al.* 1998; a special issue of Applied Linguistics 2009). Historically, CAF research traces its origins to the 1970s, when L2 researchers turned to metrics of grammatical complexity and accuracy developed in L1 acquisition research in their search for an L2 developmental index (Housen *et al.* 2012; Larsen-Freeman 2009). In mid-nineties Skehan (1998a) introduced a proficiency model that brought the three dimensions together. He claimed that initial contrast between meaning and form is complexified by a further contrast within form between control and restructuring. In this three-way set of distinctions, meaning is reflected in the fluency with which tasks are accomplished. Form-control is reflected in the accuracy with which language is used. Form-restructuring relates to the complexity of the language used, and the willingness on the part of the second language learner to take risks. Skehan proposed that these three areas are of fundamental importance, reflecting a capacity for growth and a capacity to use language effectively. Since then, the status of CAF as principal and distinct dimensions of L2 performance has been justified both empirically and theoretically (Larsen-Freeman 2006; Skehan 2003).

In order to elicit CAF measures relevant for our research, a pilot was carried out on the basis of 10 randomly chosen examples of learners' oral production and 10 examples of written production. Learners' performance was analyzed by means of well-established measures of fluency, accuracy and complexity. On the basis of the pilot, CAF measures that correspond best

to the research questions and to the peculiarities of the obtained data (Russian morphology with a high number of inflected forms; students' low level of proficiency) were chosen.

5.9.2.2 Accuracy

Accuracy is characterized as the ability to produce target-like and error-free language (Housen *et al.* 2012; Ellis & Barkhuizen 2005; Wolfe-Quintero *et al.* 1998). The notion of accuracy covers two subdimensions: grammaticality (correctness), that is, correspondence to standard prescriptive target language norms, and acceptability, that is, is “a non-standard usage fully acceptable in some social contexts” (Housen *et al.* 2012, p. 4). In our study accuracy was equated with correctness since no examples of ungrammatical but acceptable errors (or vice versa) were found. It seems that the notion of acceptability is more relevant for higher levels of proficiency.

5.9.2.2.1 General accuracy

There are two main approaches to measuring general accuracy in L2 research: (1) the percentage of error-free T-units (Gilabert 2004; Ortega 1999; Polio 1997; Rahimpour 1997; Robinson 1995a) or error-free clauses (Foster & Skehan 1996; Skehan & Foster 1997), and (2) the ratio of errors in a text to some unit of production (e.g. words, clauses, sentential units) (Bardovi-Harling & Bofman 1989; Mehnert 1998; Polio 1997).

The percentage of error-free T-units was calculated by dividing the number of error-free T-units by the total number of T-units and multiplied by 100. Following Hunt (1965), by T-unit we understand a main clause plus any subordinate clause attached to or embedded within it. Error-free T-unit in the current study is a unit that does not contain any lexical, morphological or syntactic error. Errors in spelling and punctuation were not included in error count as they are usually regarded as acceptable for this level of proficiency in common teaching practice.

In the application of this measure two problems arise. The first problem has already been discussed in SLA literature. This measure does not take into account the number of errors within one T-unit (or any other unit of analysis) and, consequently, a T-unit containing a single error is treated identically to a T-unit containing multiple errors (Bardovi-Harling & Bofman 1989; Polio

1997). With inflectional languages it becomes a serious problem because the distribution of errors can be very unfair in showing differences between learners with different levels of command of that language. Mehnert (1998) came to the same conclusion in her study of German as L2. The second problem became obvious after the analysis of the error-free units. We observed that, at this level of command, the use of prefabricated chunks influences the score, since most learners produce error-free statements almost only in these prefabricated chunks (e.g. *Как дела?* 'How are you?'; *Меня зовут* 'My name is'). Other frequent examples of error-free T-units produced by learners are short phrases with no case forms (*Я люблю читать* 'I like reading,') or copulative sentences where all the words are in the nominative case (*Это хорошо* 'It's good'; *Это большой и красивый город* 'It is a big and beautiful city'). However, these phrases do not reveal the real capacity of creating new utterances and cannot serve as the index of general accuracy.

For this reason, we chose the second approach to measure accuracy, which is the analysis of how many errors occur in relation to the total number of words produced. It was calculated by dividing the total number of errors by the total number of words and multiplied by 100.

$$\frac{\text{Number of errors}}{\text{Total number of words}} \times 100$$

There were a few specific cases in the error count that should be commented on. First, learners' production contained words with both grammatical and lexical errors. In this case the error was counted twice. For example:

Я была в Гранаде с *друга Аня.
 Ya byla v Granad'e s *druga An'a.
 I was in Granada with my *friend (male) An'a.

In this example the word *друга* [drúga] (male friend, Gen.Sing.) contain two errors: a lexical error (the word for female friend in Russian is *подруга* [padrúga]) and an error in the case form (Instrumental Case must be used after the preposition **with**: *с подругой* [spadrúgaj]).

Second, it was important to determine how to treat spelling errors since an incorrect spelling could result in a lexical error. It should be mentioned that the Cyrillic alphabet entails more difficulties in spelling than the Latin alphabet, especially at low level of proficiency, and spelling errors are quite common. As mentioned above, spelling errors were not counted.

However, there were a few cases when incorrect spelling provoked inaccurate vocabulary use. These cases received special attention as they could be interpreted in different ways. Learners of Russian who use the Roman alphabet in their L1 tend to confuse letters **п** [p] and **р** [r]; **б** [b] and **в** [v]; **и** [i], **ы** [ɨ] and **у** [u] and miss soft (**ь**) and hard signs (**Ъ**). This is the reason for mistakes like **пара** [pára] ‘a couple’ instead of **папа** [pápa] ‘dad’. All errors caused by incorrect spelling of these letters were considered spelling errors and, consequently, were not counted. However, in other cases such as **кафе** [kafé] ‘café’ instead of **кофе** [kófe] ‘coffee’ they were counted as lexical errors because they are caused by other factors- in this case the misleading equivalence with Spanish ‘café’.

5.9.2.2.2 Specific accuracy

Apart from evaluating general accuracy, specific accuracy measures were applied. For English it was target-like use of articles (Robinson 1995a; Rahimpour 1997; Gilabert 2004), verbal morphology (Wigglesworth 1997), plurals (Crookes 1989) and vocabulary (Skehan & Foster 1997). Target forms in the current study have been case forms and prepositions for Rus I, and verbs of motion with prefixes for Rus II. This predetermined the choice of specific accuracy measures.

In the analysis of oral task 1 (describe a room) and of written data collected in Rus I groups, the target-like use (TLU) of case forms and the target-like use of prepositions were calculated. Errors in gender or plural form (e.g. incorrect plural form for friends **друзи* [drúgi] instead of *друзья* [druz’já]) in the nominative case were not counted as errors in case form, rather, they were included in error count for measuring general accuracy. The TLU of case forms was calculated by dividing the number of correctly used case forms (including the nominative case) by the total number of case forms in the text produced and multiplied by 100. The following formula was applied:

$$\frac{\text{Number of accurately used case forms}}{\text{Total number of case forms}} \times 100$$

A similar formula was used to measure the TLU of prepositions. The only difference was that the total number of prepositions included not only all the prepositions produced by learners

but also missing prepositions, that is, the contexts where a preposition should have been used and was not used. In the example below the preposition **в** [v] ‘on’ is missing:

*Понедельник я посмотрел один фильм.

*Pon'ed'el'n'ik ya posmotr'el od'in f'il'm.

*Monday I watched a film.

The formula for measuring the TLU of prepositions was the following:

$$\frac{\text{Number of accurately used prepositions}}{\text{Number of prepositions produced} + \text{Number of obligatory contexts}} \times 100$$

In data obtained in Rus II groups both quantitative and qualitative aspects of the use of verbs of motion were analyzed. The target-like use of verbs of motion in written and oral production was calculated using the following formula:

$$\frac{\text{Number of accurately used verbs of motion}}{\text{Total number of verbs of motion produced}} \times 100$$

After analyzing the TLU of verbs of motion, we found out that most errors in the use of these verbs were grammatical (verbal aspect, tense, and, less frequently, person and singular/plural forms). It means that students could have learned the target items as lexical items but could not always use them correctly due to their morphological complexity and variety. For this reason, we considered it important to evaluate not only the target-like use of verbs of motion but also the total number of these verbs produced by learners in their written and oral task performance. It is not a measure of accuracy, but this number can be considered an indicator of lexical growth.

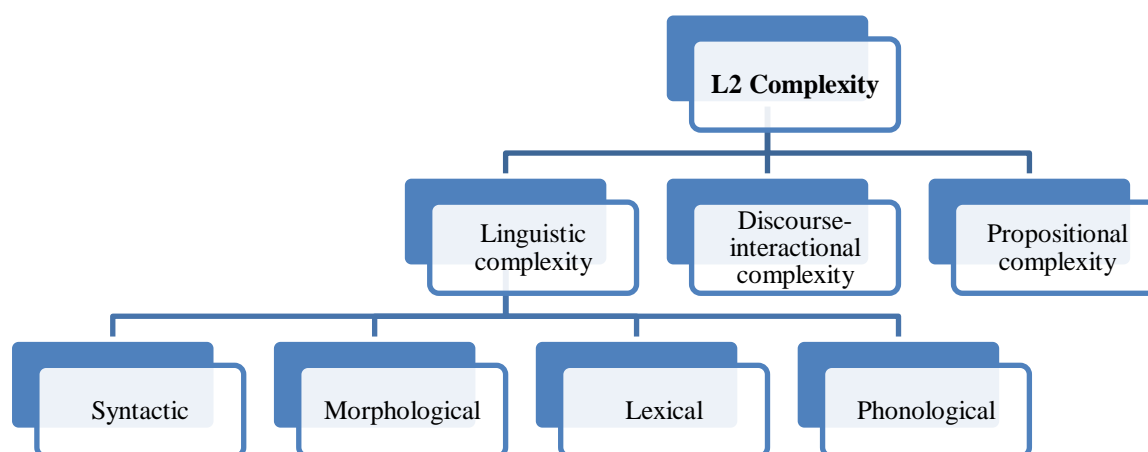
5.9.2.3 Complexity

Complexity is considered as the most problematic construct of the CAF triad because of its polysemous nature (Pallotti 2009). It is usually characterized as the ability to use a wide and varied range of sophisticated structures and vocabulary in the L2 (Housen *et al.* 2012).

Complexity refers to the extent to which learners produce “progressively more elaborated language” and “a greater variety of syntactic patterning” (Foster & Skehan 1996, p. 303), and by elaboration it is meant “learners’ willingness to use more challenging and difficult language and their preparedness to use a wide range of different structures” (Ellis & Barkhuizen 2005, p. 139).

Complexity is often viewed as a multidimensional construct with several sub-constructs (Bulté & Housen 2012; Ellis & Barkhuizen 2005; Norris & Ortega 2009). Bulté and Housen (2012, p. 23) presented a detailed and elaborated taxonomy of complexity constructs distinguishing three components: linguistic, discourse-interactional and propositional complexity (see Figure 10). Propositional complexity “refers to the number of information or idea units which a speaker/writer encodes in a given language task to convey a given message content” (Bulté & Housen 2012, p. 24). Discourse-interactional complexity has been proposed in the analysis of dialogic discourse and has been characterized in terms of the number and type of turn changes. These two types of complexity are still relatively new notions (Ellis & Barkhuizen 2005; Pallotti 2009). In our research we focused only on the notion of linguistic complexity. Linguistic complexity can be evaluated across various language domains (phonology, lexis, morphology, syntax). Lexical and syntactic complexity have the prior attention in L2 research.

Figure 10. Sub-dimensions of L2 complexity (Bulté & Housen 2012, p. 23)



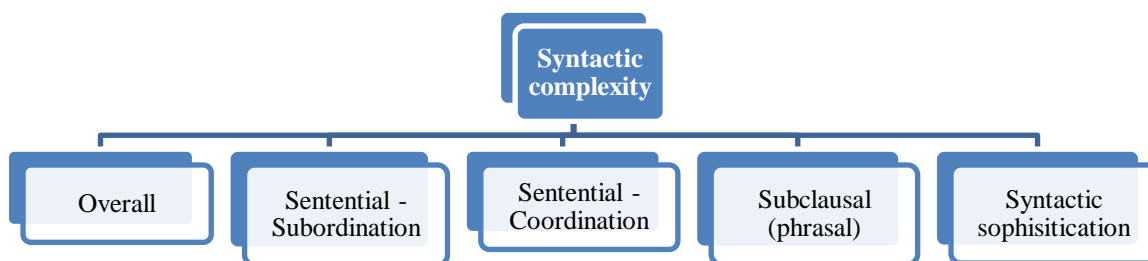
5.9.2.3.1 Syntactic complexity

Norris and Ortega (2009) identified five sub-constructs in syntactic complexity as illustrated in Figure 11:

1. Complexity via subordination, which is measured by any metric with clause in the numerator (e.g. clauses per T-unit, clauses per AS-unit, S-nodes per T-unit, etc.);
2. Overall or general complexity, which is measured by any length-based metric with a potentially multiple-clausal unit of production in the denominator (mean length of T-unit, mean length of AS-unit, mean length of turn, among others);
3. Subclausal or phrasal complexity via phrasal elaboration, which is measured by mean length of clause;
4. Complexity via coordination;
5. Complexity defined as the variety, sophistication, and acquisitional timing of forms produced.

SLA researchers have not explored the last option sufficiently, although some (e.g. Ellis & Yuan 2005) have counted the raw frequency of certain grammatical forms thought to be more sophisticated, less frequent, or later acquired, such as modals, passives, or infinitival phrases.

Figure 11. Sub-constructs of syntactic complexity (Norris & Ortega 2009)



The most commonly used ways of operationalizing syntactic complexity have focused on the amount of subordination. It has been measured by the ratio of clauses to a sentential unit of analysis, such as the terminal unit (T-unit) (Hunt 1965), the communication unit (c-unit) (Pica *et al.* 1989), or the analysis of speech unit (AS-unit) (Foster *et al.* 2000).

In our study the ratio of clauses to a T-unit was calculated. Following the definitions given by Hunt (1965) and Polio (1997), by clause we understand a structure with a subject and a finite verb. T-unit is a main clause plus any subordinate clause attached to or embedded within it (Hunt 1965). In order to measure sentential complexity, the following formula was applied:

$$\frac{\text{Number of clauses}}{\text{Number of T-units}}$$

Measuring complexity via coordination is a less common practice in SLA research. Bardovi-Harling (1992) proposed a coordination index that indicates the degree to which a learner achieves syntactic complexity through coordination. According to her formula, to calculate the coordination index the number of independent-clause coordinations must be divided by the number of combined clauses (the number of sentences subtracted from the total number of clauses) and multiplied by 100. This measure has an obvious limitation. It might be applied for analysis of written data taking into account learners' punctuation which marks the borders of sentences. However, measuring coordination in oral data obtained in the present study was problematic since in most cases it was not possible to distinguish with absolute certainty between two different sentences and two independent clauses within one sentence. For this reason, the coordination index was not calculated in our study.

Analyzing commonly used length-based measures, Norris and Ortega (2009) argued that it is important to distinguish possible sources of complexification. In the case of a potentially multiple-clausal denominator (e.g. utterance, T-unit, c-unit or AS-unit), it can become longer in several possible ways, both by the addition of subordinate clauses and by adding adjectives, prepositional phrases, etc. Norris and Ortega interpret the mean length of a potentially multi-clausal production unit as an index of overall syntactic complexity. However, the mean length of clause is different from the other length-based measures as clause length taps a more narrowly defined source of complexification, so it must be considered a specific measure that taps complexification at the phrasal level.

The pilot showed that learners barely used multiple-clausal units in their oral production which means that one T-unit often consisted of one clause. Ishikawa (1995) argued that the clause is a more sensitive unit for analysis of the data collected from beginner learners since it provides a smaller context for examining language growth in a variety of ways. For this reason, the Mean Length of Clause was chosen as a second measure of syntactic complexity. It was calculated by dividing the total number of words by the total number of clauses.

$$\frac{\text{Number of words}}{\text{Number of clauses}}$$

Finally, measures of syntactic sophistication could not be applied in our research. Due to their low level of proficiency, participants in our study were not able to produce forms which would be considered sophisticated in Russian (e.g. comparatives, participles or gerunds).

5.9.2.3.2 Lexical complexity

The concept of lexical complexity (or lexical richness) covers several aspects of vocabulary use: 1) lexical diversity (the variety of active vocabulary deployed by a speaker or writer), 2) lexical density (the ratio of lexical to function words), and 3) lexical sophistication (the number of low frequency words) (Bulté & Housen 2012; Daller & Phelan 2007; van Hout & Vermeer 2007).

The basic measure of lexical diversity is the Type-Token Ratio (TTR) (Ortega 1999; Robinson 1995b), that is, the ratio of different words used (types) to total number of words in the text (tokens). However, this measure has been strongly criticized as unreliable in contexts where texts with different lengths are compared, since the higher the number of tokens, the lower the ratio (Gilabert 2004; Skehan & Foster 2012; van Hout & Vermeer 2007). In order to correct the negative correlation between type/token results and the text length, Guiraud's Index of lexical richness has been proposed as an efficient measure because by including the square root of the tokens it compensates for differences in text length. Guiraud's Index has been calculated by means of the following formula:

$$\frac{\text{Types}}{\sqrt{\text{Tokens}}}$$

Lexical sophistication reflects on the extent to which a speaker accesses less frequent words in the second language lexicon. There are a number of frequency-based measures of lexical richness in English: Advanced Guiraud (Daller & Phelan 2007), Vermeer's measure of lexical richness (van Hout & Vermeer 2007), and P-Lex (Meara & Bell 2001), a computer program that divides a text into ten-word chunks, and then calculates the number of infrequent words used in each ten-word chunk. The concept of lexical sophistication seems to be more relevant for higher levels of proficiency and was not measured in our study.

5.9.2.4 Fluency

Fluency can be defined as the capacity to produce speech at normal rate and without interruption (Skehan 2009) or without undue pausing or hesitation (Ellis & Barkhuizen 2005;

Housen *et al.* 2012). In task-based research, fluency is often conceptualized as 1) speed fluency, which relates to rate of delivery; 2) breakdown fluency, which relates to number, length and location of pauses, and 3) repair fluency, or hesitation phenomena, which relates to dysfluency (the frequency of false starts, reformulations, repetitions and self-corrections) (Bosker *et al.* 2012; Kormos & Dénes 2004; Skehan 2003, 2009).

One of the most commonly used measures of oral fluency is speech rate which is usually calculated as a ratio of syllables produced to time taken to produce them. The main advantage of this kind of measure is that it in fact includes both the amount of speech and the length of pauses, since it takes into account the number of syllables and the total number of seconds in the narrative (Griffiths 1991). It was supported by the research of Kormos and Dénes (2004) who by means of the rank-order correlations between different fluency measures discovered that speech rate, the mean length of run (an average number of syllables produced in utterances between pauses of 0.25 seconds and above) and the mean length of pauses strongly correlate with each other.

In order to calculate speech rate in oral production, the total number of syllables produced in a given speech sample was divided by the total time required to produce the speech sample (including pause time) expressed in seconds. This figure was then multiplied by sixty to produce a figure expressed in syllables per minute. The following formula was used:

$$\frac{\text{Number of syllables}}{\text{Total number of seconds}} \times 60$$

Fluency of written production could not be measured in our study because we did not know how much time learners spent to accomplish the written task. They had one hour and a half to do both the grammar tests and the written task, and it was not observed how each participant distributed his/her time between these two parts. During the treatment sessions learners did not perform any written task within limited time; for this reason, they worked under the same conditions during data collection.

5.9.2.5 Summary

Below is the summary of CAF measures used in this study.

General accuracy

- Errors per words

Specific accuracy (oral and written tasks)

- Target-like use of prepositions
- Target-like use of case forms
- Target-like use of verbs of motion

Specific accuracy (grammar tests)

- Fill in the blanks
- Multiple choice
- Grammaticality judgment

Syntactic complexity

- Mean length of clause
- Clauses per T-unit

Lexical complexity

- Guiraud's Index of lexical richness

Fluency

- Speech rate (syllables per minute) of oral production

Besides CAF measures, the total amount of verbs of motion in written and oral production of Rus II groups was counted.

5.10 Statistical analyses

In order to analyze the obtained data, a number of statistical analyses have been performed in the current study. Descriptive statistics provided information about means, medians, standard deviations, skewness, and kurtosis. Data were screened for outliers and checked for normality of distribution, the assumption of sphericity and for homogeneity of variances. For data which were normally distributed and did not violate the assumptions of sphericity and homogeneity of variances, a Mixed Between-Within Subjects ANOVA was performed. In cases when data did not follow normal distribution, nonparametric tests (Friedman

tests, Wilcoxon signed rank tests and Mann-Whitney tests) were conducted to compare the scores at the pre-test, immediate post-test and delayed post-test within each group of participants and between the two groups.

All statistical analyses were carried out using SPSS 21. Their results will be discussed in the following chapter.

CHAPTER VI

RESULTS

6.1 Introduction

This chapter presents the results of statistical analyses of the obtained data. First, we report the results of exploratory data analyses (normality of distribution, homogeneity of variances, and sphericity) which helped us to choose the adequate statistical instruments. Then the results of group comparisons on pre-tests are displayed. The aim of these comparisons is to confirm the comparable level of learners' proficiency in task-supported and task-based groups at the moment when the experiment started.

The results of within-subjects and between-subjects comparisons on all tests are divided into three main sections which correspond to the three research questions advanced in Chapter V. First, means, standards deviations and in some cases medians for each set of dependent variables are provided. Then the results of statistical analyses are reported in order to determine whether the learners in each group have improved over time and whether one type of treatment has been more efficient than another in terms of the measures applied in this study. The results of statistical analyses are summarized in the last section. The discussion of the obtained results will be presented in the next chapter.

6.2 Preliminary analyses

In order to answer the research questions about the efficiency of task-supported (TS) and task-based (TB) treatment described in Section 5.3, we had to determine whether there was any significant difference in scores in each group over time and whether there was a significant difference between the results obtained by the TS group and the TB group. A statistical test which provides the most relevant information for our research is a mixed between-within subjects ANOVA. This test compares the mean differences between groups that have been split on two "factors" (independent variables), where one factor is a "within-subjects" factor and the other factor is a "between-subjects" factor. In our study, we have measured a number of

dependent variables summarized in Section 5.9.2.5 (e.g. errors per words, speech rate, clauses per T-unit, etc.) for two groups (the TS group and the TB group) over three time points (pre-test, immediate post-test and delayed post-test). Therefore, the independent variables in the current research are *Time* and the type of treatment (further *Method*). *Time* and *Time x Method* interaction are “within-subjects” factors in the present study. *Method* forms a “between-subjects” factor.

A number of assumptions had to be checked to make sure that a mixed between-within subjects ANOVA could be used to analyze data. This included: (a) to check that there were no significant outliers in any group of within-subjects factor or between-subjects factor; (b) to determine that the dependent variable was normally distributed; (c) to check the assumption of sphericity; and (d) to check for homogeneity of variances (Field 2009; Tabachnik & Fidell 2013).

In the course of preliminary analyses, the data were screened for outliers. Extreme outliers were removed in order to achieve the normality of score distribution.⁸ The normality of distribution was checked by means of Shapiro-Wilks test. A significant result ($p < .05$) in this test means that data are non-normality distributed. For scores that followed normal distribution, parametric tests (mixed between-within subjects ANOVA, repeated measures ANOVA, and independent samples t-tests) were used. When the results of ANOVA were found to be statistically significant, post hoc paired samples t-tests were run in order to highlight where exactly these differences occurred. The data which were not normally distributed were analyzed by means of nonparametric tests (Friedman test, Wilcoxon signed-ranks test, and Mann-Whitney test).

The sphericity assumption was checked by means of Mauchly’s sphericity test. For measures that violated the sphericity assumption (i.e., $p < .05$ in Mauchly’s sphericity test), results from multivariate test statistics (Wilks’ Lambda) automatically produced by the repeated-measures procedure in SPSS were reported (Field 2009). Levene’s tests were conducted to check the assumption of homogeneity of variances. In the cases when the assumption of homogeneity of variances was violated ($p < .05$ in Levene’s tests), the differences between groups were assessed by means of independent samples t-tests.

An alpha level of $p < .05$ was adopted for all statistical tests. Following the recommendations of Field (2009), the effect sizes were reported as r (Pearson’s correlation coefficient) for t-tests and as *partial* η^2 (partial eta squared) for ANOVA. Following Cohen

⁸ Boxplots were used to identify outliers for each measure. As a consequence of the elimination of extreme outliers, descriptive statistics may show a different n for each measure.

(1988), measures of effect size in ANOVA were interpreted by partial eta squared and classified as small ($.01 < \eta^2 < .06$), medium ($.06 < \eta^2 < .14$), or large ($\eta^2 > .14$). When effect sizes were reported as r , they were qualified in the following way: small $r = .10$, medium $r = .30$ and large $r = .50$.

6.3 Group comparability

6.3.1 TS_Rus I and TB_Rus I

In order to determine whether the TS group and the TB group were comparable and had the same level of proficiency at the beginning of the experiment, the results of pre-tests were compared. Independent samples t-tests were run for the data that were normally distributed. Mann-Whitney tests were performed for the target-like use (TLU) of prepositions and clauses per T-unit measures since the obtained results for these two measures did not follow normal distribution (see Table 31). The comparison of the two groups' pre-test scores yielded non-significant differences between the TS and TB groups for all measures ($p > .05$). These results suggest that participants in both groups did not differ at the pre-test in any of the measures of their written production and on grammar scores.

Table 31. Comparing participants on the written pre-test (Rus I)

Dependent variable		Independent samples t-test			Mann-Whitney test		
		<i>t</i>	<i>df</i>	<i>p</i>	<i>U</i>	<i>Z</i>	<i>p</i>
Accuracy	Errors per words	.09	29	.926			
	TLU of prepositions				96.00	-.35	.724
	TLU of case forms	-.40	29	.694			
Complexity	Mean length of clause	-.76	29	.455			
	Clauses per T-unit				114.50	-.18	.856
Grammar	Fill in the blanks	1.32	28.10	.197			
	Multiple choice	-.48	33	.632			
	Grammaticality judgment	.01	33	.995			

6.3.2 TS_Rus II and TB_Rus II

Rus II groups, in addition to a written task and three grammar tests, performed two oral tasks. The comparison of the two groups' written pre-test results and grammar scores yielded significant differences between the groups in three measures: TLU of verbs of motion ($U = 73.00$,

$Z=-1.96$, $p=.05$, $r=.35$), Guiraud's Index ($t(29)=2.38$, $p=.024$, $r=.40$), and multiple choice scores ($t(22,88)=3.39$, $p=.003$, $r=.54$). The effect sizes for TLU of verbs of motion ($r=.35$) and Guiraud's Index ($r=.40$) scores were medium⁹, which has allowed us to perform a reliable comparison of the two groups at the immediate and delayed post-tests. However, the difference between the two groups in multiple choice scores was large as indicated by a large effect size $r=.54$. There could be a number of reasons for this difference in scores such as the character of the test itself (a possibility of choosing a correct option by chance) and an effort made by students in both groups while completing the test. This initial advantage of the TS group over the TB group was taken into consideration when comparing the groups on post-tests.

The participants in the two groups did not statistically significantly differ on the written pre-test in other measures used in the study ($p>.05$), as shown in Table 32.

Table 32. Comparing participants on the written pre-test (Rus II)

Dependent variable		Independent samples t-test			Mann-Whitney test		
		<i>t</i>	<i>df</i>	<i>p</i>	<i>U</i>	<i>Z</i>	<i>p</i>
Accuracy	Errors per words	-1.06	28	.298			
	Total amount of verbs of motion	.02	28	.984			
	TLU of verbs of motion				73.00	-1.96	.05
Complexity	Mean length of clause	-.44	27	.662			
	Clauses per T-unit	-.30	29	.765			
	Guiraud's Index	2.38	29	.024			
Grammar	Fill in the blanks	1.21	31	.235			
	Multiple choice	3.39	22.88	.003			
	Grammaticality judgment	1.71	25.86	.100			

For oral task 1, there were non-significant differences between the TS and TB groups ($p>.05$) in all the scores (see Table 33). These results suggest that participants in the two groups demonstrated a similar level of CAF skills and knowledge of the target forms (prepositions and case forms) at the oral pre-test.

Table 33. Comparing participants on the oral pre-test, Task 1

Dependent variable		Independent samples t-test		
		<i>t</i>	<i>df</i>	<i>p</i>
Fluency	Speech rate	.92	27	.367
Accuracy	Errors per words	-.79	16.99	.441
	TLU of prepositions	-.24	27	.814
	TLU of case forms	.99	27	.333

⁹According to Cohen (1988), $r=.10$, $r=.30$ and $r=.50$ indicate small, medium, and large effect sizes, respectively.

Complexity	Mean length of clause	.01	28	.989
	Guiraud's Index	.45	29	.653

Finally, scores obtained by participants for oral task 2 were compared. Independent samples t-test showed that there was a statistically significant difference between the two groups at the pre-test ($t(29)=2.30$, $p=.029$, $r=.39$). The effect size $r=.39$ suggests that the difference between the groups at the pre-test was moderate, which has allowed us to perform a reliable analysis of the differences between the TS and TB groups at post-tests. There were non-significant differences between the TS and TB groups ($p>.05$) on other scores (see Table 34).

Table 34. Comparing participants on the oral pre-test, Task 2

Dependent variable		Independent samples t-test			Mann-Whitney test		
		<i>t</i>	<i>df</i>	<i>p</i>	<i>U</i>	<i>Z</i>	<i>p</i>
Fluency	Speech rate	.13	28	.897			
Accuracy	Errors per words	-.07	27	.943			
	Total amount of verbs of motion				93.50	-1.71	.088
	TLU of verbs of motion				121.00	-.38	.707
Complexity	Mean length of clause	-1.49	27	.146			
	Clauses per T-unit				80.00	-1.62	.105
	Guiraud's Index	2.30	29	.029			

6.4 The effects of task-supported and task-based treatment on the use of prepositions and case forms

The purpose of this section is to answer RQ1 and find out whether each type of treatment had any effect on learners' use of Russian case forms and prepositions, on the one hand, and whether one approach was more efficient than the other. In order to answer this question, we first compared the percentage of the target-like use of these forms produced by participants in the TS group and the TB group in their written and oral performance. Then we analyzed the results of three grammatical tests performed by these two groups.

6.4.1 Target-like use of prepositions in written and oral production

Descriptive statistics demonstrated consistent improvement in the use of prepositions by learners in the TB group both in their written and oral performance (see Table 35), as illustrated in Figure 12 and Figure 13, respectively.¹⁰ The best results were achieved by the TB group in the written delayed post-test. The median of this group reached the maximum score ($Mdn=100$), and the mean score was also the highest as compared to other scores on this measure ($M=93.62$, $SD=9.38$). It suggests a very positive and long-lasting effect of task-based treatment on learners' use of prepositions both in their written and oral performance.

Table 35. Descriptive statistics: TLU of prepositions

Task	Group	Pre-test		Immediate post-test		Delayed post-test	
		Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median
Written task	TS_RusI (n=17)	81.19 (19.48)	87.30	89.57 (8.68)	90.83	91.34 (7.91)	90.83
	TB_RusI (n=14)	86.76 (9.87)	86.36	93.16 (6.26)	92.86	93.62 (9.38)	100.00
Oral task 1	TS_RusII (n=13)	69.51 (18.77)	72.73	69.23 (21.77)	72.73	72.13 (14.76)	73.68
	TB_RusII (n=16)	71.13 (17.82)	66.82	72.86 (14.38)	77.75	78.31 (13.83)	82.73

Learners in the TS group only slightly improved their use of prepositions from the pre-test to the delayed post-test, which means that the task-supported treatment had almost no effect on their specific accuracy. The TB group outperformed the TS group at the immediate and delayed post-test both in oral and written performance.

¹⁰ For more descriptive statistics (means, standard errors, standard deviations, medians, skewness, and kurtosis) see Appendix D.

Figure 12. Median scores on TLU of prepositions (written task, Rus I)¹¹

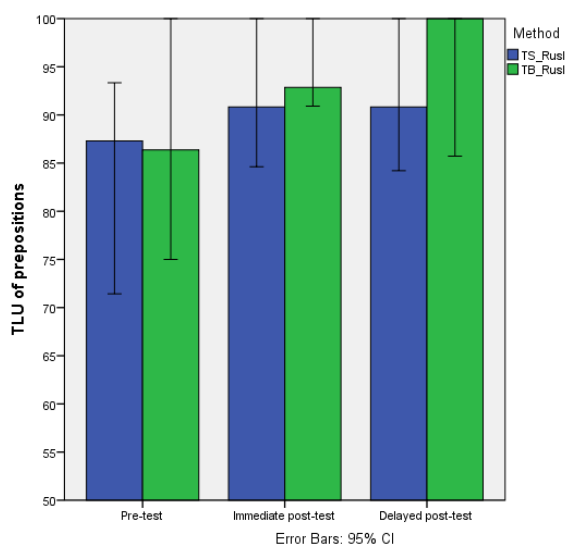
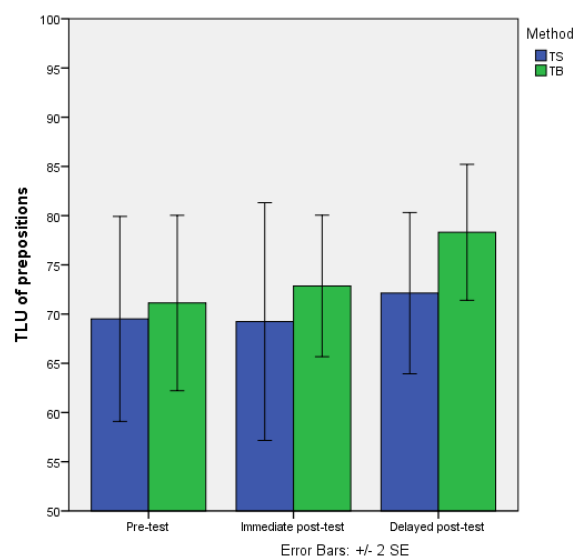


Figure 13. Mean scores on TLU of prepositions (oral task 1, Rus II)



The next step was to determine whether differences in scores within each group and between the two groups reached the level of statistical significance. Preliminary analysis showed that scores on the TLU of prepositions in written production did not follow a normal distribution. For this reason, nonparametric tests (Friedman test and Mann-Whitney test) were used. However, the analysis of oral production (oral task 1) showed that the scores on this measure were normally distributed. This allowed us to use parametric tests (mixed between-within subjects ANOVA) for statistical analysis of oral data.

Friedman tests were performed to assess whether there was a significant change in the TLU of prepositions scores within the two groups of participants over time. The results obtained suggest that although there was a tendency of positive change of these scores over time, there were no significant differences ($p > .05$) in each group of participants (see Table 36).

Table 36. Friedman test: TLU of prepositions (written task, Rus I)

Group	Chi-square	df	p
TS_RusI	3.73	2	.155
TB_RusI	3.17	2	.205

In order to assess the differences in scores on the TLU of prepositions between the TS and TB groups, a Mann-Whitney test was performed. The results showed that the TS and TB groups did not significantly differ in the pre-test ($U=96.00$, $p=.724$), the immediate post-test

¹¹ The graphs presenting the dynamic of the two groups on the TLU of prepositions in written production are based on median values because the scores did not follow a normal distribution.

($U=76.50$, $p=.223$), and the delayed post-test ($U=84.00$, $p=.351$) scores. The results are presented in Table 37.

Table 37. Mann-Whitney test: TLU of prepositions (written task, Rus I)

	<i>U</i>	<i>Z</i>	<i>p</i>
Pre-test	96.00	-.353	.724
Immediate post-test	76.50	-1.220	.223
Delayed post-test	84.00	-.932	.351

For statistical analysis of oral data, a mixed between-within subjects ANOVA was conducted. The results showed (see Table 38) that there was a statistically non-significant effect of *Time* on the use of prepositions in oral production ($F(2)=.89$, $p=.414$, *partial* $\eta^2=.032$). The *Time x Method* interaction was not significant ($F(2)=.17$, $p=.846$, *partial* $\eta^2=.006$) either. The analysis of between-subjects effects showed a statistically non-significant effect of *Method* ($F(1)=.75$, $p=.393$, *partial* $\eta^2=.027$). This result suggests that the TS group and the TB group did not significantly differ in their use of prepositions. Effect sizes (*partial eta squared*) for *Time*, *Method*, and *Time x Method* interaction were small (*partial* $\eta^2<.06$).

Table 38. Mixed between-within subjects ANOVA: TLU of prepositions (oral task 1, Rus II)

	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
<i>Time</i>	400.94	2	.89	.414	.032
<i>Time x Method</i>	74.81	2	.17	.846	.006
<i>Method</i>	311.92	1	.75	.393	.027

df= Degrees of freedom; η^2 = *partial eta squared* (effect size).

Therefore, the results obtained suggest that although there was a tendency towards positive change of scores on the TLU of prepositions over time, this change was not significant in each group of participants in their written and oral performance. The results of between-groups comparisons demonstrate that the TB group outperformed the TS group both in oral and written post-tests, although not statistically significantly. This means that the task-based treatment worked better than the task-supported treatment as far as the learning of prepositions is concerned.

6.4.2 Target-like use of case forms in written and oral production

We then explored the effect of two treatments on learners' improvement in the use of case forms. The mean scores and the standard deviations for the written task (Rus I) and for the oral task 1 (Rus II) are given in Table 39. They are visually displayed below in Figure 14 and Figure 15, respectively.

Table 39. Descriptive statistics: TLU of case forms

	Written task		Oral task 1	
	TS_Rus I (n=17)	TB_Rus I (n=14)	TS_Rus II (n=13)	TB_Rus II (n=16)
Pre-test	82.90 (9.20)*	84.24 (9.39)	58.25 (20.43)	51.01 (19.03)
Immediate post-test	87.08 (6.07)	90.62 (5.64)	49.48 (14.74)	53.45 (18.00)
Delayed post-test	84.09 (10.10)	82.53 (10.14)	60.05 (16.36)	62.61 (21.46)

*Mean (SD)

Figure 14 illustrates that both the TS and TB groups improved their use of case forms at the immediate post-test, but then their specific accuracy decreased at the delayed post-test to the level they had at the pre-test. The TB group outperformed the TS group in the correct use of case forms at the immediate post-test, but demonstrated worse results than the TS group at the delayed post-test.

Figure 15, which represents mean scores for oral production, demonstrates different dynamics. The TB group constantly improved their use of case forms and showed the best results at the delayed post-test ($M=62.61$, $SD=21.46$). Mean scores of the TS group dropped at the immediate post-test ($M=49.48$, $SD=14.74$) as compared with the pre-test. However, at the delayed post-test participants in the TS group used cases more accurately ($M=60.05$, $SD=16.36$) than at the immediate post-test. The TB group outperformed the TS group at the two post-tests.

Figure 14. Mean scores on TLU of case forms
(written task, Rus I)

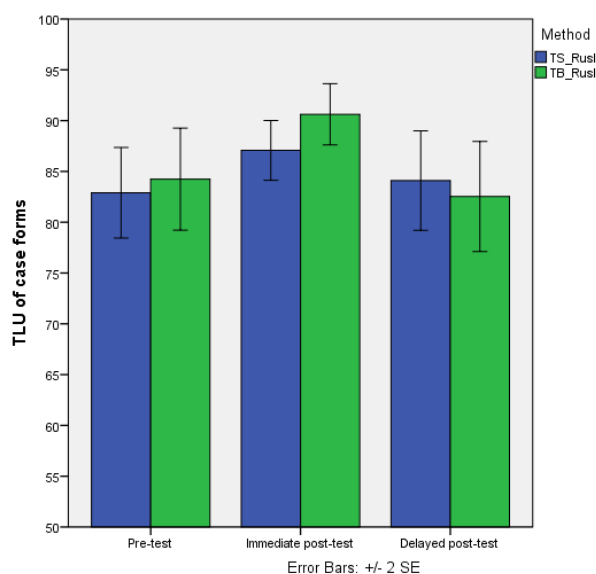
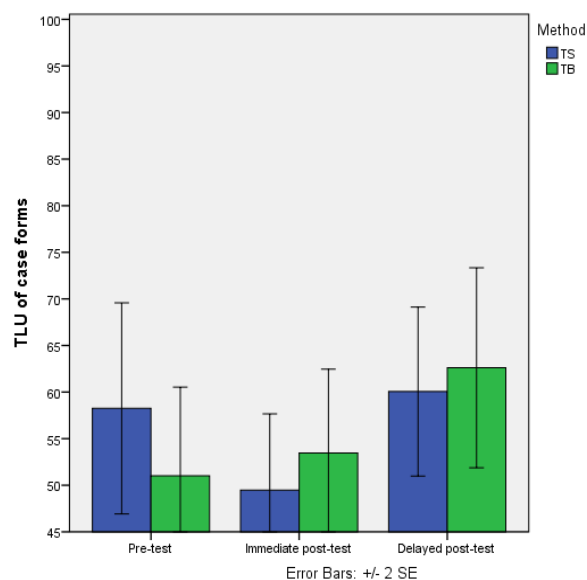


Figure 15. Mean scores on TLU of case forms
(oral task 1, Rus II)



The next step was to determine whether there was any significant improvement in the use of case forms by learners in each group over time (within-subjects effect) and whether there was a significant effect of treatment (Method) on use of these forms (between-subjects effect). A mixed between-within subjects ANOVA revealed (see Table 40) that there was a significant and large main effect of *Time* on the use of case forms in written production ($F(2,58)=5.62, p=.006, \text{partial } \eta^2=.162$). The *Time x Method* interaction was not significant ($F(2,58)=.95, p=.393, \text{partial } \eta^2=.032$), which means that the change in scores over time did not significantly differ for the two groups of participants.

Similar results were obtained in the analysis of oral data. There was a statistically significant effect of *Time* on the use of case forms ($F(2)=3.19, p=.049, \text{partial } \eta^2=.106$), whereas the *Time x Method* interaction did not reach statistical significance ($F(2)=1.17, p=.317, \text{partial } \eta^2=.042$).

Table 40. Within-subjects effects: TLU of case forms

		<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
Written task	<i>Time</i>	599.02	2,58	5.62	.006	.162
	<i>Time x Method</i>	101.15	2,58	.95	.393	.032
Oral task 1	<i>Time</i>	1455.51	2	3.19	.049	.106
	<i>Time x Method</i>	534.78	2	1.17	.317	.042

df= Degrees of freedom; η^2 = partial eta squared (effect size).

Given the significant effect of *Time* on the TLU of case forms, post hoc paired samples t-tests were conducted. The pairwise comparisons of pre-test, immediate post-test and delayed

post-test scores within each group were adjusted for multiple comparisons (Bonferroni correction). The analysis of written data showed that there were statistically non-significant differences between the TS group's scores over time ($p > .05$), which means that the participants in the TS group did not significantly improve their use of cases. In the TB group, the results showed a significant difference between the pre-test and the immediate post-test ($t(13) = -3.11$, $p = .025$, $r = .65$). The results suggest that the participants in the TB group used case forms significantly more accurately at the immediate post-test than at the pre-test. However, the comparison of the immediate post-test with the delayed post-test performed by the TB group showed a significant drop of scores ($t(13) = 3.15$, $p = .023$, $r = .65$), which means that the participants' use of case forms became less accurate in comparison with the immediate post-test. This means that a positive effect of the task-based treatment did not remain. The difference between the pre-test and the delayed post-test was statistically non-significant. This result suggests that the participants' accuracy in the use of case forms at the delayed post-test dropped to the level they had at the pre-test.

Paired samples t-tests were also conducted for oral data. The tests showed that there were statistically non-significant differences between the TB group's scores at the pre-, post- and delayed post-test ($p > .05$). This result suggests that although participants in the TB group improved their use of case forms, the difference in scores over time did not reach statistical significance. For the TS group, there was a statistically non-significant difference between the pre-test and the immediate post-test, which means that the decrease in accuracy scores we have seen in Figure 15 was not significant. We found a significant difference between the immediate post-test and the delayed post-test ($t(12) = -2.46$, $p = .030$, $r = .63$), which let us conclude that learners in the TS group significantly improved their use of case forms. The difference between the pre-test and the delayed post-test was statistically non-significant. This result suggests that the participants' accuracy in the use of case forms at the delayed post-test came back to the level they had at the pre-test.

The analysis of between-subjects effects showed a statistically non-significant effect of *Method* both in written production ($F(1,29) = .24$, $p = .626$, *partial* $\eta^2 = .008$) and oral production ($F(1) = .002$, $p = .964$, *partial* $\eta^2 = .000$). In both cases the effect size was also small (*partial* $\eta^2 < .06$). This result suggests that the TS group and the TB group did not significantly differ in the TLU of case forms. The results of the analysis of between-subjects effects are presented in Table 41.

Table 41. Between-subjects effects: TLU of case forms

	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
Written task	28.19	1, 29	.24	.626	.008
Oral task 1	1.19	1	.002	.964	.000

df= Degrees of freedom; η^2 = partial eta squared (effect size).

Taken together, the results obtained suggest that in written production both the TS and TB groups improved their use of case forms in the immediate post-test, but then in the delayed post-test learners' scores decreased to the level they had in the pre-test. This means that a positive effect of both task-supported and task-based treatment on learners' accurate use of case forms did not last long. Analysis of oral production yielded different results. The TB group demonstrated worse results than the TS group in the pre-test, but after the treatment improved their results and outperformed the TS group at the two post-tests. The TS group decreased their accuracy in the use of case forms at the immediate post-test, but then significantly improved on this measure at the delayed post-test. The TS group and the TB group did not significantly differ in their use of case forms at any of the tests.

6.4.3 Fill in the blanks

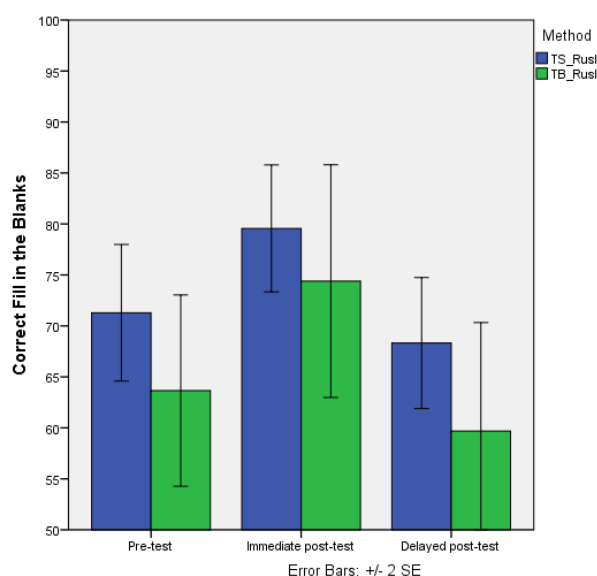
We also explored the effect of both treatments on learners' results in three grammatical tests: fill in the blanks, multiple choice, and grammaticality judgment. The mean scores and standard deviations for the fill in the blanks (FB) test are given in Table 42, and they are visually displayed in Figure 16.

Table 42. Descriptive statistics: Fill in the blanks (Rus I)

Group	Pre-test		Immediate post-test		Delayed post-test	
	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median
TS_RusI (<i>n</i> =19)	71.28 (14.60)	74.76	79.56 (13.57)	82.52	68.32 (14.02)	66.99
TB_RusI (<i>n</i> =16)	63.65 (18.78)	63.59	74.39 (22.82)	84.95	59.69 (21.27)	60.00

Both the TS and TB groups improved their results from the pre-test to the immediate post-test, but then their mean scores decreased in the delayed post-test to a slightly worse level than what they demonstrated in the pre-test, which means that the positive effect of treatment did not remain. The TS group outperformed the TB group in the pre-test and maintained this advantage in both post-tests.

Figure 16. Mean scores on fill in the blanks test (Rus I)



Preliminary analysis showed that scores on the FB test for the TS group followed normal distribution. Mauchly’s sphericity test showed nonsignificant result ($p>.05$). This allowed us to use a repeated measures ANOVA for statistical analysis of these data. However, the scores of the TB group on this measure were not normally distributed. For this reason, a Friedman test was conducted for analysis of the TB group results. The results of these two statistical tests are presented in Table 43.

Table 43. Within-subjects effects: Fill in the blanks (Rus I)

Group	Repeated measures ANOVA					Friedman test		
	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2	<i>Chi-Square</i>	<i>df</i>	<i>p</i>
TS_RusI	1290.23	2,36	16.24	.000	.474			
TB_RusI						17.37	2	.000

The results from the repeated measures ANOVA and the Friedman test showed that there was a significant main effect of *Time* for the TS group ($F(2,36)=16.24, p=.000, partial \eta^2=.474$) and for the TB group ($\chi^2(2)=17.37, p=.000$). These results suggest that there was a significant change of FB scores over time in both groups of participants.

In order to explore the differences in FB scores, pairwise comparisons of the participants’ scores in each group at the pre-test, immediate post-test and delayed post-test were conducted. Paired samples t-tests showed that the difference between the pre-test and the immediate post-test in the TS group was statistically significant ($t(18)=-4.02, p=.002, r=.47$), suggesting a significant improvement of the participants’ scores. There was a significant difference between

the immediate post-test and the delayed post-test FB scores ($t(18)=4.59$, $p=.001$, $r=.54$), which means that there was a significant drop of scores at the delayed post-test. There were statistically non-significant differences between pre-test and delayed post-test scores ($p>.05$). These results suggest that the participants' results dropped to the level they had at the pre-test.

For the TB group, Wilcoxon signed-ranks tests were conducted to compare FB scores at three time points. The results were similar to those obtained for the TS group. The tests showed a statistically significant improvement from the pre-test to the immediate post-test ($Z=-3.06$, $p=.002$, $r=.76$) and a statistically significant drop of scores in the delayed post-test ($Z=-3.05$, $p=.002$, $r=.76$). There were statistically non-significant differences between pre-test and delayed post-test scores ($p>.05$).

Since the assumption of homogeneity of variance was violated,¹² a series of independent samples t-tests and a Mann-Whitney test (for immediate post-test scores) were conducted in order to assess the differences between the two groups. The results obtained showed that there were no significant differences in FB scores between the TS and TB groups at the pre-, post- and delayed post-test as presented in Table 44.

Table 44. Between-subjects effects: Fill in the blanks (Rus I)

	Independent samples t-test			Mann-Whitney test		
	<i>t</i>	<i>df</i>	<i>p</i>	<i>U</i>	<i>Z</i>	<i>p</i>
Pre-test	1.32	28.10	.197			
Immediate post-test				145.00	-.23	.817
Delayed post-test	1.39	25.17	.177			

Taken together, the results obtained indicate that both the TS and TB groups significantly improved their scores in the FB test from the pre-test to the immediate post-test, but then demonstrated a significant drop of scores at the delayed post-test. These results suggest that the effects of both task-supported and task-based treatment did not remain. The TS group outperformed the TB group in both post-tests, although the difference between the two groups did not reach statistical significance.

¹²Levene's tests showed significant results ($p<.05$) for pre-, post- and delayed post-test scores.

6.4.4 Multiple choice

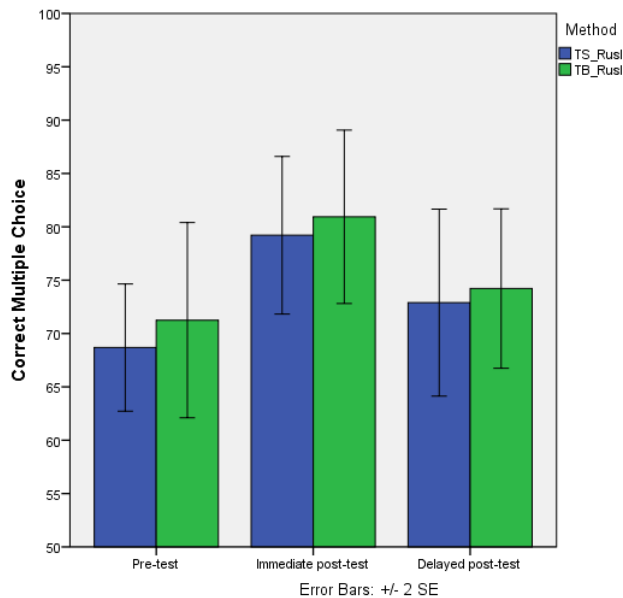
Descriptive statistics showed that both groups had a similar dynamic in the multiple choice (MC) scores (see Table 45). Learners improved their mean scores in the immediate post-test and then demonstrated a decrease in scores in the delayed post-test. The TB group slightly outperformed the TS group in all three tests as illustrated in Figure 17.

Table 45. Descriptive statistics: Multiple choice (Rus I)

Group	Pre-test	Immediate post-test	Delayed post-test
TS_RusI (n=19)	68.68 (13.00)*	79.21 (16.09)	72.89 (19.10)
TB_RusI (n=16)	71.25 (18.30)	80.94 (16.25)	74.22 (14.94)

*Mean (SD)

Figure 17. Mean scores on multiple choice test (Rus I)



In order to compare the results obtained by two groups, a mixed between-within subjects ANOVA was conducted. The results from this statistical test showed that there was a significant and large main effect of *Time* ($F(2,66)=7.19$, $p=.001$, $partial \eta^2=.179$). The *Time x Method* interaction was non-significant ($p>.05$), which means that the change in MC scores over time did not significantly differ for the two groups of participants. The analysis of between-subjects effects showed a non-statistically significant effect of *Method* ($p>.05$). This result suggests that there was no significant difference in scores on the multiple choice test between the two groups. The results of the mixed between-within subjects ANOVA are summarized in Table 46.

Table 46. Mixed between-within subjects ANOVA: Multiple choice (Rus I)

	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
<i>Time</i>	1824.11	2,66	7.19	.001	.179
<i>Time x Method</i>	6.97	2,66	.03	.973	.001
<i>Method</i>	91.34	1,33	.16	.687	.005

In order to explore the differences in MC scores, pairwise comparisons adjusted for multiple comparisons (Bonferroni correction) of the participants' scores in each group at the pre-test, immediate post-test and delayed post-test were conducted.

For the TS group, a statistically significant difference between the pre-test and the immediate post-test scores ($t(18)=-3.44$, $p=.009$, $r=.63$) was found. This suggests a significant positive effect of task-supported treatment on learners' MC scores. There was a statistically non-significant difference between the immediate and delayed post-test scores ($p>.05$), which means that although the participants had lower scores at the delayed post-test, the difference was not significant. Finally, the participants' scores at the delayed post-test were higher than those they had at the pre-test, but this difference was statistically non-significant ($p>.05$). This result suggests that learners' scores dropped at the delayed post-test to the level that learners had at the pre-test.

For the TB group, the pairwise comparisons performed showed that the differences between the the TB group's MC scores at the pre-test, immediate post-test, and delayed post-test were statistically non-significant ($p>.05$). These results suggest that although there was a tendency towards improvement from the pre-test to the delayed post-test, this improvement did not reach the level of statistical significance.

Therefore, the results of statistical tests suggest that there was a significant positive effect of the task-supported treatment on learners' use of case forms measured by means of a multiple choice test; however, this effect was not long-lasting. Analysis of scores obtained by the TB group in the two MC post-tests showed that task-based treatment had no effect on learners' use of case forms and prepositions.

6.4.5 Grammaticality judgment

Next, the results of grammaticality judgment (GJ) test were statistically analyzed. Descriptive statistics showed that both the TB and TS groups improved their mean scores in the immediate post-test as compared with the pre-test, and the TB group demonstrated slightly better

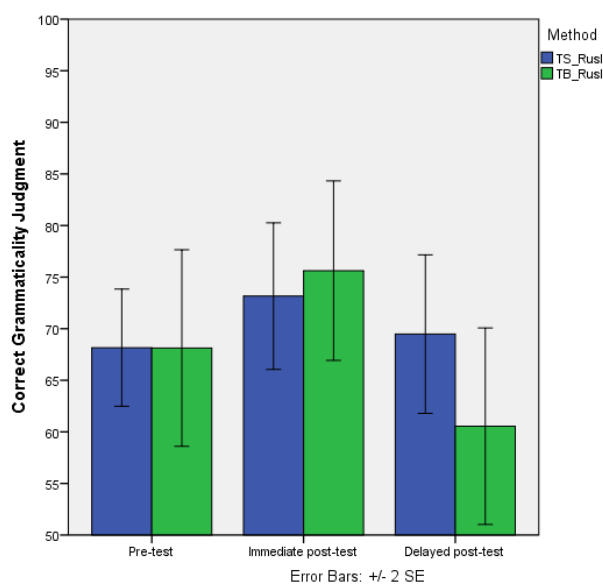
results than the TS group (see Table 47). However, there was a drop in the mean scores obtained by learners in the delayed post-test after the task-based treatment ($M=60.55$, $SD=19.05$) as illustrated in Figure 18. One of the reasons for this drop in scores could be the type of test, which requires explicit metalinguistic knowledge of specific grammatical forms and favors a traditional approach focused on forms. This possibility is confirmed by the fact that the TS group outperformed the TB group at the delayed post-test ($M=69.47$, $SD=16.74$) and obtained similar mean scores as at the immediate post-test, which means that the effect of the task-supported treatment was long-lasting.

Table 47. Descriptive statistics: Grammaticality judgment (Rus I)

Group	Pre-test	Immediate post-test	Delayed post-test
TS_RusI ($n=19$)	68.16 (12.38)*	73.16 (15.47)	69.47 (16.74)
TB_RusI ($n=16$)	68.13 (19.05)	75.62 (17.40)	60.55 (19.05)

*Mean (SD)

Figure 18. Mean scores on grammaticality judgment test (Rus I)



The next step was to check the normality of distribution and to conduct Mauchly's sphericity and Levene's tests in order to confirm that a mixed between-within subjects ANOVA could be used. A Shapiro-Wilks test showed that data were normally distributed. Mauchly's sphericity and Levene's tests showed nonsignificant results ($p>.05$), which let us use a mixed ANOVA in order to compare the results obtained by the two groups in the grammaticality judgment test.

The results from the mixed between-within subjects ANOVA (see Table 48) showed that there was a significant and large main effect of *Time* ($F(2,66)=8.47$, $p=.001$, *partial* $\eta^2=.204$).

This result suggests that there was a big change in GJ scores over time. The *Time x Method* interaction was significant ($F(2,66)=3.33, p=.042, \text{partial } \eta^2=.092$), which means that there was a moderate difference in change in scores over time for the two groups of participants. The analysis of between-subjects effects showed a non-statistically significant effect of *Method* ($p>.05$). This result suggests that there was no significant difference in grammaticality judgment scores between the two groups.

Table 48. Mixed between-within subjects ANOVA: Grammaticality judgment (Rus I)

	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>P</i>	η^2
<i>Time</i>	1585.12	2,66	8.47	.001	.204
<i>Time x Method</i>	622.98	2,66	3.33	.042	.092
<i>Method</i>	122.04	1,33	.19	.677	.006

In order to explore at what point in time the differences in GJ scores were statistically significant, pairwise comparisons of participants' scores adjusted for multiple comparisons (Bonferroni correction) in each group at the pre-test, immediate post-test and delayed post-test were conducted. There were statistically non-significant differences between pre-test, immediate post-test and delayed post-test grammaticality judgment scores in the TS group. The results suggest that the participants in this group did not significantly improve their use of case forms and showed similar results as they had before the treatment.

In the TB group, participants had higher scores at the immediate post-test than those at the pre-test, but the difference in scores did not reach a level of significance ($p>.05$). There was statistically significant difference between the immediate post-test and delayed post-test scores ($t(15)=3.91, p=.004, r=.71$), suggesting that there was a significant drop of scores at the delayed post-test. The results obtained at the delayed post-test were also worse than those at the pre-test, although difference between pre-test and delayed post-test scores was not significant ($p>.05$).

Statistical analyses of grammaticality judgment scores let us conclude that neither the TS nor the TB groups showed significant improvement at the two post-tests, which means that neither task-supported nor task-based treatment had any effect on the use of the target forms. The TB group performed slightly better than the TS group at the immediate post-test but then showed worse results than the TS group at the delayed post-test.

6.5 The effects of task-supported and task-based treatment on the use of verbs of motion

The purpose of this section is to answer RQ2 and find out which of the two approaches was more efficient for learning new lexical items – Russian verbs of motion (VM). In order to answer this question, we compared the total amount of verbs of motion produced by participants in the TS and TB groups and the amount of correctly used verbs of motion in written and oral performance. The results of three grammatical tests performed by two Rus II groups will also be analyzed in this section.

6.5.1 Total amount of verbs of motion in written and oral production

The results of descriptive statistics showed a visible dynamic in the use of verbs of motion by learners both in their oral and written production. Mean scores and standard deviations are presented in Table 49 and visually displayed in Figure 19 and Figure 20. The mean scores for the Pre-test were very low in both groups, which was an expected result since learners hadn't seen these forms in their Russian lessons before the treatment. In contrast with L2 English, learners of L2 Russian in a foreign language context are exposed to input in Russian almost exclusively in the classroom. The Immediate post-test showed the growth in number of target verbs produced by participants as a result of both task-supported and task-based treatment. This may mean that both types of treatment were beneficial for learning target forms. The mean scores at the delayed post-test were higher than the scores at the immediate post-test for both groups. However, learners in the TB group demonstrated slightly worse results at the oral delayed post-test as compared with the immediate post-test. This result suggests that the positive effect of treatment on the use of VM remained.

Table 49. Descriptive statistics: Total amount of VM in written and oral production (Rus II)

	Written task				Oral task 2			
	TS_Rus II (n=14)		TB_Rus II (n=17)		TS_Rus II (n=13)		TB_Rus II (n=18)	
	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median
Pre-test	1.54 (.88)	1.00	1.53 (1.46)	1.00	.15 (.55)	0.00	.61 (1.04)	0.00
Immediate post-test	4.38 (2.18)	4.00	5.53 (1.87)	6.00	4.23 (2.55)	5.00	9.17 (3.29)	9.50
Delayed post-test	5.00 (2.31)	5.00	6.12 (2.09)	6.00	5.92 (2.25)	5.00	8.28 (2.52)	9.00

Figure 19 and Figure 20 clearly illustrate that the TB group outperformed the TS group both at the immediate and delayed post-tests. In other words, learners in the TB group produced more VM after the treatment than learners in the TS group. The difference in results is especially visible if we compare the scores obtained for oral performance at the immediate post-test (in the TS group $M=4.23$, $SD=2.55$ and in the TB group $M=9.17$, $SD=3.29$).

Figure 19. Mean number of VM in written production (Rus II)

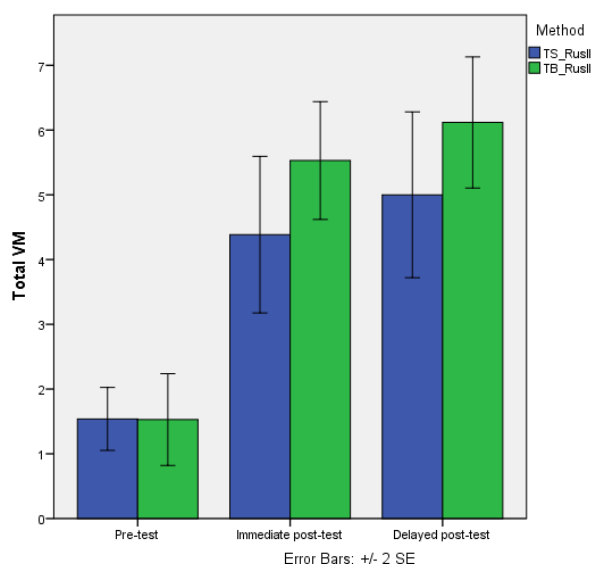
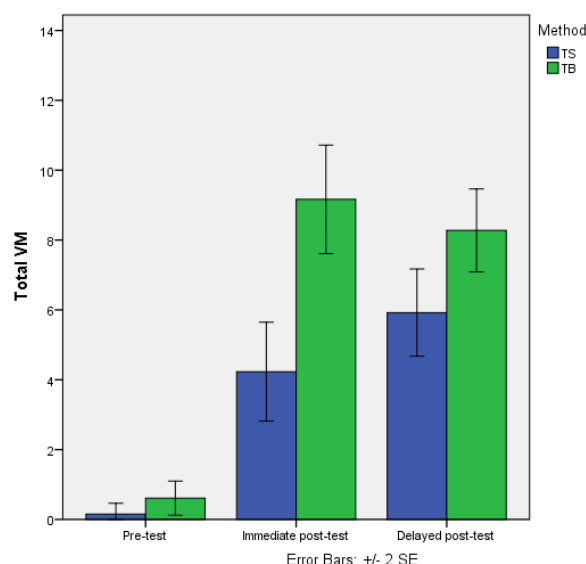


Figure 20. Mean number of VM in oral task 2 (Rus II)



Then statistical analyses were conducted in order to determine whether differences in scores within each group and between two groups reached the level of statistical significance. Preliminary analysis showed that scores on the total amount of VM in written production were normally distributed. Mauchly's sphericity and Levene's tests showed non-significant results ($p>.05$). This allowed us to use a mixed between-within subjects ANOVA for statistical analysis of written data. The analysis of oral production (oral task 2) showed that the scores on this measure did not follow normal distribution. For this reason, nonparametric tests (the Friedman test, Wilcoxon signed-rank test, and Mann-Whitney test) were used.

The results from a mixed between-within subjects ANOVA showed that there was a significant and large main effect of *Time* ($F(2,56)=55.79$, $p=.000$, $partial \eta^2=.666$). This result suggests there was a significant change in the total amount of produced verbs of motion over time. The *Time x Method* interaction was non-significant ($p>.05$), which means that this change over time did not significantly differ for the two groups of participants. The analysis of between-subjects effects showed a moderate effect size ($partial \eta^2=.076$) and a non-statistically significant

effect of *Method* ($p > .05$). This result suggests that, although there was no significant difference in the total amount of VM produced by participants in the TS and TB groups, there was a tendency of the TB group to produce more VM in their written performance. The results of the mixed between-within subjects ANOVA are summarized in Table 50.

Table 50. Mixed between-within subjects ANOVA: Total amount of VM in written production (Rus II)

	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
<i>Time</i>	277.77	2,56	55.79	.000	.666
<i>Time x Method</i>	6.39	2,56	1.28	.285	.044
<i>Method</i>	12.47	1,28	2.31	.140	.076

In order to explore the differences in the total amount of VM produced in the written performance, pairwise comparisons of the participants' scores in each group at the pre-test, immediate post-test and delayed post-test were conducted. We found out that there was a significant difference between pre-test and immediate post-test scores in the TS group ($t(12)=4.94$, $p=.001$, $r=.82$) and in the TB group ($t(16)=-7.47$, $p=.000$, $r=.88$). The results suggest that the participants in both groups produced significantly more verbs of motion at the immediate post-test than at the pre-test. There was a statistically non-significant difference between the immediate post-test and delayed post-test scores ($p > .05$) in both groups, which may mean that the positive effect of treatment remained. The difference between the pre-test and the delayed post-test was statistically significant for the TS group ($t(12)=-5.926$, $p=.000$, $r=.86$), as well as for the TB group ($t(16)=-8.91$, $p=.000$, $r=.91$). These results suggest that the participants in both groups produced significantly more VM at the delayed post-test than at the pre-test.

Since oral data did not follow normal distribution, nonparametric statistical tests were used. In particular, the Friedman test and Wilcoxon signed-ranks tests were conducted to compare within-group differences, and the Mann-Whitney test was performed to assess between-groups differences.

The results of the Friedman test suggest that there was a statistically significant change in the total amount of VM produced both within the TS group ($\chi^2(2)=20.94$, $p=.000$) and the TB group ($\chi^2(2)=29.65$, $p=.000$) over time.

Wilcoxon signed-ranks tests were conducted to compare the differences in the amount of VM produced by participants in each group at the pre-test, immediate post-test and delayed post-test. In the TS group, there was a significant difference between pre-test and immediate post-test scores ($Z=-2.95$, $p=.003$, $r=.82$), which means that learners in the TS group significantly increased the use of verbs of motion in their oral performance. There was a statistically

significant difference between the immediate post-test and delayed post-test scores ($Z=-2.62$, $p=.009$, $r=.72$). The difference between the pre-test and delayed post-test scores was also statistically significant ($Z=-3.19$, $p=.001$, $r=.89$). This result suggests that the participants in the TS group produced significantly more target forms at the delayed post-test than at the pre-test.

Wilcoxon signed-ranks tests for the TB group showed that there was a significant difference between pre-test and immediate post-test scores ($Z=-3.73$, $p=.000$, $r=.88$). The results suggest that the participants in the TB group produced significantly more verbs of motion at the immediate post-test than at the pre-test. There was a statistically non-significant difference between immediate and delayed post-test scores ($p>.05$), which may mean that the positive effect of task-based treatment remained. The difference between the pre-test and the delayed post-test was statistically significant ($Z=-3.74$, $p=.000$, $r=.88$). This result suggests that the participants in the TB group, as well as in the TS group, produced significantly more target forms at the delayed post-test than at the pre-test. The results of Wilcoxon signed-ranks tests for the two groups are summarized in Table 51.

Table 51. Wilcoxon signed-ranks tests: Total amount of VM in oral production (Rus II)

	TS_RusII			TB_RusII		
	Pre-test – Immediate post-test	Immediate post-test – Delayed post-test	Delayed post-test – Pre-test	Pre-test – Immediate post-test	Immediate post-test – Delayed post-test	Delayed post-test – Pre-test
Z	-2.95 ^c	-2.62 ^c	-3.19 ^c	-3.73 ^c	-1.81 ^d	-3.74 ^c
p	.003	.009	.001	.000	.070	.000

In order to assess between-groups differences, two types of tests were performed since the pre-test scores violated the assumption of normality of distribution (see Table 52). The results of a Mann-Whitney test showed that there were no statistically significant differences between the TS and TB groups at the pre-test ($p>.05$). Independent samples t-tests were used to assess the differences in immediate and delayed post-test scores between the two groups of participants. The difference in immediate post-test scores was statistically significant ($t(29)=-4.51$, $p=.000$, $r=.64$) suggesting that the TB group produced significantly more target forms than the TS group did. The analysis also yielded statistically significant differences in delayed post-test scores between the groups ($t(29)=-2.68$, $p=.012$, $r=.45$).

Table 52. Between-subjects effects: Total amount of VM in oral production (Rus II)

	Independent samples t-test			Mann-Whitney test		
	<i>t</i>	<i>df</i>	<i>p</i>	<i>U</i>	<i>Z</i>	<i>p</i>
Pre-test				93.50	-1.71	.088
Immediate post-test	-4.51	29	.000			
Delayed post-test	-2.68	29	.012			

Taken together, the results of statistical analyses of oral data suggest that both task-supported and task-based treatment were beneficial for learning verbs of motion with prefixes. However, participants in the TB group produced significantly more verbs of motion than the TS group did at the immediate and delayed post-test. This result may mean that the task-based treatment was more efficient for learning new lexical items than the task-supported treatment. The analysis of written data showed similar results. In their written performance, participants in both groups produced significantly more VM at the immediate post-test and delayed post-test than at the pre-test. Although there was no significant difference between the two groups, the TB group tended to produce more verbs of motion than the TS group.

6.5.2 Target-like use of verbs of motion in written and oral production

The next step was to compare the amount of accurately used verbs of motion by learners in both groups. As well as with other measures, descriptive statistics were first calculated. The results are presented in Table 53. The TS and TB groups demonstrated a similar dynamic as reported in the previous section. Mean and median scores at the pre-test were very low ($Mdn=1.00$ for written performance of the TS group, and $Mdn=.00$ for written production of the TB group and oral performance of both groups), which was an expected result since the target forms were new for learners. There was an improvement in scores at the immediate post-test, especially in scores obtained by the TB group for their oral production ($M=.11$, $SD=.32$ at the pre-test and $M=6.61$, $SD=3.07$ at the immediate post-test). These results suggest that both task-supported and task-based treatment had a positive effect on the accurate use of verbs of motion. Delayed post-test scores were slightly higher than the scores for the immediate post-test (except for oral performance of the TB group who demonstrated only slightly lower results), which means that the positive effect of treatment remained.

Table 53. Descriptive statistics: TLU of verbs of motion in written and oral production (Rus II)

	Written task				Oral task 2			
	TS_Rus II (n=14)		TB_Rus II (n=17)		TS_Rus II (n=13)		TB_Rus II (n=18)	
	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median
Pre-test	1.21 (1.05)	1.00	.71 (1.31)	.00	.08 (.28)	.00	.11 (.32)	.00
Immediate post-test	3.00 (2.07)	2.50	4.12 (2.09)	4.00	3.23 (2.71)	3.00	6.61 (3.07)	6.50
Delayed post-test	3.14 (2.28)	2.50	4.65 (2.18)	5.00	4.23 (2.39)	4.00	6.17 (2.06)	6.50

The results of descriptive statistics are displayed in Figure 21 (for written production) and Figure 22 (for oral production). While both groups showed improvement in the TLU of verbs of motion, the TB group achieved better results than the TS group both on written and oral post-tests. This allows us to conclude that the task-based treatment was more beneficial for learning verbs of motion than the task-supported treatment.

Figure 21. Mean scores on TLU of verbs of motion in written production (Rus II)

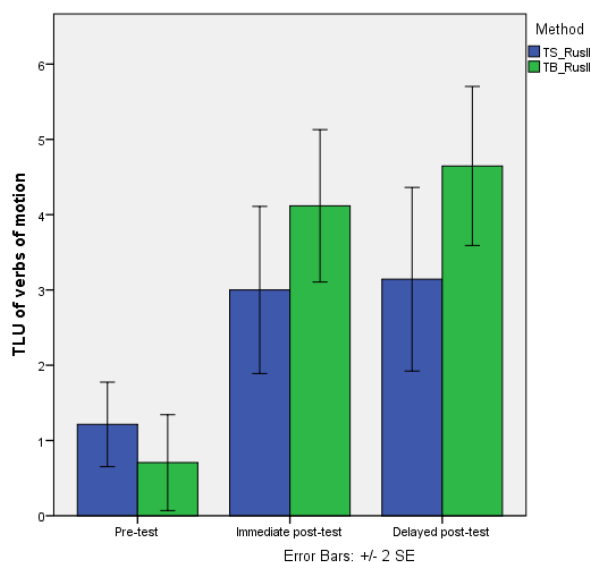
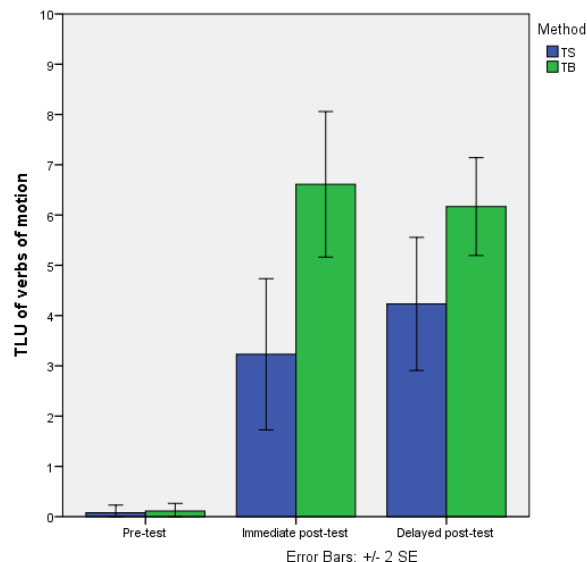


Figure 22. Mean scores on TLU of verbs of motion in oral task 2 (Rus II)



The next step was to choose adequate statistical tests for analysis of written and oral data. Preliminary analysis showed that scores on the TLU of verbs of motion both in written and oral production did not follow a normal distribution. For this reason, nonparametric tests were performed. A Friedman test and Wilcoxon signed-rank tests were conducted to analyze within-subjects differences, and Mann-Whitney tests and Independent samples t-tests were performed to analyze between-subject differences.

The results of the Friedman test suggest that there was a statistically significant change in the TLU of verbs of motion in written production both within the TS group ($\chi^2(2)=10.36$, $p=.006$) and the TB group ($\chi^2(2)=26.16$, $p=.000$) over time.

In order to find out at what time point these differences occurred, Wilcoxon signed-ranks tests were conducted. There was a significant difference between the pre-test and the immediate post-test scores in the TS group ($Z=-2.37$, $p=.018$, $r=.63$) and in the TB group ($Z=-3.53$, $p=.000$, $r=.86$). The results suggest that the participants in both groups correctly produced significantly more verbs of motion at the immediate post-test than at the pre-test. There was a statistically non-significant difference between immediate and delayed post-test scores ($p>.05$) both in the TS and TB groups, which means that the positive effect of treatment remained. The difference between the pre-test and the delayed post-test was statistically significant in the TS group ($Z=-2.62$, $p=.009$, $r=.70$) and in the TB group ($Z=-3.63$, $p=.000$, $r=.88$). This result suggests that the participants in both groups correctly produced significantly more verbs of motion at the delayed post-test than at the pre-test. The results of Wilcoxon signed-ranks tests for written data are summarized in Table 54.

Table 54. Wilcoxon signed-ranks tests: TLU of verbs of motion in written production (Rus II)

	TS_RusII			TB_RusII		
	Pre-test – Immediate post-test	Immediate post-test – Delayed post-test	Delayed post-test – Pre-test	Pre-test – Immediate post-test	Immediate post-test – Delayed post-test	Delayed post-test – Pre-test
Z	-2.37	-.040	-2.62	-3.53	-1.03	-3.63
p	.018	.968	.009	.000	.305	.000

Mann-Whitney tests to assess the differences in pre-test and delayed post-test scores, and independent samples t-test to assess the differences in immediate post-test scores between the TS and TB groups were performed (see Table 55). The results show that there were moderate differences between the TS and TB groups at the pre-test ($U=73.00$, $Z=-1.96$, $p=.05$, $r=.35$). Although there was a tendency of the TB group to outperform the TS group at the immediate post-test, the difference in scores was statistically non-significant ($p>.05$). The Mann-Whitney test yielded statistically significant moderate differences between the two groups in delayed post-test scores ($U=69.00$, $Z=-2.01$, $p=.044$, $r=.36$), suggesting that learners in the TB group correctly produced significantly more VM than learners in the TS group.

Table 55. Between-subjects effects: TLU of verbs of motion in written production (Rus II)

	Independent samples t-test			Mann-Whitney test		
	<i>t</i>	<i>df</i>	<i>p</i>	<i>U</i>	<i>Z</i>	<i>p</i>
Pre-test				73.00	-1.96	.05
Immediate post-test	-1.49	29	.063			
Delayed post-test				69.00	-2.01	.044

In the analysis of oral data Friedman tests were performed to assess whether there was a significant change in TLU of verbs of motion over time within the TS and TB groups. The results suggest that there was a significant change in scores within both the TS ($\chi^2(2)=19.86$, $p=.000$) and the TB group ($\chi^2(2)=28.80$, $p=.000$).

Wilcoxon signed-ranks tests were conducted to compare the amount of correctly produced verbs of motion. There was a significant difference between pre-test and immediate post-test scores in the TS group ($Z=-2.82$, $p=.005$, $r=.78$) and in the TB group ($Z=-3.73$, $p=.000$, $r=.88$). The results suggest that participants in both groups correctly produced significantly more verbs of motion at the immediate post-test than at the pre-test. There was a statistically non-significant difference between immediate and delayed post-test scores ($p>.05$), which may mean that the positive effect of both types of treatment remained. The difference between the pre-test and the delayed post-test was statistically significant both in the TS group ($Z=-3.08$, $p=.002$, $r=.85$) and in the TB group ($Z=-3.74$, $p=.000$, $r=.88$). These results suggest that participants in the two groups correctly produced significantly more target forms at the delayed post-test than at the pre-test. The results of Wilcoxon signed-ranks tests for oral data are summarized in Table 56.

Table 56. Wilcoxon signed-ranks tests: TLU of verbs of motion in oral production (Rus II)

	TS_RusII			TB_RusII		
	Pre-test – Immediate post-test	Immediate post-test – Delayed post-test	Delayed post-test – Pre-test	Pre-test – Immediate post-test	Immediate post-test – Delayed post-test	Delayed post-test – Pre-test
Z	-2.82	-1.61	-3.08	-3.73	-1.29	-3.74
p	.005	.968	.002	.000	.198	.000

Between-groups differences were analyzed by means of a Mann-Whitney test (for pre-test scores) and independent samples t-tests (to assess the differences in immediate and delayed post-test scores). The results are presented in Table 57. The Mann-Whitney test showed that there were no significant differences between the TS and TB groups at the pre-test ($U=121$, $Z=-.38$, $p=.707$). The difference in scores obtained at the immediate post-test was statistically

significant ($t(29)=-3.17$, $p=.004$, $r=.51$), with a large effect size, suggesting that the TB group accurately produced significantly more target forms than the TS group did. The analysis yielded moderate differences between the groups in delayed post-test scores ($t(29)=-2.41$, $p=.022$, $r=.41$), suggesting that learners correctly produced more verbs of motion after the task-based treatment than after the task-supported treatment.

Table 57. Between-subjects effects: TLU of verbs of motion in oral production (Rus II)

	Independent samples t-test			Mann-Whitney test		
	<i>t</i>	<i>df</i>	<i>p</i>	<i>U</i>	<i>Z</i>	<i>p</i>
Pre-test				121	-.38	.707
Immediate post-test	-3.17	29	.004			
Delayed post-test	-2.41	29	.022			

A summary of the results obtained for TLU of verbs of motion let us conclude that both task-supported and task-based treatment led to improvement in learners' use of VM in their oral and written production. There were significant differences between pre-test and immediate post-test scores and between pre-test and delayed post-test scores obtained by the two groups. The statistical analyses of between-subjects differences showed that participants in the TB group accurately produced significantly more verbs of motion than learners in the TS group did at the oral immediate and delayed post-test and at the written delayed post-test. These results suggest that the task-based treatment was more efficient for learning verbs of motion with prefixes than the task-supported treatment.

6.5.3 Fill in the blanks

The mean scores and standard deviations for the fill in the blanks test performed by Rus II groups are provided in Table 58, and are visually displayed in Figure 23. Both the TS and TB group showed constant improvement of their results from the pre-test to delayed post-test. The mean scores at the delayed post-test were higher than at the immediate post-test, which could be explained by the fact that this test was a part of the final exam and participants probably put more effort into taking the test.

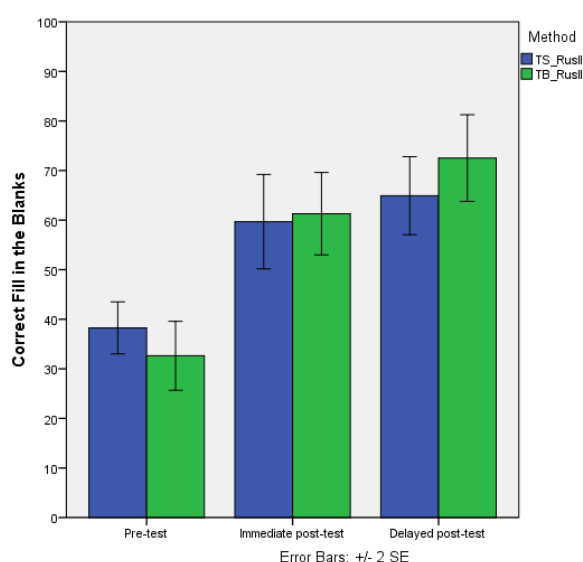
Table 58. Descriptive statistics: Fill in the blanks (Rus II)

Group	Pre-test	Immediate post-test	Delayed post-test
TS_RusII (n=14)	38.25 (9.83)	59.68 (17.82)	64.92 (14.74)
TB_RusII (n=16)	32.63 (15.14)	61.29 (18.12)	72.51 (19.09)

*Mean (SD)

Figure 23 illustrates the dynamics of the two groups over time. Although the TB group had lower scores than the TS group at the pre-test, both groups achieved similar results at the immediate post-test. The TB group outperformed the TS group at the delayed post-test.

Figure 23. Mean scores on fill in the blanks test (Rus II)



Preliminary analysis showed that FB scores were normally distributed, and Mauchly's sphericity and Levene's tests showed non-significant results ($p > .05$). This allowed us to use a mixed between-within subjects ANOVA for statistical analysis of obtained data.

The results from a mixed between-within subjects ANOVA (see Table 59) showed that there was a significant large main effect of *Time* ($F(2,62)=88.66, p=.000, partial \eta^2=.741$). This result suggests that there was a significant change in FB scores over time. The *Time x Method* interaction was also significant ($F(2,62)=3.23, p=.046, partial \eta^2=.094$), with a moderate effect size, which means that there were moderate differences in the way the FB scores changed in the TB and TS groups. The analysis of between-subjects effects showed a statistically non-significant effect of *Method* ($p > .05$). This result suggests that task-supported and task-based treatments did not significantly differ in their effect on learners' scores in a fill in the blanks test.

Table 59. Mixed between-within subjects ANOVA: Fill in the blanks (Rus II)

	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
<i>Time</i>	19367.91	2,62	88.66	.000	.741
<i>Time x Method</i>	706.04	2,62	3.23	.046	.094
<i>Method</i>	34.35	1,31	.06	.810	.002

In order to explore the differences in FB scores, pairwise comparisons adjusted for multiple comparisons (Bonferroni correction) of participants' scores in each group at the pre-test, immediate and delayed post-test were conducted. There was a statistically significant difference between pre-test and immediate post-test scores in the TS group ($t(13)=-5.75$, $p=.000$, $r=.85$), which means that participants in the TS group significantly improved their use of VM as a result of task-supported treatment. There was a statistically non-significant difference between immediate post-test and delayed post-test scores ($p>.05$), whereas the difference between pre-test and delayed post-test scores was statistically significant ($t(13)=-8.60$, $p<.001$, $r=.92$). These results suggest that the positive effect of the TS treatment remained.

The pairwise comparisons performed for the TB group yielded significant differences between the pre-test and immediate post-test ($t(18)=-7.78$, $p=.000$, $r=.88$), immediate post-test and delayed post-test ($t(18)=-3.22$, $p=.014$, $r=.60$), and pre-test and delayed post-test FB scores ($t(18)=-9.55$, $p=.000$, $r=.91$). The results suggest that participants in the TB group significantly improved their grammar skills from the pre-test to the delayed post-test. This may mean that there was a positive and long-lasting effect of task-based treatment on learners' knowledge of verbs of motion measured by means of fill in the blanks test.

Taken together, the results of statistical analysis of FB scores showed that participants both in the TS and TB groups significantly improved their use of verbs of motion from the pre-test to the delayed post-test. The results of the mixed between-within subjects ANOVA also indicated that the task-supported and task-based treatment did not significantly differ in their effect on learners' scores in the fill in the blanks test.

6.5.4 Multiple choice

Descriptive statistical tests were conducted in order to analyze MC scores. Mean scores and standard deviations which are reported in Table 60 indicate that learners in both groups improved their results in the immediate post-test. For the TB group, this increase in scores was larger than for the TS group. Learners in each group demonstrated similar scores at the

immediate and delayed post-tests, which may mean that the positive effect of the treatment remained.

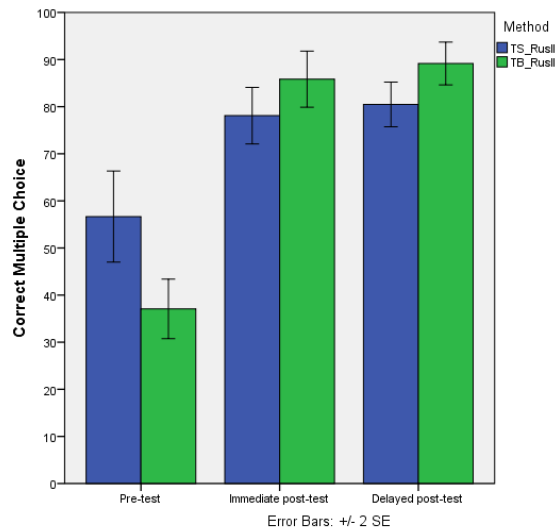
Table 60. Descriptive statistics: Multiple choice (Rus II)

Group	Pre-test	Immediate post-test	Delayed post-test
TS_RusII (n=14)	56.67 (18.07)*	78.09 (11.22)	80.48 (8.85)
TB_RusII (n=16)	37.08 (12.64)	85.83 (11.89)	89.17 (9.07)

*Mean (SD)

Figure 24 illustrates that although participants in the TB group had lower mean scores at the pre-test as compared with the TS group, they managed to outperform the TS group at the immediate and delayed post-test. This means that although both types of treatment helped to improve learners' knowledge of VM measured by means of the multiple choice test, the task-based treatment was more efficient than the task-supported treatment.

Figure 24. Mean scores on multiple choice test (Rus II)



Preliminary tests showed that both the assumption of sphericity ($p < .05$ in Mauchly's test for the TS group) and the assumption of homogeneity of variances ($p < .05$ in Levene's tests) were violated. Therefore, a mixed between-within subjects ANOVA could not be performed. A repeated measures ANOVA was performed in order to assess the differences within the TB group. The results of multivariate tests are reported for the TS group. The results showed that there was a significant main effect of *Time* for the TS group ($F(2,12)=10.64$, $p=.000$, *partial* $\eta^2=.639$) and the TB group ($F(2,30)=132.49$, $p=.000$, *partial* $\eta^2=.898$). The results suggest that there was a significant change in MC scores over time in both groups of participants (see Table 61).

Table 61. Within-subjects effects: Multiple choice (Rus II)

TS_RusII	Multivariate tests				
	<i>Wilks' Lambda</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
	.361	2,12	10.64	.000	.639
TB_RusII	Repeated measures ANOVA				
	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
	7367.93	2,30	132.49	.000	.898

The pairwise comparisons adjusted for multiple comparisons (Bonferroni correction) suggest that there was a large significant difference between pre-test and immediate post-test MC scores in the TS group ($t(13)=-3.42$, $p=.014$, $r=.68$) and in the TB group ($t(15)=-13.40$, $p=.000$, $r=.96$). This means that learners significantly improved their scores in MC test both after the task-supported and task-based treatment. There was a statistically non-significant difference between immediate and delayed post-test scores in both groups ($p>.05$), which may mean that the positive effect of treatment remained. The difference between pre-test and delayed post-test scores was statistically significant in the TS group ($t(13)=-4.52$, $p=.002$, $r=.78$) and in the TB group ($t(15)=-13.49$, $p=.000$, $r=.96$). These results suggest that there was a significant long-lasting effect of two types of treatment on learners' grammar scores.

The differences in MC scores between the TS group and the TB group were assessed by means of independent samples t-tests which report values corrected for the case when the homogeneity of variances is not assumed (see Table 62). The results showed that there were significant moderate differences between the groups in pre-test ($t(22,88)=3.39$, $p=.003$, $r=.47$) and delayed post-test ($t(28)=-2.65$, $p=.013$, $r=.45$) scores. These results suggest that although participants in the TB group showed statistically significantly worse results at the pre-test than learners in the TS group, they achieved a significant improvement at the delayed post-test. This proves the efficiency of the task-based approach as far as learning verbs of motion is concerned. There was no significant difference in immediate post-test scores between the two groups ($p>.05$).

Table 62. Between-subjects effects: Multiple choice (Rus II)

Time	Independent samples t-tests		
	<i>t</i>	<i>df</i>	<i>p</i>
Pre-test	3.39	22,88	.003
Immediate post-test	-1.82	28	.079
Delayed post-test	-2.65	28	.013

Summarizing the results, we can conclude that participants in both groups demonstrated a significant improvement in their MC scores from the pre-test to the delayed post-test. Between-

groups comparison showed that learners in the TB group outperformed the TS group at the immediate post-test and obtained significantly better scores at the delayed post-test as compared with the scores obtained by TS learners. This means that the task-based treatment was more efficient for improvement of learners' grammar skills than the task-supported treatment.

6.5.5 Grammaticality judgment

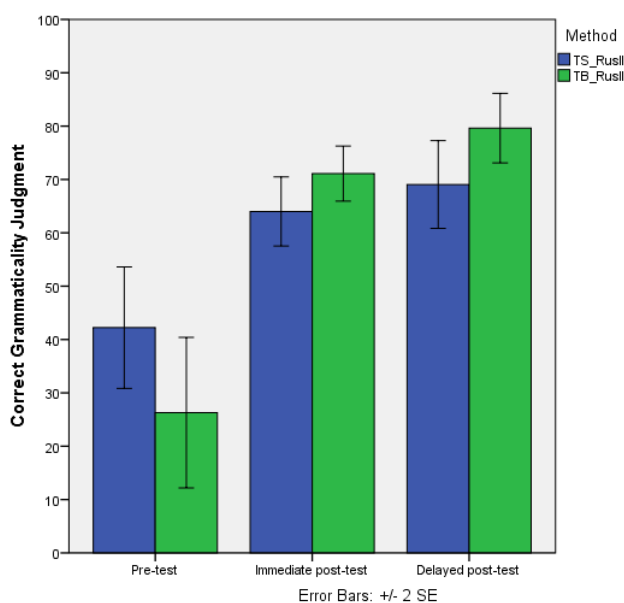
Analysis of GJ scores showed similar results as those obtained in the multiple choice test. There was an improvement of mean scores from the pre-test to the immediate post-test and from the immediate post-test to the delayed post-test in both groups (see Table 63). Although the TB group had lower pre-test scores, learners in this group managed to outperform the TS group at the two post-tests, as illustrated in Figure 25. This result suggests that the task-based treatment had a positive and long-lasting effect on learning verbs of motion.

Table 63. Descriptive statistics: Grammaticality judgment (Rus II)

Group	Pre-test	Immediate post-test	Delayed post-test
TS_RusII (n=14)	45.24 (19.42)*	63.81 (13.00)	69.05 (15.38)
TB_RusII (n=16)	29.58 (30.21)	71.25 (11.60)	81.67 (12.76)

*Mean (SD)

Figure 25. Mean scores on grammaticality judgment test (Rus II)



Preliminary tests showed that both the assumption of sphericity ($p < .05$ in Mauchly's test in the TB group) and the assumption of homogeneity of variances ($p < .05$ in Levene's tests) were violated. Therefore, a repeated measures ANOVA and multivariate tests were performed in order to assess the differences within each group of participants, and the differences between groups were assessed by means of independent samples t-tests.

There was a significant main effect of *Time* for the TS group ($F(2,26)=13.35$, $p=.000$, *partial* $\eta^2=.507$) and the TB group ($F(2,14)=32.43$, $p=.000$, *partial* $\eta^2=.822$). The results suggest that there was a significant change in GJ scores over time in both groups of participants (see Table 64).

Table 64. Within-subjects effects: Grammaticality judgment (Rus II)

TS_RusII	Repeated measures ANOVA				
	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	<i>η²</i>
	4383.07	2,26	13.35	.000	.507
TB_RusII	Multivariate tests				
	<i>Wilks' Lambda</i>	<i>df</i>	<i>F</i>	<i>p</i>	<i>η²</i>
	.18	2,14	32.43	.000	.822

In order to explore the differences in GJ scores at each time point, pairwise comparisons of participants' scores for each group were conducted. There was a significant difference between pre-test and immediate post-test scores in the TS group ($t(13)=-3.91$, $p=.005$, $r=.73$). The results suggest that there was a significant positive effect of the task-supported treatment on learners' knowledge of verbs of movement. There was a statistically non-significant difference between immediate and delayed post-test scores, which indicates that the positive effect of the treatment remained. The difference between pre-test and delayed post-test scores was statistically significant ($t(13)=-4.17$, $p=.003$, $r=.75$), which proves that learners in the TS group performed significantly better at the delayed post-test than at the pre-test.

For the TB group, the pairwise comparisons yielded significant differences between the pre-test and the immediate post-test ($t(15)=-5.47$, $p=.000$, $r=.82$), the immediate post-test and the delayed post-test ($t(15)=-3.42$, $p=.011$, $r=.66$), and the pre-test and the delayed post-test ($t(15)=-7.53$, $p=.000$, $r=.89$). These results suggest that participants in the TB group significantly improved their grammar scores from the pre-test to the delayed post-test.

In order to analyze the differences in GJ scores between the two groups, a series of independent samples t-tests were conducted. The results obtained showed that there were no significant differences between the two groups' scores at the pre-test and immediate post-test ($p < .05$). The analyses yielded significant differences between the groups at the delayed post-test

($t(28)=-2.42$, $p=.021$, $r=.41$). As suggested by the effect size $r=.41$, the differences were moderate. The results of between-subjects comparisons are presented in Table 65.

Table 65. Between-subjects effects: Grammaticality judgment (Rus II)

Time	Independent samples t-tests		
	<i>t</i>	<i>df</i>	<i>p</i>
Pre-test	1.71	25,86	.100
Immediate post-test	-1.66	28	.109
Delayed post-test	-2.42	28	.021

Taken together, the results obtained showed that learners both in the TS group and in the TB group performed significantly better at the delayed post-test than at the pre-test, which means that there was a positive and long-lasting effect of both types of treatment on the use of verbs of motion measured by means of the grammaticality judgment test. The TB group outperformed the TS group at the immediate post-test and performed significantly better at the delayed post-test than the TS group. This result suggests that the task-based treatment was more efficient for learning new target forms than the task-supported treatment.

6.6 The effects of task-supported and task-based treatment on learners' accuracy, complexity and fluency

After examining the effect of two types of treatment on the use of case forms and verbs of motion, we next examined learners' production by looking at changes in the domains of CAF and using the measures described in Section 5.9 of the previous chapter. We explored these changes first in learners' written production and then in their oral performance. The results are described in the following subsections. As in the previous sections, we first present descriptive statistics for all the measures, including the mean scores, medians (in cases where data did not follow normal distribution) and standard deviations at each of the three time points (pre-test, immediate post-test and delayed post-test). Then the results of parametric and nonparametric statistical tests are presented.

6.6.1 Written production

6.6.1.1 General accuracy

General accuracy was measured by calculating the percentage of errors in the text produced by a learner. This means that the lower the scores, the better the result, since it demonstrates that the learner made fewer errors. The results of descriptive statistics (mean scores and standard deviations) are presented in Table 66. These results show that learners in both groups improved their accuracy at the immediate post-test. The TB group obtained the same mean scores at the delayed post-test as at the immediate post-test, which means that the task-based treatment had a positive and long-lasting effect on learners' accuracy. Learners in the TS group decreased their level of accuracy at the delayed post-test.

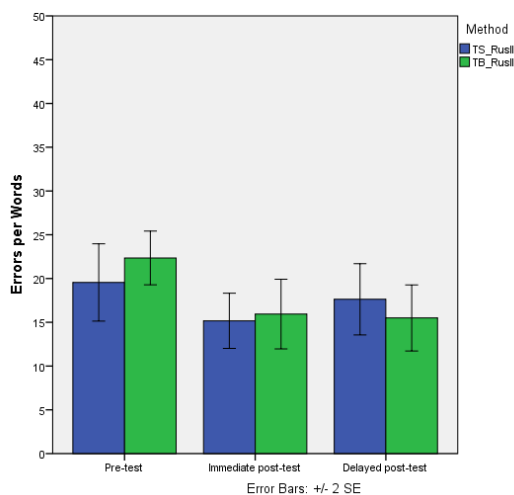
Table 66. Descriptive statistics: Errors per words (written task, Rus II)

Group	Pre-test	Immediate post-test	Delayed post-test
TS_RusII (<i>n</i> =14)	19.55 (8.26)*	15.16 (5.90)	17.63 (7.61)
TB_RusII (<i>n</i> =16)	22.35 (6.14)	15.94 (7.95)	15.50 (7.56)

*Mean (SD)

The results of measuring general accuracy are visually displayed in Figure 26. It illustrates that participants in the TB group had a higher percentage of lexical and grammatical errors than participants in the TS group at the pre-test. However, learners in the two groups obtained the same mean accuracy scores immediately after the treatment, and the TB group outperformed the TS group at the delayed post-test.

Figure 26. Mean scores on errors per words (written task, Rus II)



The next step was to conduct analyses in order to determine whether differences in scores within each group and between two groups reached the level of statistical significance. The results from a mixed between-within subjects ANOVA showed that there was a significant main effect of *Time* ($F(2,27)=11.18, p<.001, \text{partial } \eta^2=.453$). This result suggests that there was a significant change in scores over time for both groups. The *Time x Method* interaction was not significant ($p>.05$). The analysis of between-subjects effects showed a statistically non-significant effect of *Method* ($p>.05$). This result suggests that there were non-significant differences in accuracy scores between the two groups. The results from a mixed between-within subjects ANOVA are summarized in Table 67.

Table 67. Mixed between-within subjects ANOVA: Errors per words (written task, Rus II)

	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
<i>Time</i>	.55	2,27	11.18	.000	.453
<i>Time x Method</i>	.86	2,27	2.17	.133	.139
<i>Method</i>	5.17	1,28	.04	.839	.002

In order to explore the differences in accuracy scores, pairwise comparisons of the participants' scores in each group at the pre-test, immediate post-test and delayed post-test were performed by means of paired samples t-tests.

There was a significant difference between pre-test and immediate post-test accuracy scores both in the TS group ($t(13)=3.32, p=.017, r=.68$) and in the TB group ($t(15)=3.55, p=.009, r=.68$). The results suggest that the participants were significantly more accurate (i.e., they made fewer errors) at the immediate post-test than they were at the pre-test. There was a statistically non-significant difference between immediate and delayed post-test scores, which means that although learners made more errors at the delayed post-test, the positive effect of both task-supported and task-based treatment remained. Although participants in the TS group made fewer errors at the delayed post-test than at the pre-test, the difference in accuracy scores was not statistically significant, suggesting that learners did not significantly improve their general accuracy as a results of the task-supported treatment. In the TB group, the difference between pre-test and delayed post-test was statistically significant ($t(15)=3.47, p=.010, r=.66$). This result suggests that participants made significantly fewer errors at the delayed post-test than at the pre-test, suggesting a long-lasting positive effect of the task-based treatment.

Taken together, the results suggest that there was a significant positive effect of both task-supported and task-based treatment on learners' accuracy. However, this effect was not long-lasting for learners in the TS group. Their accuracy decreased at the delayed post-test to the

level they had at the pre-test. Participants in the TB group made significantly fewer errors at the two post-tests as compared with the pre-test, suggesting a long-lasting positive effect of task-based treatment on their general accuracy. There were no significant differences between the two groups, which means that TS and TB approaches did not differ as far as their effect on learners' accuracy is concerned.

6.6.1.2 Syntactic complexity

Syntactic complexity was measured by calculating the mean length of clause and the ratio of clauses per T-unit. Mean scores and standard deviations on both measures are presented in Table 68. The results showed that participants in the TS group produced shorter clauses in both post-tests as compared with the pre-test, whereas learners in the TB group produced longer clauses at the immediate post-test than at the pre-test and then decreased their syntactic complexity at the delayed post-test. As far as the ratio of clauses per T-unit is concerned, the TB group demonstrated a constant growth of syntactic complexity from the pre-test to the delayed post-test. The TS group used less subordinate clauses at the immediate post-test than at the pre-test and then the ratio of clauses per T-unit increased at the delayed post-test.

Table 68. Descriptive statistics: Syntactic complexity (written task, Rus II)

	Mean length of clause		Clauses per T-unit	
	TS_Rus II (n=13)	TB_Rus II (n=16)	TS_Rus II (n=14)	TB_Rus II (n=17)
Pre-test	6.20 (.74)*	6.34 (.95)	1.18 (.12)	1.19 (.17)
Immediate post-test	6.04 (.96)	6.43 (.78)	1.17 (.14)	1.23 (.20)
Delayed post-test	5.87 (.77)	6.17 (.86)	1.28 (.20)	1.30 (.17)

*Mean (SD)

Figure 27 and Figure 28 illustrate the changes in mean scores on the two measures of syntactic complexity. Figure 27 illustrates that mean length of clauses produced by learners in the TB group was higher than in the TS group. In Figure 28, we can observe that the difference between the two groups was minimal at the pre-test and the delayed post-test. However, at the immediate post-test the TB group outperformed the TS group on the ratio of clauses per T-unit.

Figure 27. Mean scores on mean length of clause (written task, Rus II)

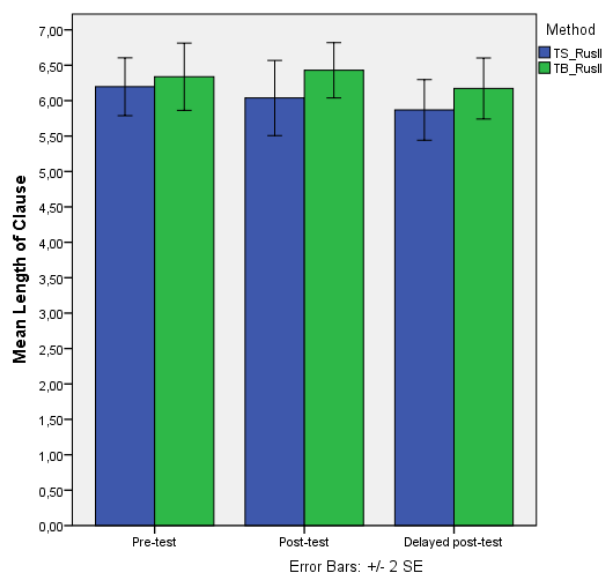
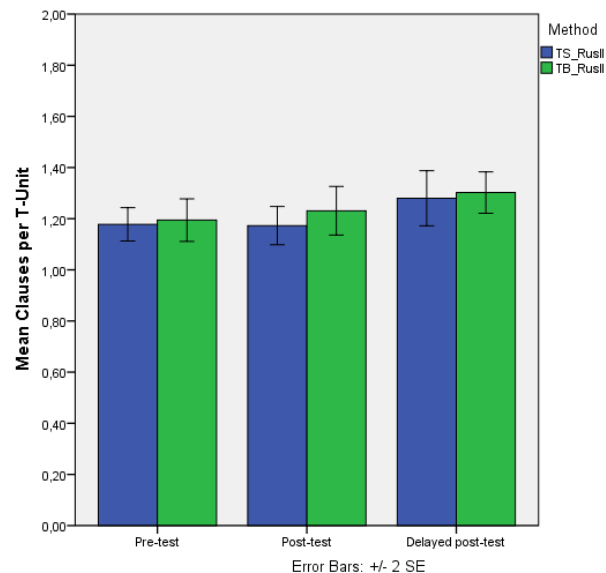


Figure 28. Mean scores on clauses per T-unit (written task, Rus II)



In order to find out whether the observed differences in scores reached the level of statistical significance, a mixed between-within subjects ANOVA was performed. The results (see Table 69) showed that the main effect of *Time*, as well as the *Time x Method* interaction were statistically non-significant ($p > .05$). This result suggests there was no significant change in mean length of clause over time in both groups. The analysis of between-subjects effects showed a statistically non-significant effect of *Method* ($p > .05$). This result suggests that there were no significant differences between task-based and task-supported treatments regarding their effects on learners' syntactic complexity.

Table 69. Mixed between-within subjects ANOVA: Mean length of clause (written task, Rus II)

	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
<i>Time</i>	.93	2,26	.95	.399	.068
<i>Time x Method</i>	.66	2,26	.51	.608	.038
<i>Method</i>	1.68	1,27	1.16	.290	.041

The analysis of scores on the ratio of clauses per T-unit showed that there was a significant main effect of *Time* ($F(2,58)=12.81, p=.000, partial \eta^2=.244$), which means that there was a significant change in a number of clauses per T-unit over time. The *Time x Method* interaction and the main effect of *Method* were not significant ($p > .05$). This result suggests that there was no significant difference in the scores between the two groups. The results of the mixed between-within subjects ANOVA are presented in Table 70.

Table 70. Mixed between-within subjects ANOVA: Clauses per T-unit (written task, Rus II)

	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
<i>Time</i>	.20	2,58	9.35	.000	.244
<i>Time x Method</i>	.02	2,58	.37	.692	.013
<i>Method</i>	.02	1,29	.37	.548	.013

In order to explore the differences in syntactic complexity scores over time, pairwise comparisons of the participants' scores in each group at the pre-test, immediate post-test and delayed post-test were conducted. The results of paired samples t-tests showed that there was no significant difference between pre-test and immediate post-test scores either in the TS group or the TB group. The results suggest that participants did not produce more clauses per T-unit (i.e., their written production was not more syntactically complex) at the immediate post-test than at the pre-test. There was a statistically significant difference between immediate post-test and delayed post-test scores ($t(13)=-2.89$, $p=.038$, $r=.62$) in the TS group, which means that participants produced significantly more clauses per T-unit at the delayed post-test than at the immediate post-test. In the TB group, participants also produced more clauses per T-unit at the delayed post-test than at the immediate post-test, but this difference was not statistically significant. The difference between pre-test and delayed post-test scores was statistically significant both in the TS group ($t(13)=-3.27$, $p=.018$, $r=.67$) and in the TB group ($t(16)=-2.93$, $p=.029$, $r=.59$). These results suggest that participants in both groups demonstrated significantly higher syntactic complexity at the delayed post-test than at the pre-test.

Taken together, the results indicate that, on the one hand, participants in both groups produced shorter clauses at the delayed post-test than at the pre-test, but there was no significant change in mean length of clauses over time in any of the groups. On the other hand, participants in both groups produced significantly more clauses per T-unit at the delayed post-test than at the pre-test, which means that their syntactic complexity measured via subordination significantly increased after the treatment. There were no significant differences between the task-based treatment and the task-supported treatment regarding their effect on learners' syntactic complexity.

6.6.1.3 Lexical complexity

Lexical complexity of learners' production was measured by means of Guiraud's Index of lexical richness. The results of descriptive statistics presented in Table 71 suggest that both

groups improved their lexical complexity (e.g., used a greater variety of vocabulary) from the pre-test to the delayed post-test, which means that both types of treatment were beneficial for vocabulary learning.

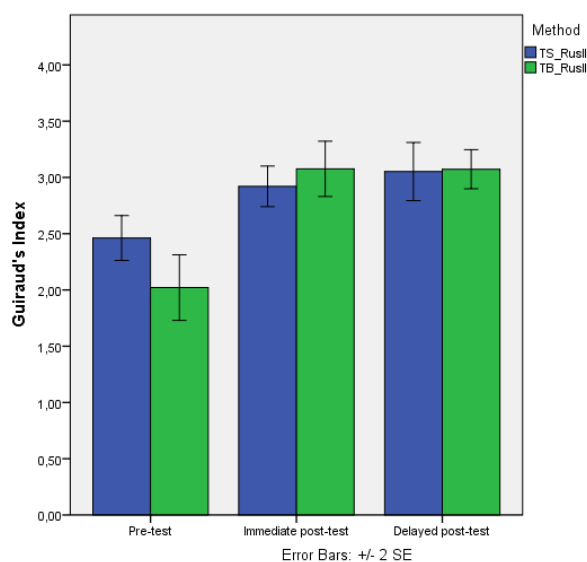
Table 71. Descriptive statistics: Guiraud’s Index (written task, Rus II)

Group	Pre-test	Immediate post-test	Delayed post-test
TS_RusII (n=14)	2.46 (.37)*	2.92 (.34)	3.05 (.48)
TB_RusII (n=17)	2.03 (.60)	3.08 (.51)	3.07 (.36)

*Mean (SD)

As illustrated in Figure 29, the TB group demonstrated a lower lexical complexity than the TS group at the pre-test. However, learners managed to improve their results after the task-based treatment and outperformed participants in the TS group at the immediate post-test. At the delayed post-test the two groups obtained similar mean scores.

Figure 29. Mean scores on Guiraud’s Index (written task, Rus II)



Preliminary analysis showed that data were normally distributed. Mauchly’s sphericity and Levene’s tests showed non-significant result ($p>.05$), which allowed us to perform a mixed between-within subjects ANOVA. The results are summarized in Table 72.

Table 72. Mixed between-within subjects ANOVA: Guiraud’s Index (written task, Rus II)

	Sum of Squares	df	F	p	η^2
Time	12.81	2,58	30.03	.000	.509
Time x Method	1.50	2,58	3.52	.036	.108
Method	.18	1,29	.87	.360	.029

The results from the mixed between-within subjects ANOVA showed that there was a significant main effect of *Time* ($F(2,58)=12.81, p=.000, \text{partial } \eta^2=.509$). This result suggests that there was a significant change in lexical complexity scores over time. The *Time x Method* interaction was significant ($F(2,58)=1.50, p=.036, \text{partial } \eta^2=.108$), which means that the change in scores over time significantly differed for the two groups of participants.

Paired samples t-tests showed that there was a significant difference between pre-test and immediate post-test scores both in the TS group ($t(13)=-4.24, p=.003, r=.76$) and in the TB group ($t(16)=-5.20, p=.000, r=.79$). The results suggest that participants in the two groups demonstrated higher lexical richness immediately after the treatment than they had at the pre-test. There was a statistically non-significant difference between immediate and delayed post-test scores, which may mean that the positive effect of both types of treatment remained. The difference between pre-test and delayed post-test scores was statistically significant in the TS group ($t(13)=-4.01, p=.004, r=.74$) and in the TB group ($t(16)=-5.76, p=.000, r=.82$). This result suggests that participants in both groups demonstrated greater lexical complexity at the delayed post-test than at the pre-test.

The analysis of between-subjects effects showed a non-statistically significant effect of *Method* ($p>.05$). This result suggests that there was no significant difference in lexical complexity scores between the two groups. The follow-up independent samples t-tests showed that there were statistically significant differences between the two groups at the pre-test ($t(29)=2.38, p=.024, r=.40$). The effect size $r=.40$ suggests that the differences between the groups at the pre-test were moderate, which allowed us to perform a reliable analyses of the differences between the TS and TB groups at the post-tests. The groups did not significantly differ at the immediate and delayed post-tests ($p>.05$), suggesting that the task-supported and the task-based treatment did not differ in their effect on lexical complexity.

Taken together, the results obtained showed that participants in both groups demonstrated significantly higher lexical richness at the two post-tests than they did at the pre-test. This result suggests that both the task-supported and the task-based treatment had a positive and long-lasting effect on learners' lexical complexity measured by means of Guiraud's Index. There were no significant differences between two groups. However, mean scores indicated that learners in the TB group increased their lexical complexity to a greater extent than learners in the TS group.

6.6.2 Oral production

Participants in both groups (Rus II) performed two different oral tasks, “Describe a room” (further oral task 1) and the map task (oral task 2) (see Appendix C). CAF scores obtained in these two tasks were statistically analyzed. The results obtained for each task were quite different. For this reason, we report them in a comparative way (i.e., comparing the results of oral task 1 and oral task 2 for each of the chosen measures).

6.6.2.1 General accuracy

The results of descriptive statistics (mean scores and standard deviations) for general accuracy of oral data are presented in Table 73. These results show that while performing oral task 1 learners in both groups made more errors at the immediate post-test and then improved their accuracy at the delayed post-test. The TS group outperformed the TB group on the three tests.

The results obtained from the performance of oral task 2 were different. Both groups improved their accuracy equally (i.e. made less grammatical and lexical errors) at the immediate post-test. However, at the delayed post-test participants in the TB group achieved better results and outperformed learners in the TS group whose level of accuracy slightly decreased. These differences between the two groups’ mean accuracy scores are visually displayed in Figure 30 and Figure 31.

Table 73. Descriptive statistics: Errors per words (oral production)

	Task 1		Task 2	
	TS_RusII (n=13)	TB_RusII (n=18)	TS_RusII (n=13)	TB_RusII (n=16)
Pre-test	21.95 (9.83)*	24.31 (5.28)	28.06 (8.24)	28.29 (9.22)
Immediate post-test	23.57 (8.83)	25.77 (11.06)	17.43 (5.05)	17.95 (8.70)
Delayed post-test	20.89 (8.72)	21.70 (9.76)	18.13 (7.74)	13.98 (6.55)

*Mean (SD)

Figure 30. Mean scores on errors per words
(oral task 1)

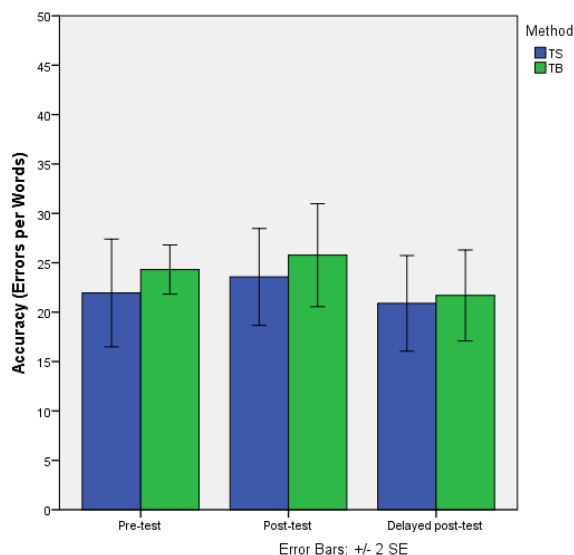
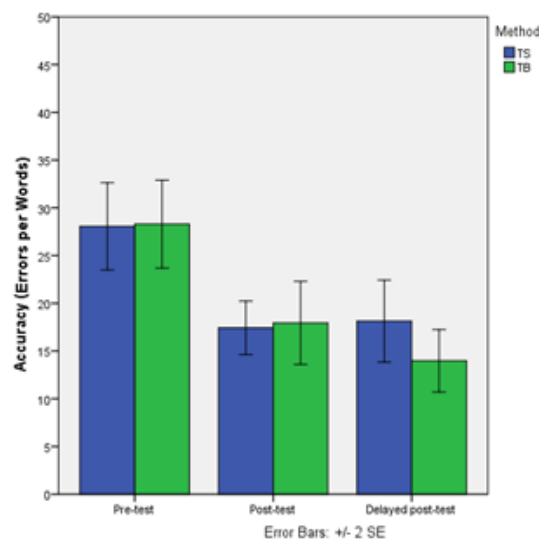


Figure 31. Mean scores on errors per words
(oral task 2)



For oral task 1, preliminary analyses showed that in the TS group Mauchly's sphericity tests yielded significant results ($p < .05$), whereas in the TB group these results were non-significant ($p > .05$). Therefore, for the TS group results from multivariate tests are reported. Assumption of homogeneity of variances has been violated for pre-test scores ($p < .05$ in Levene's tests). For this reason, the differences between groups were assessed by means of independent samples t-tests.

The results from the repeated measures ANOVA and multivariate tests showed that the main effect of *Time* for the TS group and for the TB group was statistically non-significant ($p > .05$), although the effect size was moderate in both groups (*partial* $\eta^2 = .134$ in the TS group and *partial* $\eta^2 = .109$ in the TB group). These results which are presented in Table 74 suggest that although there was a tendency of improvement of learners' general accuracy at the delayed post-test, there was no significant change in scores across the three periods of time in both groups.

Table 74. Within-subjects effects: Errors per words (oral task 1)

TS_RusII	Multivariate tests				
	<i>Wilks' Lambda</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
	.87	2,11	.85	.453	.134
TB_RusII	Repeated measures ANOVA				
	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
	153.30	2,34	2.09	.139	.109

The independent samples t-tests conducted to assess the differences between the two groups of participants showed that there were no significant differences ($p > .05$) between the two

groups' pre-test, immediate post-test and delayed post-test scores, suggesting that there were no differences between the two treatments as far as general accuracy is concerned. The results are reported in Table 75.

Table 75. Independent samples t-tests: Errors per words (oral task 1)

Time	Independent samples t-tests		
	<i>t</i>	<i>df</i>	<i>p</i>
Pre-test	-.79	16,99	.441
Immediate post-test	-.59	29	.557
Delayed post-test	-.24	29	.815

For oral task 2, Mauchly's sphericity tests showed nonsignificant results ($p > .05$). Assumption of homogeneity of variances for immediate post-test scores has been violated ($p < .05$ in Levene's tests). For this reason, the differences within each group were assessed by means of a repeated measures ANOVA, and the differences between the two groups were assessed by means of independent samples t-tests.

The results from the repeated measures ANOVA (see Table 76) showed that there was a significant main effect of *Time* for the TS group ($F(2,24)=16.21$, $p=.000$, *partial* $\eta^2=.575$) and for the TB group ($F(2,30)=19.93$, $p=.000$, *partial* $\eta^2=.571$). This result suggests that there was a significant change in scores over time for both groups.

Table 76. Repeated measures ANOVA: Errors per words (oral task 2)

Within-subjects effects	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
TS_RusII	918.09	2,24	16.21	.000	.575
TB_RusII	1747.58	2,30	19.93	.000	.571

In order to explore the differences in accuracy scores, pairwise comparisons of the participants' scores in each group were conducted by means of paired samples t-tests. There was a significant difference between pre-test and immediate post-test accuracy scores both in the TS group ($t(12)=4.36$, $p=.003$, $r=.78$) and in the TB group ($t(15)=3.69$, $p=.007$, $r=.68$). The results suggest that participants were significantly more accurate (i.e. made fewer errors) at the immediate post-test than they were at the pre-test. There was a statistically non-significant difference between immediate and delayed post-test scores, which means that the positive effect of both types of treatment remained. The difference between pre-test and delayed post-test scores was statistically significant in the TS group ($t(12)=4.84$, $p=.001$, $r=.81$) and in the TB group ($t(15)=6.31$, $p=.000$, $r=.85$). This result suggests that participants in the two groups made

significantly fewer errors at the delayed post-test than at the pre-test, suggesting a long-lasting positive effect of the task-supported and the task-based treatment.

The results from a series of independent samples t-tests showed that, as in the case of oral task 1, there were no statistically significant differences between the two groups' accuracy scores obtained for oral task 2 ($p > .05$) as reported in Table 77.

Table 77. Independent samples t-tests: Errors per words (oral task 2)

Time	Independent samples t-tests		
	<i>t</i>	<i>df</i>	<i>p</i>
Pre-test	-.07	27	.943
Immediate post-test	-.20	24,72	.843
Delayed post-test	1.57	27	.129

Summarizing the results of statistical analysis of accuracy scores, we found different dynamics in the performance of oral task 1 and oral task 2. In oral task 1, there was no significant change in scores over time in the two groups, although there was a tendency towards improvement of learners' general accuracy at the delayed post-test as compared with the pre-test. In performance of oral task 2, participants in both groups made significantly fewer errors at the two post-tests than at the pre-test, suggesting a long-lasting positive effect of the task-supported and the task-based treatment. Between-groups comparisons showed that there were no significant differences between the two types of treatment regarding their effect on general accuracy of learners' oral performance.

6.6.2.2 Syntactic complexity

Syntactic complexity of oral production was measured by calculating mean length of clause. We also calculated the ratio of clauses per T-unit, but for oral task 1 there were no cases of subordination in learners' performance, and for oral task 2, we found very few cases of subordination. This result was expected, since subordination is less common for oral production than written production, especially at this level of command of L2 Russian. In this section, only the results of statistical analyses of scores for mean length of clause are included.

Descriptive statistics showed that while performing oral task 1, at the immediate post-test participants in the TS group produced clauses of the same length as at the pre-test. At the delayed post-test their syntactic complexity decreased and they produced shorter clauses than those produced at the immediate post-test and the pre-test. Learners in the TB group produced

the longest clauses at the pre-test and then decreased their syntactic complexity at the two post-tests. For oral task 2, different results were obtained. Learners in the TS group increased the mean length of clauses produced from the pre-test to the delayed post-test. The TB group showed higher scores at the immediate post-test than at the pre-test and then their syntactic complexity decreased at the delayed post-test. The results of descriptive statistics are presented in Table 78.

Table 78. Descriptive statistics: Mean length of clause (oral production)

	Task 1		Task 2			
	TS_RusII (n=13)	TB_RusII (n=17)	TS_RusII (n=13)		TB_RusII (n=16)	
	Mean (SD)	Mean (SD)	Mean (SD)	Median	Mean (SD)	Median
Pre-test	5.51 (.61)	5.51 (.79)	5.86 (.79)	5.65	6.37 (.99)	6.16
Immediate post-test	5.52 (.58)	5.41 (.53)	6.14 (.95)	6.30	7.30 (2.08)	6.82
Delayed post-test	5.27 (.25)	5.45 (.44)	6.25 (1.13)	5.90	6.72 (.94)	6.57

Figure 32 and Figure 33 illustrate the dynamics of scores on syntactic complexity in the two groups. In oral task 1, the TS group outperformed the TB group at the immediate post-test. However, at the delayed post-test learners in the TS group produced shorter clauses than learners in the TB group. In oral task 2, the TB group outperformed the TS group (i.e. produced longer clauses) on the three tests.

Figure 32. Mean scores on mean length of clause (oral task 1)

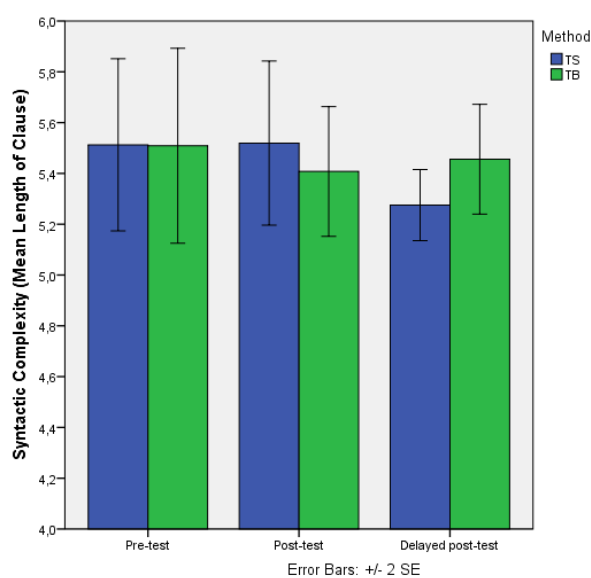
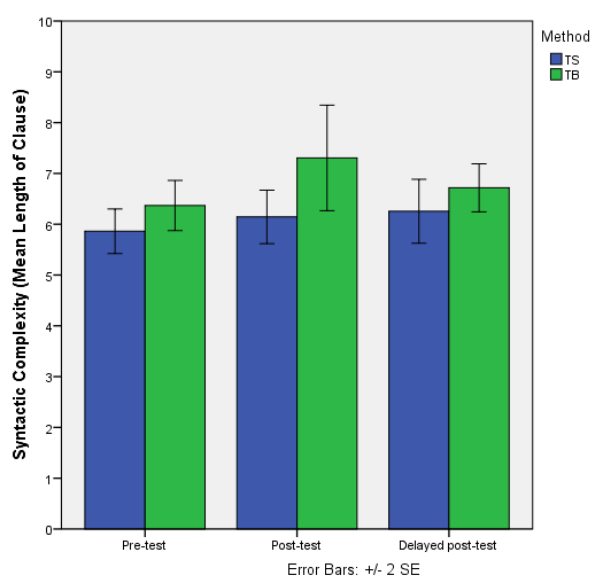


Figure 33. Mean scores on mean length of clause (oral task 2)



For oral task 1, preliminary analyses showed that Mauchly's sphericity tests produced non-significant results ($p > .05$). However, assumption of homogeneity of variances has been violated for delayed post-test scores ($p < .05$ in Levene's tests). For this reason, within-subjects differences were assessed by means of a repeated measures ANOVA and the differences between groups were assessed by means of independent samples t-tests.

The results from repeated measures ANOVA (see Table 79) showed that the main effect of *Time* both for the TS group and for the TB group was not significant ($p > .05$). These results suggest that there was no significant change in syntactic complexity scores across the three periods of time in both groups.

Table 79. Repeated measures ANOVA: Mean length of clause (oral task 1)

Within-subjects effects	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
TS_RusII	.50	2,24	1.34	.281	.100
TB_RusII	.09	2,32	.16	.852	.010

The independent samples t-tests conducted to assess the differences between the two groups of participants showed that there were no significant differences between the two groups' pre-, post- and delayed post-test scores ($p > .05$) as reported in Table 80.

Table 80. Independent samples t-tests: Mean length of clause (oral task 1)

Time	Independent samples t-tests		
	<i>t</i>	<i>df</i>	<i>p</i>
Pre-test	.01	28	.989
Immediate post-test	.55	28	.587
Delayed post-test	-1.40	26,10	.173

The scores obtained in oral task 2 by participants in the TB group did not follow normal distribution. For this reason, a Friedman test was used to assess within-group differences for the TB group. Data obtained from the TS group were normally distributed, and a repeated measures ANOVA was conducted for statistical analysis of these data.

The results from the repeated measures ANOVA showed that the main effect of *Time* was non-significant ($F(2,24)=.90$, $p > .05$, *partial* $\eta^2=.070$). The Friedman test showed there was a non-significant change in scores within the TB group ($\chi^2(2)=1.62$, $p > .05$). Taken together, the results suggest that there was no significant change in syntactic complexity scores over time in the two groups (see Table 81).

Table 81. Within-subjects effects: Mean length of clause (oral task 2)

Group	Repeated measures ANOVA					Friedman test		
	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2	<i>Chi-Square</i>	<i>df</i>	<i>p</i>
TS_RusII	1.07	2,24	.90	.419	.070			
TB_RusII						1.62	2	.444

The differences in pre-test and delayed post-test scores between the two groups were assessed by means of independent samples t-tests. A Mann-Whitney test was performed to assess the differences in immediate post-test scores. The results showed that there were no statistically significant differences between the TS and TB groups (see Table 82). The results obtained suggest that the two treatments did not differ in terms of the participants' improvement in syntactic complexity.

Table 82. Between-subjects effects: Mean length of clause (oral task 2)

Time	Independent samples t-test			Mann-Whitney test		
	<i>t</i>	<i>df</i>	<i>p</i>	<i>U</i>	<i>Z</i>	<i>p</i>
Pre-test	-1.50	27	.146			
Immediate post-test				70	-1.49	.136
Delayed post-test	-1.20	27	.242			

Summarizing the results, we can conclude that there were no significant changes in mean length of clauses produced by the TS group or the TB group before and after the treatment. While performing oral task 2, learners in the TB group produced longer clauses than learners in the TS group at the three tests, although between-subjects differences were not significant. The results suggest that both task-supported and task-based treatment did not result in advancement of participants' syntactic complexity.

6.6.2.3 Lexical complexity

Lexical complexity was measured by means of Guiraud's Index of lexical richness. Descriptive statistics showed that dynamics of scores on this measure obtained by participants in two groups differed for oral task 1 and oral task 2. In both tasks, the two groups improved their lexical complexity (i.e. used a greater variety of vocabulary) from the pre-test to the delayed post-test. While performing oral task 1, learners in the TS group outperformed the TB group at the immediate and delayed post-test. In oral task 2, learners in the TB group showed lower

lexical complexity than learners in the TS group at the pre-test. Despite this fact, the TB group outperformed the TS group on lexical complexity at the two post-tests (see Table 83).

Table 83. Descriptive statistics: Guiraud’s Index (oral production)

	Task 1		Task 2	
	TS_RusII (n=13)	TB_RusII (n=18)	TS_RusII (n=13)	TB_RusII (n=18)
Pre-test	1.51 (.44)*	1.44 (.37)	1.76 (.46)	1.36 (.48)
Immediate post-test	1.89 (.23)	1.68 (.39)	2.32 (.47)	2.68 (.41)
Delayed post-test	1.94 (.27)	1.82 (.37)	2.60 (.48)	2.79 (.39)

*Mean (SD)

Dynamics of mean scores on lexical complexity obtained by participants in the TS and the TB groups is visually displayed in Figure 34 and Figure 35.

Figure 34. Mean scores on Guiraud’s Index (oral task 1)

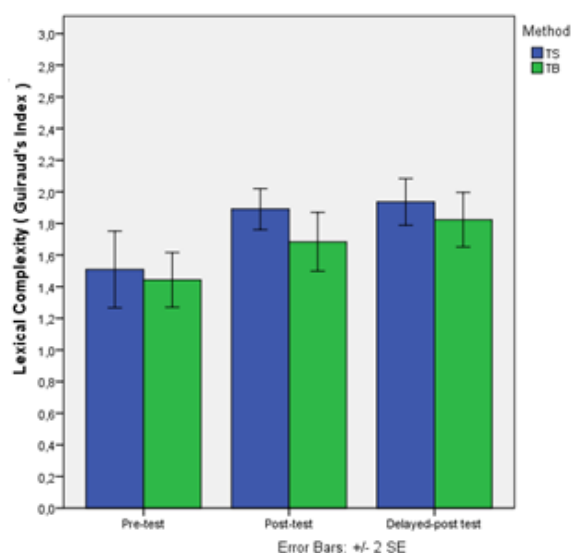
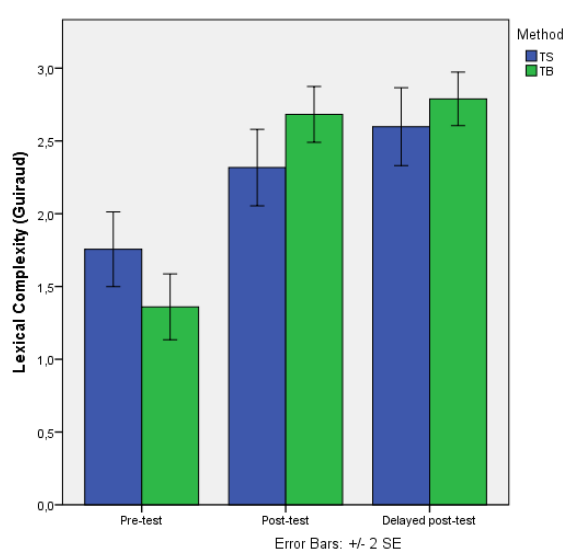


Figure 35. Mean scores on Guiraud’s Index (oral task 2)



First, lexical complexity scores obtained for performance of oral task 1 were analyzed. The results from a mixed between-within subjects ANOVA showed that there was a significant (moderate) effect of *Time* ($F(2,58)=14.92, p=.000, partial \eta^2=.340$). This result suggests that there was a significant change in scores over time. The *Time x Method* interaction was not significant ($p>.05$), which means that the change in scores over time did not significantly differ for the two groups of participants. The analysis of between-subjects effects showed a statistically non-significant effect of *Method* ($p>.05$). This result suggests that there were no significant differences between the TB and TS approaches as far as learners’ lexical complexity is concerned. The results are summarized in Table 84.

Table 84. Mixed between-within subjects ANOVA: Guiraud's Index (oral task 1)

	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
<i>Time</i>	2.71	2,58	14.92	.000	.340
<i>Time x Method</i>	.08	2,58	.43	.654	.015
<i>Method</i>	.37	1,29	1.88	.180	.061

In order to explore the differences in lexical complexity scores, pairwise comparisons of the participants' scores in each group at the pre-test, immediate post-test and delayed post-test were conducted by means of paired samples t-tests.

There was a significant difference between pre-test and immediate post-test lexical complexity scores both in the TS group ($t(12)=-3.96$, $p=.006$, $r=.75$) and in the TB group ($t(17)=-3.34$, $p=.012$, $r=.63$). The results suggest that participants in both groups demonstrated greater lexical complexity at the immediate post-test than that at the pre-test. There was a statistically non-significant difference between immediate post-test and delayed post-test scores, which means the positive effect of the task-supported and the task-based treatment remained. The difference between pre-test and delayed post-test scores was statistically significant both in the TS group ($t(12)=-2.75$, $p=.049$, $r=.62$) and in the TB group ($t(17)=-3.76$, $p=.005$, $r=.67$). The effect size was large for both groups. These results may mean that there was a positive effect of the two types of treatment on learners' lexical complexity.

For statistical analysis of scores obtained for performance of oral task 2, a mixed between-within subjects ANOVA was conducted. The results showed that there was a significant main effect of *Time* ($F(2,58)=11.14$, $p=.000$, *partial* $\eta^2=.663$). This result suggests that there was a significant change in lexical complexity scores over time. The *Time x Method* interaction was also significant ($F(2,58)=1.20$, $p=.004$, *partial* $\eta^2=.175$), which means that the change in lexical complexity scores over time differed for the two groups of participants. The results of the mixed between-within subjects ANOVA are presented in Table 85.

Table 85. Mixed Between-Within Subjects ANOVA: Guiraud's Index (oral task 2)

	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
<i>Time</i>	22.28	2,58	11.14	.000	.663
<i>Time x Method</i>	2.40	2,58	1.20	.004	.175
<i>Method</i>	.06	1,29	.31	.580	.011

In order to explore the differences in lexical complexity scores within each group of participants, pairwise comparisons of their scores were conducted by means of paired samples t-tests. There was a significant difference between pre-test and immediate post-test scores both in the TS group ($t(12)=-2.97$, $p=.035$, $r=.65$) and in the TB group ($t(17)=-8.35$, $p=.000$, $r=.89$). The

results suggest that participants in the two groups demonstrated higher lexical richness at the immediate post-test as compared with the pre-test. There was a statistically non-significant difference between immediate and delayed post-test scores ($p>.05$), which may mean that the positive effect of the treatment remained. The difference between the pre-test and the delayed post-test was statistically significant in the TS group ($t(12)=-5.30, p=.001, r=.84$) and in the TB group ($t(17)=-8.81, p=.000, r=.90$). This result suggests that both task-supported and task-based treatment had a positive and long-lasting effect on learners' lexical complexity measured by means of Guiraud's Index.

The analysis of between-subjects effects showed a statistically non-significant effect of *Method* ($p>.05$). However, the follow-up independent samples t-tests showed that there were statistically significant differences between the two groups at the pre-test ($t(29)=2.30, p=.029, r=.39$). The effect size $r=.39$ suggests that this difference was moderate, which may allow us to perform a reliable analysis of the differences between the TS and TB groups for post-tests. The groups significantly differed at the immediate post-test ($t(29)=-2.30, p=.029, r=.39$) suggesting that the TB group had significantly higher lexical complexity scores than the TS group. The differences in the delayed post-test scores were non-significant ($p>.05$), as reported in Table 86.

Table 86. Independent samples t-tests: Guiraud's Index (oral task 2)

Time	Independent samples t-tests		
	<i>t</i>	<i>df</i>	<i>p</i>
Pre-test	2.30	29	.029
Immediate post-test	-2.30	29	.029
Delayed post-test	-1.22	29	.233

Taken together, the results indicate that participants in both groups had significantly higher lexical complexity at the immediate post-test than that at the pre-test, which means that both task-supported and task-based treatments were beneficial for improving learners' lexical richness. In oral task 1, both groups demonstrated greater lexical complexity at the delayed post-test than at the pre-test. The difference between pre-test and delayed post-test scores did not reach the significance level in the TS group and was statistically significant in the TB group. The TS group outperformed the TB group on lexical complexity at the two post-tests, although this difference did not reach statistical significance. Analysis of scores for oral task 2 showed that both groups statistically significantly improved their scores on lexical complexity from the pre-test to the delayed post-test. Participants in the TB group performed significantly worse than learners in the TS group at the pre-test but managed to improve their results after the task-based treatment and significantly outperformed the TS group at the immediate post-test.

6.6.2.4 Fluency

Oral fluency was measured by calculating speech rate, that is, the number of syllables per minute. Analysis of mean scores which are presented in Table 87 showed that both groups' fluency increased from the pre-test to the delayed post-test, which means that the treatment learners received was beneficial for their speech rate. The TS group slightly outperformed the TB group in oral task 1 both at the immediate and delayed post-test. While performing oral task 2, learners in the TB group were less fluent than learners in the TS group at the immediate post-test, but then both groups demonstrated the same level of fluency at the delayed post-test. These results are visually displayed in Figure 36 (for oral task 1) and Figure 37 (for oral task 2).

Table 87. Descriptive statistics: Speech rate (oral production)

	Task 1		Task 2	
	TS_RusII (n=12)	TB_RusII (n=17)	TS_RusII (n=12)	TB_RusII (n=18)
Pre-test	49.42 (12.35)*	45.24 (11.94)	66.36 (17.01)	65.41 (20.72)
Immediate post-test	58.33 (8.45)	54.22 (13.28)	75.11 (14.06)	70.19 (17.01)
Delayed post-test	62.21 (10.10)	57.09 (13.55)	79.29 (7.13)	79.43 (24.33)

*Mean (SD)

Figure 36. Mean scores on speech rate
(oral task 1)

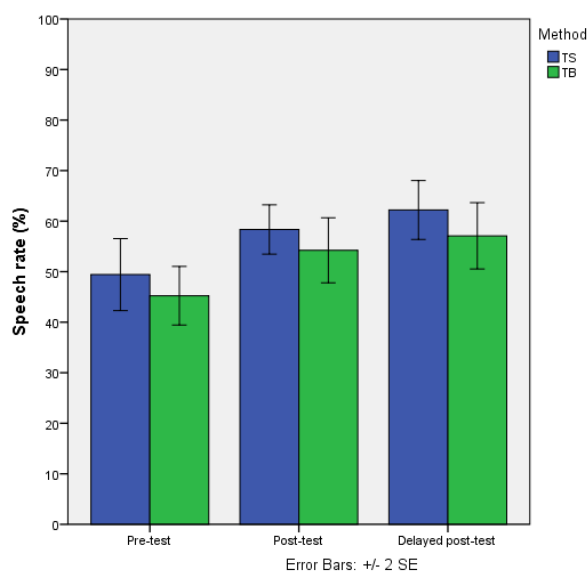
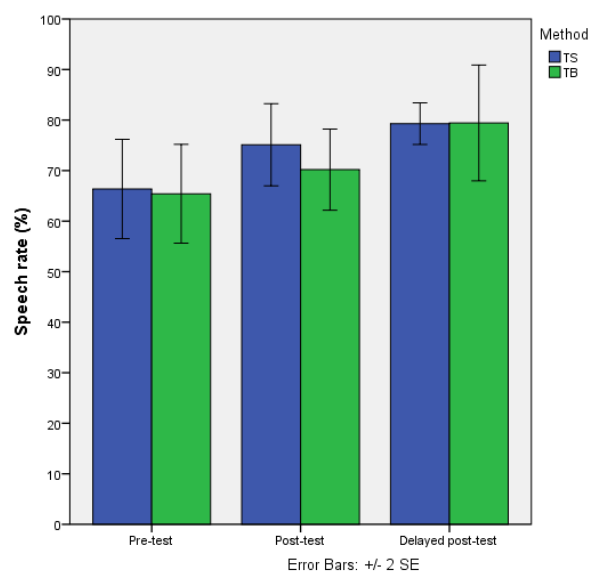


Figure 37. Mean scores on speech rate
(oral task 2)



For oral task 1, Mauchly's sphericity test and Levene's tests showed non-significant results ($p > .05$); therefore, a mixed between-within subjects ANOVA was performed to assess the differences in fluency scores within each group and between the two groups. There was a significant main effect of *Time* ($F(2,54)=19.08, p=.000, partial \eta^2=.414$). The *Time x Method*

interaction was not significant ($p>.05$), which means that the change in fluency scores over time did not significantly differ for the two groups of participants. The analysis of between-subjects effects showed a statistically non-significant effect of *Method* ($p>.05$). This result suggests that there were no significant differences between the TB and TS approaches as far as learners' fluency is concerned. The results of within-subjects and between-subjects comparisons are presented in Table 88.

Table 88. Mixed between-within subjects ANOVA: Speech rate (oral task 1)

	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
<i>Time</i>	2280.45	2,54	19.08	.000	.414
<i>Time x Method</i>	4.41	2,54	.04	.964	.001
<i>Method</i>	421.86	1,27	1.36	.254	.048

Paired samples t-tests showed that there was a significant difference between pre-test and immediate post-test fluency scores in the TS group ($t(11)=-3.13$, $p=.029$, $r=.68$) and in the TB group ($t(16)=-3.29$, $p=.014$, $r=.63$). The results suggest that participants in both groups were significantly more fluent at the immediate post-test than that they were at the pre-test. There was a statistically non-significant difference between immediate and delayed post-test scores ($p>.05$), which means the positive effect of the two types of treatment remained. The difference between pre-test and delayed post-test scores was statistically significant both in the TS group ($t(11)=-4.21$, $p=.004$, $r=.79$) and in the TB group ($t(16)=-3.76$, $p=.005$, $r=.68$). This result suggests that participants in the two groups were significantly more fluent at the delayed post-test than that they were at the pre-test.

For oral task 2, Mauchly's sphericity tests produced non-significant results ($p>.05$), which let us perform a repeated measures ANOVA in order to assess within-groups differences. The differences between groups were assessed by means of independent samples t-tests which report values corrected for the case when the homogeneity of variances is not assumed ($p<.05$ in Levene's tests).

The results from the repeated measures ANOVA showed that there was a significant main effect of *Time* for the TS group ($F(2,22)=3.74$, $p=.040$, *partial* $\eta^2=.254$) and the TB group ($F(2,34)=4.22$, $p=.023$, *partial* $\eta^2=.199$). These results suggest that there was a large change in fluency scores over time in both groups (see Table 89).

Table 89. Repeated measures ANOVA: Speech rate (oral task 2)

Within-subjects effects	<i>Sum of Squares</i>	<i>df</i>	<i>F</i>	<i>p</i>	η^2
TS_RusII	1044.64	2,22	3.74	.040	.254
TB_RusII	1826.94	2,34	4.22	.023	.199

Paired samples t-tests yielded statistically non-significant differences between pre-test and immediate post-test scores ($p>.05$), which may mean that the treatment did not have an immediate effect on learners' fluency. The difference between immediate post-test and delayed post-test fluency scores for both groups of participants was also statistically non-significant. However, there was a significant difference between the pre-test and delayed post-test scores both in the TS group ($t(11)=-2.57$, $p=.026$, $r=.61$) and in the TB group ($t(17)=-2.42$, $p=.027$, $r=.51$). This result suggests that both types of treatment had a positive effect on learners' speech rate.

The results from a series of independent samples t-tests showed that there were no significant differences between the two groups' fluency scores at any point in time ($p>.05$), as reported in Table 90.

Table 90. Independent samples t-tests: Speech rate (oral task 2)

Time	Independent samples t-tests		
	<i>t</i>	<i>df</i>	<i>p</i>
Pre-test	.13	28	.897
Immediate post-test	.83	28	.414
Delayed post-test	-.02	21.12	.982

Taken together, the results obtained in oral task 1 suggest that participants in both groups were significantly more fluent at the immediate post-test than that they were at the pre-test, which means that both task-supported and task-based treatment had a positive effect on learners' fluency. This effect was long-lasting since participants obtained significantly higher fluency scores at the delayed post-test than at the pre-test. While performing oral task 2, both groups demonstrated significantly higher fluency at the delayed post-test as compared with the pre-test. Analysis of mean scores showed that learners in the TS group were slightly more fluent than learners in the TB group at the two post-tests, but there was no significant difference between the two groups. This means that the two types of treatment had a similar effect on fluency of oral production.

6.7 Summary of the results

In this section the results of statistical analyses are summarized according to the three research questions which the present study aims at answering. Table 91 presents the summary of results for each measure applied in this study. These results will be discussed in the next chapter.

***RQ 1:** Do the task-supported approach and the task-based approach help to promote the accurate use of case forms and prepositions?*

Learners in both groups improved their use of prepositions in written and oral performance over time (from the pre-test to the delayed post-test), but this change was not significant. As far as the accurate use of case forms is concerned, learners in both groups improved their scores at the immediate post-test, but then demonstrated a drop in scores at the delayed post-test both in their written production and on three grammatical tests. Within-subjects comparisons showed that at the delayed post-test learners in both groups reached the level that they had at the pre-test, which means that the positive effect of the treatment on the use of case forms was not long-lasting. However, in oral performance a different tendency was observed. Learners in the TB group improved their use of case forms over time, although not statistically significantly. Learners in the TS group showed decreased accuracy in the use of case forms at the immediate post-test, but then significantly improved on this measure at the delayed post-test and reached the level they had at the pre-test. The TS group and the TB group did not significantly differ on any of the measures. Taken together, the results obtained indicate that neither task-supported nor task-based treatment led to a significant improvement in the use of prepositions and case forms in the long term.

***RQ 2:** Which of the two approaches is more efficient for learning new lexical items (Russian verbs of motion with prefixes)?*

In terms of the use of the verbs of motion, participants in both groups demonstrated a statistically significant improvement in their scores obtained for oral and written performance and for grammatical tests at the two post-tests as compared with the pre-test. These results suggest that both task-supported and task-based treatment had a positive and long-lasting effect on learning verbs of motion. Between-groups comparisons showed that participants in the TB group outperformed the TS group on all post-tests. For some measures, this difference was statistically significant (for example, the difference in delayed post-test scores on TLU of verbs of motion, multiple choice and grammaticality judgment tests); for other measures, the difference did not reach statistical significance. These results let us conclude that task-based

treatment was more efficient for learning new target items (verbs of motion) than task-supported treatment.

RQ 3: *How does each approach affect learners' oral and written production in terms of general accuracy, syntactic and lexical complexity, and fluency?*

Analysis of CAF scores showed that in written production learners' syntactic complexity measured via subordination (clauses per T-unit) and lexical complexity measured by means of Guiraud's Index significantly increased after the treatment. There was no significant change in phrasal complexity (mean length of clause) over time. General accuracy significantly improved in both groups immediately after the treatment. However, learners in the TS group showed decreased accuracy at the delayed post-test to the level they had at the pre-test, whereas participants in the TB group made significantly fewer errors at the delayed post-test as compared with the pre-test. There were no significant differences between the two groups in their written accuracy and complexity, although learners in the TB group outperformed the TS group on the post-tests.

Analysis of oral data detected different dynamics of CAF scores obtained for oral task 1 (room description) and oral task 2 (map task). While performing oral task 1, learners in both groups demonstrated no significant changes in accuracy and syntactic complexity scores over time. There were positive significant changes in lexical complexity and fluency both after the task-supported and the task-based treatment. The TS group outperformed the TB group on all measures at the post-tests, although this difference did not reach statistical significance.

While performing oral task 2, both groups showed statistically significant improvement in scores on general accuracy, lexical complexity and fluency from the pre-test to the delayed post-test. There were no significant differences in syntactic complexity within each group over time. Participants in the TB group significantly outperformed the TS group on lexical complexity at the immediate post-test and demonstrated higher lexical complexity than the TS group at the delayed post-test. The TB group also outperformed the TS group on accuracy and syntactic complexity. However, learners in the TS group were slightly more fluent than learners in the TB group.

Table 91. Summary of the results

RQ	Measure	Within-subjects differences	Between-subjects differences
The effects of the task-supported and the task-based	TLU of prepositions	There was a tendency towards a positive change of scores over time, but this change was not significant in the two groups.	The TB group outperformed the TS group both in oral and written post-tests, although not significantly.
	TLU of case forms	<u>In written production:</u> Participants in the TS group did not significantly improve	The TS group and the TB group did not significantly differ on any of the

treatment on the use of case forms and prepositions		<p>their use of case forms. Learners in the TB group significantly improved their scores at the immediate post-test, but then demonstrated a significant drop in scores at the delayed post-test to the level they had at the pre-test.</p> <p><u>In oral production:</u> Learners in the TS group decreased their accuracy in the use of case forms at the immediate post-test, but then significantly improved on this measure at the delayed post-test and reached the level they had at the pre-test. Learners in the TB group showed improved results both at the immediate and delayed post-tests, but not statistically significantly.</p>	<p>oral or written tests.</p> <p><u>In written production:</u> The TB group outperformed the TS group at the immediate post-test and showed worse results than the TS group at the delayed post-test.</p> <p><u>In oral production:</u> The TB group outperformed the TS group in the two post-tests.</p>
	FB	Both the TS and TB groups statistically significantly improved their scores from the pre-test to the immediate post-test, but then demonstrated a significant drop of scores at the delayed post-test. In both groups, learners' scores decreased at the delayed post-test to the level they had at the pre-test.	The TS group outperformed the TB group in both post-tests, although the difference between the two groups did not reach statistical significance.
	MC	Learners in the TS group significantly improved their scores at the immediate post-test, but then demonstrated a drop of scores at the delayed post-test. Participants in the TB group did not significantly improve their scores.	The TB group slightly outperformed the TS group in the two post-tests.
	GJ	Neither the TS nor the TB groups showed significant improvement at the two post-tests.	The TB group performed slightly better than the TS group at the immediate post-test but then showed worse results than the TS group at the delayed post-test.
The effects of the task-supported and the task-based treatment on the use of verbs of motion	Total amount of VM	Participants in the two groups produced significantly more VM at the two post-tests than at the pre-test both in their written and oral performance.	<p><u>In written production:</u> Although there was no significant difference between the two groups, there was a tendency of the TB group to produce more verbs of motion than in the TS group.</p> <p><u>In oral production:</u> Participants in the TB group produced significantly more verbs of motion than learners in the TS group at the immediate and delayed post-tests.</p>
	TLU of VM	Participants in the two groups accurately produced significantly more VM at the immediate post-test and delayed post-test than at the pre-test both in their written and oral performance.	Participants in the TB group accurately produced significantly more verbs of motion than learners in the TS group at the oral immediate and delayed post-tests and at the written delayed post-test.
	FB	Participants both in the TS and TB groups significantly improved their use of verbs of motion from the pre-test to the immediate post-test and from the pre-test to the delayed post-test.	The TS group and the TB group did not significantly differ in their scores. The TB group outperformed the TS group at the delayed post-test.
	MC	Participants in both groups demonstrated a statistically significant improvement in their scores at the two post-tests as compared with the pre-test.	Learners in the TB group outperformed the TS group at the immediate post-test and obtained significantly better scores at the

			delayed post-test than the TS learners.
	GJ	Learners in both groups performed significantly better at the immediate and delayed post-tests than at the pre-test.	The TB group outperformed the TS group at the immediate post-test and performed significantly better at the delayed post-test than the TS group.
The effects of the task-supported and the task-based treatment on learners' written production	General accuracy	The TS group significantly improved their accuracy at the immediate post-test, but then their scores decreased at the delayed post-test to the level they had at the pre-test. The TB group made significantly fewer errors at the two post-tests as compared with the pre-test.	There were no statistically significant differences between the two groups. Learners in the TB group made fewer errors than learners in the TS group at the delayed post-test.
	Syntactic complexity (mean length of clause)	Participants in both groups produced shorter clauses at the delayed post-test than at the pre-test, but there was no significant change in mean length of clause over time.	There were no significant differences between the task-based treatment and the task-supported treatment. The TB group slightly outperformed The TS group at the two post-tests.
	Syntactic complexity (clauses per T-unit)	Significantly increased in both groups	
	Lexical complexity	Significantly increased in both groups	There were no statistically significant differences between two groups. The TB group outperformed the TS group at the two post-tests.
The effects of the task-supported and the task-based treatment on learners' oral production	General accuracy	<u>In oral task 1:</u> There was no significant change in scores over time in the two groups, although there was a tendency towards improvement of learners' general accuracy at the delayed post-test as compared with the pre-test. <u>In oral task 2:</u> Participants in both groups made significantly fewer errors on the two post-tests than the pre-test.	There were no statistically significant differences between two groups. <u>In oral task 1:</u> The TS group outperformed the TB group at the two post-tests. <u>In oral task 2:</u> Participants in the TB group outperformed learners in the TS group at the delayed post-test.
	Syntactic complexity (mean length of clause)	There was no significant change in scores over time.	There were no statistically significant differences between the two groups. <u>In oral task 1:</u> The TS group outperformed the TB group at the immediate post-test. At the delayed post-test learners in the TS group produced shorter clauses than learners in the TB group. <u>In oral task 2:</u> The TB group produced longer clauses than the TS group at the three tests.
	Lexical complexity	Participants in both groups demonstrated significantly higher lexical complexity at the immediate post-test than at the pre-test. <u>In oral task 1:</u> Learners in the TS group increased their lexical complexity on the delayed post-test as compared with the pre-test, but not statistically significantly. The TB groups demonstrated significantly higher lexical complexity at the delayed post-test than at the pre-test. <u>In oral task 2:</u> Both groups statistically significantly improved their scores on lexical complexity from the pre-test to the delayed post-test.	<u>In oral task 1:</u> The TS group outperformed the TB group on lexical complexity at the two post-tests, although this difference did not reach statistical significance. <u>In oral task 2:</u> Participants in the TB group performed significantly worse than learners in the TS group at the pre-test, but significantly outperformed the TS group at the immediate post-test and demonstrated higher lexical complexity than the TS group at the delayed post-test.
	Fluency	Participants in both groups were	<u>In oral task 1:</u> The TS group

		<p>significantly more fluent at the delayed post-test than that they were at the pre-test.</p>	<p>outperformed the TB group on the two post-tests, although this difference did not reach the level of statistical significance.</p> <p><u>In oral task 2:</u> The TS and TB groups did not significantly differ in their fluency scores. The TS group outperformed the TB group at the immediate post-test.</p>
--	--	--	---

CHAPTER VII

DISCUSSION AND CONCLUSION

7.1 Introduction

In the previous chapter we presented the results obtained after analyzing oral and written data collected from fifty four learners of Russian before and after task-supported and task-based treatment. In this chapter, we elaborate further on the results obtained and consider them in relation to previous research and our hypotheses about the effects of two types of treatment on learners' production and the use of the target forms. This chapter is organized around three research questions outlined in Chapter V (Section 5.3). We compare the effects of task-supported and task-based treatment on the use of case forms, verbs of motion and on CAF dimensions. When appropriate, extracts from learners' written and oral performance are included in order to improve our understanding of the measures used and the progress made. Then the implications of the results of this study for teaching Russian as a foreign language are discussed. Finally, some limitations of the study are acknowledged, and some directions for further research are proposed.

7.2 The effects of task-supported and task-based treatment on the use of prepositions and case forms

We hypothesized that both groups would improve their use of prepositions after the treatment, without any significant difference between the two groups. Our hypothesis was confirmed by the results of statistical analyses. **Learners in both groups used prepositions more accurately in their written and oral performance at the two post-tests, although the change in scores was not statistically significant.** The lack of significant differences can be explained by the fact that participants could use most of the prepositions correctly before the treatment. This suggestion is confirmed by high mean scores on the TLU of prepositions at the pre-test: learners accurately produced more than 80% of prepositions in their written performance and about 70% of prepositions while performing oral task 1. There were no

significant differences between the two groups. This result suggests that **task-supported and task-based treatment had a similar effect on accuracy in using Russian prepositions.**

As far as the TLU of case forms is concerned, we suggested that learners in both groups would show improvement in the use of case forms as a result of the treatment and that after the task-supported treatment learners will use case forms more accurately than after the task-based treatment. This hypothesis was only partially confirmed. Learners in both groups improved their scores at the immediate post-test, but then demonstrated a drop in scores at the delayed post-test in their written production and on three grammatical tests. Within-subjects comparisons showed that **at the delayed post-test, three months later, learners in both groups reached the level that they had at the pre-test**, which means that **the positive effect of treatment on the use of case forms did not last.** These results can be partially explained by the experiment conditions and learners' motivation. The immediate post-test coincided with learners' final exam in Russian and they likely worked harder in order to perform well and obtain a good mark, whereas the delayed post-test was a part of current control and learners were not given a mark. It resulted in lower motivation and worse scores.

While performing oral task 1, learners in the TB group improved their use of case forms over time, although not statistically significantly. Learners in the TS group decreased their accuracy in the use of case forms at the immediate post-test, but then significantly improved on this measure at the delayed post-test and reached the level they had at the pre-test. Therefore, **learners in both groups demonstrated no significant improvement in the use of case forms in their oral performance over a long term perspective.**

We did not statistically compare the results obtained for written and oral data. However, the mean scores show that the number of errors in case forms was much higher in oral production than in written performance. About 39% of all case forms were produced inaccurately at the oral delayed post-test, whereas only 17% of all case forms produced in written performance contained an error. This result can be explained by the fact that while performing the written task, learners had more time to think and choose an appropriate grammar form. Oral performance requires the use of procedural (implicit) knowledge to a greater extent (Ellis 2005), and case forms had not become part of learners' implicit knowledge yet. It is also possible that two modes of performance, written and oral, affect learners' level of anxiety in a different way. A simple observation demonstrated that learners were more nervous while speaking than during the performance of the written task.

Between-groups comparisons showed that the TS group and the TB group did not significantly differ in their use of case forms measured by calculating the TLU of case forms and by fill in the blanks, multiple choice, and grammaticality judgment tests. This result suggests that **two types of treatment had a similar short-term effect on learning case forms in Russian.**

We are not aware of any study which compares the effect of two types of treatment on learners' accuracy in the use of prepositions and case forms. However, there are a number of studies which have focused on the process of learning the Russian case system within one (usually task-supported or grammar-based) teaching approach (Denissenko 2016; Rubinstein 1995; Quero Gervilla 2005, among others).

Our results are similar to those obtained by other researchers. Denissenko (2016) demonstrated that a number and variety of errors in case forms stayed high after many hours of treatment (150 hours). She found that four groups of learners with different levels of proficiency (A1-B1) showed no significant improvement in the use of case forms. Quero Gervilla (2005) focused on errors in each of the case forms in Russian. She calculated the percentage of errors for each case in grammatical tests and in written production. She found that learners with higher levels of proficiency were more accurate with the case forms while completing the grammatical test. However, the analysis of learners' compositions showed that the process of case acquisition was characterized by progress and setbacks. Interestingly, an average percentage of errors in case forms in written production reported by Quero Gervilla (2005) is 16,48%, almost the same as in our study (16,69%).

Therefore, our results obtained after 14 hours of task-supported and task-based treatment are similar to those reported in other studies with more participants and more hours of instruction. TBLT did not make a difference as far as the use of case forms is concerned.

7.3 The effects of task-supported and task-based treatment on the use of verbs of motion

Our hypothesis was that learners in the TB group would produce more verbs of motion and be more accurate in their use than learners in the TS group while performing oral and written tasks, but the TS group would demonstrate higher scores on grammar tests than the TB group.

The first part of our hypothesis was confirmed by results of statistical tests. Between-groups comparisons showed that **participants in the TB group outperformed participants in the TS group on the total amount of verbs of motion produced and on the amount of**

accurately produced verbs of motion at the immediate and delayed post-tests. At the delayed post-test, the difference between groups was statistically significant, which means that learners who had received the task-based treatment accurately produced significantly more verbs of motion than learners who had been exposed to the task-supported treatment.

The second part of our hypothesis was rejected since the **TB group outperformed the TS group on the three grammatical tests.** Difference in delayed post-test scores on multiple choice and grammaticality judgment tests between the two groups was statistically significant. For fill in the blanks scores, this difference did not reach statistical significance. Taken together, these results indicate that **the task-based treatment was more efficient for learning verbs of motion with their complex morphological characteristics than the task-supported treatment.**

Our results differ from those obtained by De la Fuente (2006) who compared the effect of three types of treatment (PPP, task-supported and task-based) on learning 15 Spanish nouns for food items. She found that there was no difference between the performance of the TB group and the TS group on word retrieval and that the TS group was more accurate with gender and number morphology than the TB group. One of the reasons for the difference in the results obtained by De la Fuente and in our study could be a difference in experiment conditions and in measures applied in each study. The duration of each session in her research was 50 minutes, whereas treatment in our study included 14 hours. De la Fuente used vocabulary discrete-point tests to measure outcomes, and participants in the current research performed oral and written tasks and completed three grammar tests.

The fact that learners in the TB group obtained significantly better results than learners in the TS group doesn't mean that the task-supported approach was not efficient. Within-subjects comparisons showed that **participants in both groups demonstrated statistically significant improvement in their grammar scores and scores obtained for the TLU of verbs of motion in their oral and written performance** at the two post-tests as compared with the pre-test. In order to illustrate to what extent learners' performance improved after the treatment, four examples of oral production (two participants performing oral task 2) are presented in Table 92.

Table 92. Examples of learners' oral performance (oral task 2)

	Pre-test	Delayed post-test
Learner 1 (TS group)	Хорошо я должна идти направо в улице Попова. А потом мы видим спортивный зал слева. А ты должна идти там рядом. Когда ты спортивный зал и ты с улицы Пушкина, мы будем... Когда ты в банк, иди направо.	Ты в метро. Тебе надо выйти (TF, correct) из метро и повернуть направо и перейти (TF, correct) улица Попова. До конца ты увидишь слева спортивный зал. Повернуть ты должен повернуть слева и идти прямо по улице

	<p>Мы видим направо кафе, а рядом кафе физический факультет. Напротив физический факультет библиотека, поэтому ты должна идти улица Чехова. Ты будешь в библиотеке. Сейчас ты должна идти площади Сахарова. А это между школа языков и институт языкознания. А там ты увидишь филологический факультет.</p> <p>Total TF: 0 Correct TF: 0</p>	<p>Пушкина до банка. А потом перейти (TF, correct) улицу и ты должен дойти (TF, correct) до библиотеку. Перейти (TF, incorrect) через библиотеку и пройти (TF, correct) мимо площадь Сахарова. И ты заканчиваешь филологический факультете.</p> <p>Total TF: 6 TF which are not repeated: 4 Correct TF: 5</p>
<p>Learner 2 (TB group)</p>	<p>Мы на улице Попова. Мы хотим идти на банк. И мы идти на улице Попова. Это справа от метро. Потом мы банке, мы идем на улице Пушкины до библиотеки. Потом библиотеки мы идем на улице Чехове. И мы идем слева до филологического факультета.</p> <p>Total TF: 0 Correct TF: 0</p>	<p>Ты в станции метро. Ты выйдешь (TF, correct) из метро. Ты должен повернуть направо до улицы Пушкина. Там ты должен повернуть еще раз, но налево. Ты зайдешь (TF, correct) в банк, ты пройдешь (TF, incorrect) улицу и войдешь (TF, correct) в библиотеку. Потом ты выйдешь (TF, correct) из библиотеки, ты перейдешь (TF, correct) через площадь Сахарова и ты дойдешь (TF, correct) до филологического факультета.</p> <p>Total TF: 7 TF which are not repeated: 6 Correct TF: 6</p>

Learner 1 was exposed to the task-supported treatment, and Learner 2 received the task-based treatment. The examples demonstrate that both learners did not know the target forms prior to the treatment and correctly produced various VM with prefixes after the experimental sessions. These results suggest that both task-supported and task-based treatment had a positive and long-lasting effect on learning verbs of motion.

7.4 The effects of task-supported and task-based treatment on learners' accuracy, complexity and fluency

In Chapter V, we stated four hypotheses regarding the effect of treatment on four CAF dimensions. The results obtained for each measure applied in this study are discussed in the following subsections.

7.4.1 Accuracy

As far as accuracy is concerned, we hypothesized that learners would be more accurate both after task-supported and task-based treatment and learners in the TS group would demonstrate higher accuracy than learners in the TB group. The results differed for oral and written production, on the one hand, and for oral task 1 and oral task 2, on the other hand. **General accuracy of written production significantly improved at the immediate post-test in both groups.** However, learners in the TS group decreased their accuracy at the delayed post-test to the level they had at the pre-test, which means that **the positive effect of the task-supported treatment on learners' accuracy did not last.** Participants in the TB group made significantly fewer errors at the delayed post-test as compared with the pre-test, which means that **the task-based treatment had a positive long-lasting effect on learners' written accuracy.** This means that the task-based teaching approach was more beneficial for improving accuracy of participants' written performance than TSLT. Therefore, written data refuted our hypothesis as far as the effect of two types of treatment on accuracy is concerned.

The lack of significant changes in accuracy after many hours of formal instruction was reported in other studies. For example, Denissenko (2016) found that participants who had A1 level of language proficiency in Russian showed no significant improvement in accuracy after 150 hours of formal (task-supported) instruction. Accuracy in her study was measured by counting a number of errors in verbs per total number of verbs produced and counting errors in different case forms.

Henry (1996) analyzed 67 short autobiographies written by English-speaking learners of Russian from beginner to intermediate level. In this study, all types of errors were counted including errors in letter writing, spelling, punctuation, syntax, morphology, and vocabulary. The research found that there were no significant differences in accuracy measured by means of error-free T-units between four groups of learners. Henry (1996) concluded that the lack of significant differences between groups on the accuracy measures indicate that a grammar-based curriculum does not necessarily yield accurate writers. The results reported by Denissenko (2016) and Henry (1996) are similar to those obtained in our research.

Analysis of oral data shows that **in the performance of oral task 1** (the room description), **neither the TS group nor the TB group demonstrated a significant change in accuracy scores** over time. However, **in the performance of oral task 2** (the map task), **both groups significantly improved their accuracy** and made significantly fewer grammatical and

lexical errors on the delayed post-test as compared with the pre-test. Different results which were obtained from the two oral tasks can be explained by different language material required by each task. If we compare these results with learners' improvement on specific accuracy measures, we can see their direct correlation with general accuracy scores. The room description in Russian involved the use of various prepositions and the corresponding case forms. The results on the TLU of case forms discussed above indicated that participants did not master these forms after both types of treatment. It could have affected general accuracy scores. The map task required mostly the use of verbs of motion. We found that both teaching methods, task-based and task-supported, were beneficial for learning these forms. As a consequence, general accuracy scores obtained in this task improved after the treatment.

Between-groups comparisons showed that **the two groups did not significantly differ in accuracy of their oral and written performance**. Similarly, González-Lloret and Nielson (2015) did not find significant differences in accuracy after TBLT and grammar-based teaching of L2 Spanish. Lai, Zhao, and Wang (2011) came to the same conclusion for L2 Chinese. This means that task-supported and task-based treatments do not differ as far as their effect on accuracy is concerned.

7.4.2 Syntactic complexity

We did not expect significant differences in syntactic complexity for any of the two groups. Our hypothesis was partially confirmed. The analysis of written and oral data showed that **there were no significant changes in phrasal complexity (mean length of clause) over time either in the TS or the TB group**. In other words, learners did not produce longer clauses as a result of treatment.

Syntactic complexity of learners' written production measured via subordination (clauses per T-unit) significantly increased after the treatment. This means that learners wrote significantly more complex sentences as a result of both task-supported and task-based treatment. Denissenko (2016) also reported that participants in her study significantly improved their written syntactic complexity as measured by the ratio of clauses per sentence. Taken together, these results indicate that both task-supported and task-based treatments are beneficial for increasing subordinational complexity (the amount of subordinate clauses used) of learners' written production, but not their phrasal complexity (the length of clauses produced).

In the oral performance of both groups, the ratio of clauses per T-unit did not change after the treatment, confirming our hypothesis. In the performance of oral task 1, there were no cases of subordination at all. The performance of oral task 2 contained very few cases of subordination. These results were expected, since, generally, subordination is not frequently used in oral production, especially by learners with a low level of proficiency. Moreover, one cannot expect that these types of task favor subordination, since descriptions and map directions do not require complex syntactic structures.

As far as between-groups differences are concerned, **the two groups in our study did not significantly differ on any of the measures of syntactic complexity**. This means that the task-supported and task-based treatments are not that different with respect to their effect on learners' syntactic complexity. The results reported in other studies are contradictory. Lai, Zhao, and Wang (2011) who investigated L2 Chinese learning obtained the same results as those in our study. They measured syntactic complexity by calculating mean length of T-units and found no significant differences between the two groups, one of which completed a task-based online course while the other group completed a grammar-based online course. However, González-Lloret and Nielson (2015) reported that learners in the TB group produced significantly larger utterances in Spanish than students from the grammar-based group.

7.4.3 Lexical complexity

We suggested that learners would use a higher variety of vocabulary after the experimental sessions. We also hypothesized that the task-based treatment would result in higher lexical complexity than the task-supported treatment. Analysis of oral and written data completely confirmed our hypothesis. **Learners in both groups demonstrated significantly higher lexical complexity at the two post-tests as compared with the pre-test**. That is, they used a greater variety of vocabulary immediately after the treatment and demonstrated the same level of lexical richness a few months after the treatment. This result shows that participants broadened their vocabulary and successfully learned new lexical items both after task-supported and task-based sessions.

Significant lexical growth was reported in other studies. For example, Denissenko (2016) used the same measure of lexical complexity in her study (Guiraud's Index). She found that participants used significantly richer vocabulary in Russian after 150 hours of formal (task-

supported) instruction. This means that formal instruction can have a positive effect on vocabulary development.

Between-groups comparisons showed that the task-based treatment was more beneficial for increasing learners' lexical diversity than the task-supported treatment. **Participants in the TB group significantly outperformed the TS group on lexical complexity at the immediate post-test** (oral task 2) and demonstrated higher lexical complexity than the TS group at the oral delayed post-test (oral task 2) and two written post-tests. However, the TS group obtained higher scores on Guiraud's Index than the TB group while performing oral task 1, but the difference in scores between the two groups did not reach statistical significance. This means that the task-supported treatment was ultimately more efficient for learning prepositions of space and lexical items related to the description of a room, but produced worse results for verbs of motion and the corresponding prepositions as compared with the task-based treatment.

7.4.4 Fluency

Our hypothesis was that learners in the TB group would improve their fluency, whereas learners in the TS group would show no significant differences in their fluency scores. We also suggested that the task-based treatment would result in higher oral fluency than the task-supported treatment. Only the first part of our hypothesis regarding the TB group's progress was confirmed. Learners in the TB group significantly improved their fluency after the treatment as we expected. Learners in the TS group also demonstrated significantly higher fluency on the post-tests as compared with their pre-tests. This means that **both task-supported and task-based treatment resulted in positive significant changes in oral fluency**.

Denissenko (2016) obtained similar results for learners' written production in L2 Russian. She reported that participants significantly improved their written fluency measured by dividing a number of words produced by the total time of the test (40 minutes). Taken together, these results indicate that formal instruction which uses tasks not as units of syllabus design but as an instrument to promote production may result in fluent L2 speakers and writers, as well as communicative language teaching.

Comparisons of the scores on speech rate obtained by the two groups show that **there were no significant differences in oral fluency between the groups**. According to the mean scores on speech rate, learners in the TS group were slightly more fluent than learners in the TB

group in the performance of both tasks. This means that the second part of our hypothesis was refuted. Two types of treatment had a similar positive effect on learners' fluency.

Our results differ from those obtained in other comparative studies. Lai, Zhao, and Wang (2011) reported that their TBLT group demonstrated significantly higher oral fluency in Chinese than their counterparts in the control group. González-Lloret and Nielson (2015) also found that a TBLT group performed significantly better on fluency than students from a GB group. One of the possible reasons for these differences in results could be the nature of form-focused treatment in each study. In the current research, the use of tasks at the production stage of the task-supported treatment gave learners a possibility for communication and elaboration of their output. Grammar-based language teaching which was described in González-Lloret and Nielson's study limits this possibility and often replaces production with drills and repetitions of predetermined models. Therefore, it resulted in lower fluency as compared with the communicative approach. Another possible explanation is the difference in the duration of the courses and students' profile. We suggest that the shorter the course, the more similar the results of different types of treatment will be. The treatment in the present research lasted 4 weeks, whereas participants in González-Lloret and Nielson's study (2015) and Lai, Zhao, and Wang's study (2011) had 8 and 12 weeks of treatment respectively. In longer courses the differences between the two types of treatment are likely to be more visible than in a short course.

7.5 Conclusion

The main objective of the present research was to compare two approaches to language teaching in order to determine whether one of them was more efficient for learning morphologically complex target forms (Russian case forms and verbs of motion) and for improving learners' oral and written production in terms of accuracy, fluency, syntactic complexity and lexical diversity. This objective determined the focus of the study on some theoretical issues and practical problems of task design and implementation. First, we had to reconsider the question of the efficiency of a task-supported approach which was criticized by some advocates of communicative language teaching. Second, we had to revisit some of the widely accepted TBLT assumptions regarding task design and task sequencing in light of high code complexity and to question the efficiency of a task-based approach for learning a

morphologically rich language (Russian). After having carried out the whole experiment and analyzing the results, the following conclusions have been drawn:

1. In spite of the criticism that form-focused instruction has received over the last few decades, it has proved to have a positive effect on L2 learning. The results obtained both from grammatical tests and analyses of oral and written data indicated that task-supported treatment was beneficial for the improvement of learners' oral fluency, written syntactic complexity (the ratio of clauses per T-units) and written and oral lexical diversity. However, TSLT did not significantly affect learners' written and oral accuracy. Only in the performance of oral task 2 was there a significant change in accuracy scores after the treatment. These results support a conclusion made by Henry (1996), who argued that formal instruction does not necessarily yield accurate speakers and writers.

2. The results of the present research discredit a very common belief that Russian cannot be learned unless explicit focus on forms is provided given the complexities of the language for non-Russian speakers. In fact, most Russian courses in Russia, Europe and the United States do heavily focus on forms. Communicative activities are usually used to support language practice. Almost all Russian textbooks, especially for the lower levels, are based on the PPP principle. However, the results drawn from the present analysis show that not only is TBLT as efficient as focus-on-forms approaches for improving learners' accuracy, but it also promotes the development of metalinguistic knowledge and significantly contributes to the growth of learners' vocabulary (lexical complexity), oral fluency and written syntactic complexity.

In spite of these positive findings, TBLT cannot be considered a "panacea" for different kinds of problems and challenges that learners face in the process of L2 learning. In particular, in the current research, task-based treatment did not result in significantly more accurate use of prepositions and case forms.

3. As far as the impact of the two types of treatment on the use of the target forms is concerned, the results varied for different target forms. Neither task-supported nor task-based treatment helped to achieve the desired results in the use of prepositions and case forms. Learners in both groups did not demonstrate a significant change in the use of these forms after the treatment. However, both types of treatment proved to have a positive long-lasting effect on learning verbs of motion. Participants in both groups demonstrated significant improvement in their grammar scores and in the accurate use of verbs of motion in their oral and written performance.

The difference in the results cannot be explained by higher difficulty of one of the forms. Both case forms and verbs of motion are considered complex forms, and their acquisition entails a long process, which usually progresses in a zigzag fashion. In my opinion, the results obtained in the present study can be explained in line with VanPatten's theory of input processing (2002). He argued that, firstly, learners process input for meaning before they process it for form and, secondly, they prefer processing "more meaningful" morphology before "less" or "nonmeaningful" morphology. Most Russian case forms are not communicatively meaningful forms, that is, an error in the case form does not usually lead to misunderstanding. Verbs of motion, on the other hand, are lexical items with "meaningful" morphological characteristics (verbal tense, masculine/feminine forms in the past tense, among others). This could explain why participants in our study successfully learnt these verbs, but did not demonstrate a significant improvement in the proper use of case endings. As VanPatten (2002, p. 758) claimed, in order to process form that is not meaningful, learners "must be able to process informational or communicative content at no (or little) cost to attention." This did not happen in our study due to the low level of language proficiency of the participants.

4. Comparing the effects that two types of treatment had on different aspects of learners' production showed that the TS and TB groups did not significantly differ on accuracy, syntactic complexity and fluency of their oral and written performance. The task-supported and task-based treatment also had a similar effect on accuracy in using Russian prepositions and case forms. However, TBLT gave significantly better results than TSLT as far as lexical complexity and the target-like use of verbs of motion are concerned. These seem to be important and interesting findings, because they discredit another common belief about form-focused approaches and CLT. Communicative approaches are usually associated with higher fluency which is reached at the cost of accuracy, and this was not the case in our study. The significant difference between the two approaches involved the dimension of lexical diversity, which is, without any doubt, a crucial part of language learning.

5. Based on real classroom experience and pilot results, we would like to suggest that the degree of linguistic difficulty should be given more careful consideration when designing tasks. Unless carefully controlled during design, linguistic demands that are overwhelmingly high or too far from the learner's possibilities may pose problems. Following Gilabert and Castellví (in press), we argue that for morphologically complex languages such as Russian linguistic complexity should be an integral part of task design that is carefully considered and integrated in order to guarantee successful task completion. The design of meaningful, realistic and interesting

tasks that present meticulously crafted pre-tasks, tasks and post-tasks could solve a lot of the problems of integrating meaning and form in tasks for the teaching of L2 Russian.

7.6 Implications

We believe that the present research has various implications at the theoretical, methodological and pedagogical levels. To our knowledge, our research stands as one of the first studies aimed at comparing TBLT and TSLT on the basis of L2 Russian, a morphologically complex language which has generally received little attention in SLA studies. Most studies have overwhelmingly focused on task features affecting cognitive complexity with linguistic difficulty largely remaining outside the picture. The findings of the current research provide evidence for the importance of linguistic complexity in a framework of task complexity when teaching a morphologically rich language.

The current research should hopefully stand as a step in the development of the methodology of teaching Russian as a foreign language. It discredits a common belief about communicative language teaching which exists among Russian teachers. Many teachers believe that Russian cannot be learned unless explicit focus on forms is provided given the complexities of the language for non-Russian speakers, and communicative activities are usually designed to support such explicit teaching and learning. The findings of this study provide empirical evidence for the positive effects of a task-based approach on learning complex target forms in Russian and on learners' accuracy, fluency and complexity.

Finally, the findings of the present research have some pedagogical value and are applicable for the design of task-based syllabi. Our study provides an example of how tasks could be designed and organized in order to be included in the classroom setting. It also contains the results of a pilot on the elaboration of assessment tasks which could be applied in the design of evaluation tools.

7.7 Limitations and further research

Inevitably, some limitations have been imposed on this study which should be acknowledged at this point. They have to do with the conditions under which the experiment was carried out and the need to incorporate information from learners' variables.

In the first place, classroom conditions imposed some limitations which we had to assume and take into account while interpreting the results of the experiment. The first problem we had to deal with was missing data. The number of participants was not the same in two groups since some learners left the Russian course before the treatment had ended or missed one of the post-tests. Additionally, as was explained in Chapter VI, we had to exclude extreme outliers at the stage of preliminary analyses. Pairwise deletion (available-case analysis) was used in order to handle missing data (Peugh & Enders 2004). It permitted minimizing the loss that occurs in listwise deletion, which removes all data for a case that has one or more missing values. Obviously, it is desirable to compare groups with equal and larger number of participants, since it allows for obtaining more reliable data. This is not an easy task with L2 Russian in a foreign language setting. Applying TBLT and TSLT approaches for teaching Russian in a different context (in Russian-speaking countries, for example) and under different conditions (a whole semester or even a whole academic year of treatment) could be an area for further research.

The second limitation was caused by the integration of the experiment into the university curriculum and the schedule it required. As a result, two post-tests (the immediate post-test for Rus I and the delayed post-test for Rus II) coincided with learners' final exams at the end of the spring term (in June) and at the end of the autumn term (in January). This could have affected learners' motivation and the results of these post-tests, because participants made more effort to perform well in these tests in order to obtain a good mark. Future studies should be designed in such a way that would avoid this limitation.

The third limitation of the present study which was discussed in Section 5.9.2.4 is the lack of data for measuring written fluency. This could be easily avoided in future research by introducing time limits for learners' written performance.

Finally, it is important to keep in mind that the TS group and the TB group were formed by different participants. This study has controlled for learners' initial level of proficiency in Russian by comparing pre-test results. However, it has largely ignored other learner factors such as ability variables (working memory, reasoning, aptitude, etc.) and affective variables

(motivation, level of anxiety, willingness to communicate, etc.) (Robinson 2005). It is quite possible that these factors had some kind of impact on learners' production and could be the reason for differences in performance of the two groups. As suggested by other researchers (Gilabert 2004; Skehan 1996b; Robinson 2002, among others), much more research is needed in the area of individual differences. Future research should incorporate information about learners' affective and ability variables in order to achieve a wider picture of what goes on with L2 performance. It would be particularly interesting to compare how each approach, task-based and task-supported, affects learners' motivation. Feedback from learners after the sessions showed that there were learners who enjoyed working with pedagogical tasks and found it more motivating than doing grammar exercises. However, there were learners who asked for traditional focus-on-forms activities and explicit metalinguistic explanation in order to understand how the language works. This means that various language teaching approaches are justified and each approach works better or worse for different learners.

REFERENCES

1. Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259–302). Honolulu: University of Hawai'i Press.
2. Alegría de la Colina, A., & García Mayo, M. P. (2007). Attention to form across collaborative tasks by low-proficiency learners in an EFL setting. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 91–116). Clevedon: Multilingual Matters.
3. Alexander, L. G. (1967). *First Things First*. London: Longman.
4. Alexander, L. G. (1975). Grammatical inventory of T-L English. In J. A. van Ek (Ed.), *The Threshold Level*. Strasbourg: Council of Europe.
5. Allen, L. Q. (2000). Form-meaning connections and the French causative: An experiment in processing instruction. *Studies in Second Language Acquisition*, 22, 69–84.
6. Amiantova, I. E., Bitekhtina, G. A., Vsevolodova, M. V., & Klobukova, L. P. (2001). Funktsionalno-kommunikativnaja lingvodidakticheskaja model' jazyka kak sostavljajuschaja sovremennoj lingvisticheskoj paradigmy. *Vestnik Moskovskogo universiteta*, Serija 9. Filologija, №6, s. 124–148. Амиантова И.Э., Битехтина Г.А., Всеволодова М.В., Клобукова Л.П. (2001). Функционально-коммуникативная лингводидактическая модель языка как составляющая современной лингвистической парадигмы // *Вестник Московского университета*, Серия 9. Филология, №6, с. 124–148.
7. Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, 51(4), 355–365.
8. Antonova, I. A., & Khristiansen, I. A. (2003). O metodicheskikh printsypakh sozdanija professionalno-orientirovannogo posobia po russkomu jazyku dlja inostrannykh uchashchiksja. V. V. Krasnykh & A. I. Izotov (red.), *Jazyk. Soznanije. Kommunikatsija*. Moskva: Maks Press, Выр. 25, s. 189–196. Антонова И.А., Христиансен И.А. (2003). О методических принципах создания профессионально-ориентированного пособия по русскому языку для иностранных учащихся // В.В. Красных, А.И. Изотов (ред.) *Язык. Сознание. Коммуникация*. Москва: МАКС Пресс, Вып. 25, с. 189–196.
9. Arnaudet, M., & Barrett, M. (1984). *Approaches to academic reading and writing*. Englewood Cliffs, NJ: Prentice Hall.
10. Asher, J. J. (1969). The total physical response approach to second language learning. *The Modern Language Journal*, 53(1), 3–17.
11. Baralt, M., Gilabert, R., & Robinson, P. (2014). *Task sequencing and instructed second language learning*. London: Bloomsbury Publishing.
12. Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly*, 26(2), 390–395.
13. Bardovi-Harling, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, 11, 17–34.

14. Beljaev, B. V. (1965). *Oчерки по психологии обучения иностранным языкам*. Moskva: Izdatel'stvo Prosveschenie. Беляев Б.В. (1965). *Очерки по психологии обучения иностранным языкам*. Москва, Издательство «Просвещение».
15. Benati, A. (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(1), 95–127.
16. Beretta, A., & Davies, A. (1985). Evaluation of the Bangalore Project. *ELT Journal*, 39(2), 121–127.
17. Bialystok, E. (1981). The role of conscious strategies in second language proficiency. *Modern Language Journal*, 65(1), 24–35.
18. Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & de Jong, N. (2012). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 30(2), 159–175.
19. Breen, M. P., & Candlin, C. N. (1980). The essentials of a communicative curriculum in language teaching. *Applied Linguistics*, 1(2), 89–102.
20. Breen, M. P. (1984). Process syllabuses for the language classroom. In C. J. Brumfit (Ed.), *General English syllabus design* (pp. 47–60). Oxford: Pergamon Press.
21. Breen, M. P. (1987a). Contemporary paradigms in syllabus design (Parts 1 and 2). *Language Teaching*, 20, 81–92 and 157–174.
22. Breen, M. P. (1987b). Learner contribution to task design. In C. N. Candlin & D. Murphy (Eds.), *Lancaster practical papers in English language education: Vol. 7. Language learning tasks* (pp. 23–46). Englewood Cliffs, NJ: Prentice Hall.
23. Brindley, G. (1994). Task-centred assessment in language learning programs: The promise and the challenge. In N. Bird, P. Falvey, A. Tsui, D. Allison, & A. McNeill (Eds.), *Language and learning* (pp. 73–94). Hong Kong: Institute of Language in Education.
24. Brinton, D. M., Snow, M. A., & Wesche, M. B. (1989). *Content-based second language instruction*. Boston: Heinle & Heinle.
25. Brown, G., Anderson, A., Shillcock, R., & Yule, G. (1984). *Teaching talk: Strategies for production and assessment*. Cambridge: Cambridge University Press.
26. Brumfit, C. J. (1981). Notional syllabuses revisited: A response. *Applied Linguistics*, 2(1), 90–92.
27. Brumfit, C. J. (1984). The Bangalore procedural syllabus. *ELT Journal*, 38(4), 233–241.
28. Bruton, A. (2002). From tasking purposes to purposing tasks. *ELT Journal*, 56(3), 280–288.
29. Bruton, A. (2005). Task-based language teaching: For the state secondary FL classroom? *Language Learning Journal*, 31(1), 55–68.
30. Bruton, A. (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System*, 39, 523–532.
31. Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). Amsterdam: John Benjamins.

32. Burrows, C. (2008). An evaluation of task-based learning (TBL) in the Japanese classroom. *English Today*, 24, 11–16.
33. Bygate, M. (1996). Effects of task repetition: appraising the developing language of learners. In D. Willis & J. Willis (Eds.), *Challenge and change in language teaching*. London: Heinemann.
34. Bygate, M. (1999). Task as context for the framing, reframing and unframing of language. *System*, 27(1), 33–48.
35. Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks. Second language learning, teaching, and testing* (pp. 23–48). Harlow: Longman.
36. Bygate, M., Skehan, P., & Swain, M. (Eds.) (2001). *Researching pedagogic tasks. Second language learning, teaching, and testing*. Harlow: Longman.
37. Cadierno, T. (1995). Formal instruction from a processing perspective: An investigation into the Spanish past tense. *Modern Language Journal*, 79, 179–193.
38. Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.
39. Candlin, C. N. (1984). Syllabus design as a critical process. In C. J. Brumfit (Ed.), *General English syllabus design: ELT documents No. 118* (pp. 29–46). London: Pergamon Press & The British Council.
40. Candlin, C. N. (1987). Towards task-based language learning. In C. N. Candlin & D. Murphy (Eds.), *Lancaster practical papers in English language education: Vol.7. Language learning tasks* (pp. 23–46). Englewood Cliffs, NJ: Prentice Hall.
41. Candlin, C. N. (2001). Afterword: Taking the curriculum to task. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks. Second language learning, teaching, and testing* (pp. 229–243). Harlow: Longman.
42. Carless, D. (2004). Issues in teachers' reinterpretation of a task-based innovation in primary schools. *TESOL Quarterly*, 38(4), 639–662.
43. Carless, D. (2007). The suitability of task-based approaches for secondary schools: Perspectives from Hong Kong. *System*, 35, 595–608.
44. Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, 15, 357–386.
45. Carroll, S., Roberge, Y., & Swain, M. (1992). The role of feedback in second language acquisition: Error correction and morphological generalization. *Applied Linguistics*, 13, 173–198.
46. Castellví, J. & Markina, E. (2015). *Input and instruction in morphologically complex languages: The case of Russian*. Paper given at the Task-based Language Teaching Conference, Leuven, Belgium.
47. Chernyshov, S. I., & Chernyshova, A. V. (2009). *Poekhali-2! Russkij jazyk dlja vzroslykh. Bazovyj kurs*. T.1. Sankt-Peterburg: Zlatoust. Чернышов С.И., Чернышова А.В. (2009). *Поехали-2! Русский язык для взрослых. Базовый курс*. Т.1. Санкт-Петербург: Златоуст.

48. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates.
49. Comer, W. J. (2007). Implementing task-based language teaching from the ground up: Consideration for lesson planning and classroom practice. *Russian Language Journal*, 57, 181–203.
50. Common European Framework of Reference for Languages: Learning, Teaching, Assessment (2001). Council of Europe, Strasbourg: Cambridge University Press.
51. Cook, G. (2000). *Language play, language learning*. Oxford: Oxford University Press.
52. Cook, V. J. (1974). *English topics*. Oxford: Oxford University Press.
53. Coyle, D. (2008). CLIL – a pedagogical approach. In N. Van Deusen-Scholl & N. Hornberger (Eds.), *Encyclopedia of language and education* (pp. 97–111). Springer Science.
54. Crombie, W. (1985). *Relational syllabuses*. Oxford: Oxford University Press.
55. Crookes, G. (1986). *Task classification: A cross-disciplinary review*. Technical report №4. Honolulu: Center for Second Language Classroom Research, Social Science Research Institute, University of Hawaii at Manoa.
56. Crookes, G. (1989). Planning and interlanguage variability. *Studies in Second Language Acquisition*, 11, 367–383.
57. Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Clevedon: Multilingual Matters.
58. Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting*, 5(1), 1–18.
59. Daller, H., & Phelan, D. (2007). What is in a teacher's mind? Teacher ratings of EFL essays and different aspects of lexical richness. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 234–244). Cambridge: Cambridge University Press.
60. Dalton-Puffer, C. (2011). Content-and-Language Integrated Learning: From practice to principles? *Annual Review of Applied Linguistics*, 31, 182–204.
61. De Graaff, R., & Housen, A. (2009). Investigating the effects and effectiveness of L2 instruction. In M. H. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 726–755). Oxford: Blackwell.
62. De Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, 19, 249–276.
63. De la Fuente, M. J. (2006). Classroom L2 vocabulary acquisition: Investigating the role of pedagogical tasks and form-focused instruction. *Language Teaching Research*, 10(3), 263–295.
64. DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, 17(3), 379–410.
65. DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, 19, 167–189.
66. DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on*

- form in classroom second language acquisition* (pp. 42–63). Cambridge: Cambridge University Press.
67. DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22(4), 499–533.
 68. DeKeyser, R. M. (2005). What makes second-language grammar difficult? A review of issues. *Language Learning*, 55(1), 1–25.
 69. DeKeyser, R. M., & Sokalski, K. (1996). The differential role of comprehension and production practice. *Language Learning*, 46, 613–642.
 70. Denissenko, A. (2016). L'adquisició del rus com a llengua estrangera en un context d'instrucció formal a l'aula: Complexitat, correcció i fluïdesa i ús del cas en la producció escrita. Unpublished Ph.D. thesis, University of Pompeu Fabra, Spain.
 71. Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition*, 13, 431–469.
 72. Doughty, C., & Pica, T. (1986). Information gap tasks: An aid to second language acquisition? *TESOL Quarterly*, 20, 305–325.
 73. Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114–138). Cambridge: Cambridge University Press.
 74. Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–262). Cambridge: Cambridge University Press.
 75. Duff, P. (1986). Another look at interlanguage talk: Taking task to task. In R. Day (Ed.), *Talking to learn. Conversation in second language acquisition* (pp. 147–181). Cambridge: Newbury House.
 76. Dulay, H., & Burt, M. (1974). Natural sequences in child second language acquisition. *Language Learning*, 24, 37–53.
 77. Dulay, H., Burt, M., & Krashen, S. (1982). *Language two*. New York: Oxford Press.
 78. Duran, G., & Ramaut, G. (2006). Tasks for absolute beginners and beyond: Developing and sequencing tasks at basic proficiency levels. In K. Van den Branden (Ed.), *Task-based language education: From theory to practice* (pp. 47–75). Cambridge: Cambridge University Press.
 79. Ellis, D. P. (2011). *The role of task complexity in the linguistic complexity of native speaker output*. Unpublished Qualifying Paper. University of Maryland, USA.
 80. Ellis, N. (1994). Implicit and explicit language learning. An overview. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1–32). London: Academic Press.
 81. Ellis, R. (1990). *Instructed second language acquisition: learning in the classroom*. Oxford: Basil Blackwell.
 82. Ellis, R. (1991). Grammar teaching practice or consciousness-raising? In R. Ellis (Ed.), *Second language acquisition and second language pedagogy* (pp. 232–241). Clevedon, Avon: Multilingual Matters.
 83. Ellis, R. (1993). The structural syllabus and second language acquisition. *TESOL Quarterly*, 27(1), 91–113.

84. Ellis, R. (1994). A theory of instructed second language acquisition. In N. Ellis (Ed.), *Implicit and explicit learning of language* (pp. 79–114). London: Academic Press.
85. Ellis, R. (1997). *Second language acquisition*. Oxford: Oxford University Press.
86. Ellis, R. (2000). Task-based research and language pedagogy. *Language Teaching Research*, 4(3), 193–220.
87. Ellis, R. (2001). Investigating form-focused instruction. In R. Ellis (Ed.), *Form-focused instruction and second language learning* (pp. 1–46). Malden: Blackwell Publishers.
88. Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
89. Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, 54(2), 227–275.
90. Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27(2), 141–172.
91. Ellis, R. (2006). The methodology of task-based teaching. *The Asian EFL Journal Quarterly*, 8(3), 19–45.
92. Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, 19(3), 222–246.
93. Ellis, R. (2012). *Language teaching research and language pedagogy*. Malden: Wiley-Blackwell.
94. Ellis, R. (2016). Focus on form: A critical review. *Language Teaching Research*, 20(3), 405–428.
95. Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.
96. Ellis, R., & Takashima, H. (1999). Output enhancement and the acquisition of the past tense. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 173–188). Amsterdam: John Benjamins.
97. Ellis, R., & Yuan, F. (2005). The effects of careful within-task planning on oral and written task performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 167–192). Philadelphia: John Benjamins.
98. Esmantova, T. L. (2011). *Russkyj jazyk: 5 elementov. Uroven' A1 (elementarnyj)*. Sankt-Peterburg: Zlatoust. Эсмантова, Т.Л. (2011). *Русский язык: 5 элементов. Уровень А1 (элементарный)*. Санкт-Петербург: Златоуст.
99. Farley, A. P. (2001). The effects of processing instruction and meaning-based output instruction. *Spanish Applied Linguistics*, 5, 57–94.
100. Felix, S. W. (1981). The effect of formal instruction on second language acquisition. *Language Learning*, 31(1), 87–112.
101. Felix, S. W., & Weigl, W. (1991). Universal Grammar in the classroom: The effects of formal instruction on second language acquisition. *Second Language Research*, 7, 162–180.
102. Field, A. (2009). *Discovering statistics using SPSS: And sex and drugs and Rock'n'roll*. London: Sage publications.

103. Finocchiaro, M., & Brumfit, C. (1983). *The functional-notional approach: From theory to practice*. Oxford: Oxford University Press.
104. Foster, P. (1998). A classroom perspective on the negotiation of meaning. *Applied Linguistics*, 19(1), 1–23.
105. Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299–323.
106. Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354–375.
107. Fotos, S. (1993). Consciousness-raising and noticing through focus on form: Grammar task performance vs. formal instruction. *Applied Linguistics*, 14(4), 385–407.
108. Fotos, S. (1994). Integrating grammar instruction and communicative language use through grammar consciousness-raising tasks. *TESOL Quarterly*, 28, 323–351.
109. Fotos, S., & Ellis, R. (1991). Communicating about grammar: A task-based approach. *TESOL Quarterly*, 25, 608–628.
110. García Mayo, M. P. (Ed.) (2007). *Investigating tasks in formal language learning*. Clevedon: Multilingual Matters.
111. Gascoigne, C. (2006). Explicit input enhancement: Effects on target and non-target aspects of second language acquisition. *Foreign Language Annals*, 39(4), 551–564.
112. Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum Associates.
113. Gass, S. M. (2003). Input and interaction. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 224–255). Oxford: Blackwell Publishing Ltd.
114. Gass, S. M., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review*, 19, 3–17.
115. Gass, S. M., Mackey, A., & Ross-Feldman, L. (2005). Task-based interactions in classroom and laboratory settings. *Language Learning*, 55(4), 575–611.
116. Gattegno, C. (1963). *Teaching foreign languages in schools: The silent way*. UK: Educational Explorers.
117. Geddes, M., & Sturtridge, G. (1979). *Listening links*. London: Heinemann.
118. Gilabert, R., & Castellví, J. (in press). Task and syllabus design for morphologically complex languages. In J. Schwieter & A. Benati (Eds), *The Cambridge handbook of language learning*. Cambridge: Cambridge University Press.
119. Gilabert, R. (2004). *Task complexity and L2 narrative oral production*. Unpublished Ph.D. dissertation, University of Barcelona, Spain.
120. Gilabert, R. (2005). Evaluating the use of multiple sources and methods in needs analysis: A case study of journalists in the Autonomous Community of Catalonia (Spain). In M. H. Long (Ed.), *Second language needs analysis* (pp. 182–199). Cambridge: Cambridge University Press.
121. Gilabert, R. (2013). Teaching English as a foreign language II. Course given at the University of Barcelona, Spain.

122. Gilabert, R., Barón, J., & Llanes, A. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching*, 47, 367–395.
123. González-Lloret, M., & Nielson, B. (2015). Evaluating TBLT: The case of a task-based Spanish program. *Language Teaching Research*, 19(5), 525–549.
124. Goo, J., Granena, G., Novella, M., & Yilmaz, Y. (2009). *Implicit and explicit instruction in L2 learning: Norris and Ortega (2000) revisited and updated*. Paper presented at the Second Language Research Forum, East Lansing, MI.
125. Gor, K., & Jackson, S. (2013). Morphological decomposition and lexical access in a native and second language: A nesting doll effect. *Language and Cognitive Processes*, 28, 1065–1091.
126. *Gosudarstvennyj standart po russkomu jazyku kak inostrannomu. Bazovyj uroven'* (2001). Red. Nakhabina, M. M., & Soboleva, N. I. Sankt-Peterburg: Zlatoust. *Государственный стандарт по русскому языку как иностранному. Базовый уровень* (2001). Ред. Нахабина М.М., Соболева Н.И. Санкт-Петербург: Златоуст.
127. Gregg, K. (1984). Krashen's monitor and Occam's razor. *Applied Linguistics*, 5, 79–100.
128. Griffiths, R. (1991). Language classroom speech rates: A descriptive study. *TESOL Quarterly*, 25(1), 189–194.
129. Halliday, M. A. (1973). *Explorations in the functions of language*. London: Edward Arnold.
130. Halliday, M. A., McIntosh, A., & Stevens, P. (1964). *The Linguistic sciences and language teaching*. London: Longman.
131. Hammerly, H. (1987). The immersion approach: Litmus test of second language acquisition through classroom communication. *Modern Language Journal*, 71, 395–401.
132. Han, Z., Park, E. S., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics*, 29(4), 597–618.
133. Harley, B. (1988). Effects of instruction on SLA: Issues and evidence. *Annual Review of Applied Linguistics*, 9, 165–178.
134. Harley, B. (1989). Functional grammar in French immersion: A classroom experiment. *Applied Linguistics*, 10(3), 331–359.
135. Harley, B. (1992). Patterns of second language development in French immersion. *French Language Studies*, 2, 159–183.
136. Harley, B. (1993). Instructional strategies and SLA in early French immersion. *Studies in Second Language Acquisition*, 15(2), 245–259.
137. Harley, B. (1998). The role of focus-on-form tasks in promoting child L2 acquisition. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 156–176). Cambridge: Cambridge University Press.
138. Harley, B., & Swain, M. (1985). The interlanguage of immersion students and its implications for second language teaching. In A. Davies, C. Criper, & A. P. R. Howatt (Eds.), *Interlanguage* (pp. 291–311). Edinburgh: Edinburgh University Press.

139. Henry, K. (1996). Early L2 writing development: A study of autobiographical essays by university-level students of Russian. *Modern Language Journal*, 80(3), 309–326.
140. Hernández, T. A. (2011). Re-examining the role of explicit instruction and input flood on the acquisition of Spanish discourse markers. *Language Teaching Research*, 15(2), 159–182.
141. Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461–473.
142. Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurements and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA* (pp. 1–20). Amsterdam: John Benjamins.
143. Howatt, A. P. (1984). *A history of English language teaching*. Oxford: Oxford University Press.
144. Hulstijn, J. (1995). Not all grammar rules are equal: Giving grammar instruction its proper place in foreign language teaching. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 359–386). Honolulu: University of Hawai'i Press.
145. Hulstijn, J., & De Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? In J. Hulstijn & R. Schmidt (Eds.), *Consciousness in second language learning* (pp. 97–112). *AILA Review*, Vol. 11.
146. Hunt, K. (1965). *Grammatical structures written at three grade levels*. Champaign, IL: The National Council of Teachers of English.
147. Hymes, D. (1966). Two types of linguistic relativity. In W. Bright (Ed.), *Sociolinguistics* (pp. 114–158). The Hague: Mouton.
148. Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, 4, 51–70.
149. Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language Learning*, 51(3), 401–436.
150. Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study of ESL relativization. *Studies in Second Language Acquisition*, 24, 541–577.
151. Johnson, K. (1979). Communicative approaches and communicative processes. In C. J. Brumfit & K. Johnson (Eds.), *The communicative approach to language teaching*. Oxford: Oxford University Press.
152. Johnson, K. (1982). *Communicative syllabus design and methodology*. Oxford: Pergamon.
153. Johnson, K. (1996). *Language teaching and skill learning*. Oxford: Basil Blackwell.
154. Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. (1995). Does textual enhancement promote noticing? A think-aloud protocol analysis. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 183–216). Honolulu: University of Hawai'i Press.

155. Kellerman, E. (1985). If at first you *do* succeed... In S. M. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 345–353). Rowley, MA: Newbury House.
156. Kempe, V., & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition*, 20, 543–587.
157. Kim, Y. (2006). Effects of input elaboration on vocabulary acquisition through reading by Korean learners of English as a foreign language. *TESOL Quarterly*, 40(2), 341–373.
158. Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 145–164.
159. Kowal, M., & Swain, M. (1997). From semantic to syntactic processing: How can we promote it in the immersion classroom? In M. Swain & R. K. Johnson (Eds.), *Immersion education: International perspectives* (pp. 284–309). Cambridge: Cambridge University Press.
160. Krashen, S. & Terrell, T. (1983). *The Natural approach: Language acquisition in the classroom*. Oxford: Pergamon Press.
161. Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.
162. Krashen, S. (1982). *Principles and practice in second language acquisition*. New York: Pergamon Press.
163. Krashen, S. (1985). *The Input Hypothesis: Issues and implications*. London: Longman.
164. Krashen, S. (1992). Formal grammar instruction: Another educator comments. *TESOL Quarterly*, 26(2), 409–411.
165. Krashen, S. (1993). The effect of formal grammar teaching: Still peripheral. *TESOL Quarterly*, 27(4), 722–725.
166. Krashen, S. (1994). The Input Hypothesis and its rivals. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45–78). London: Academic Press.
167. Kuiken, F., & Vedder, I. (2004). *How does the cognitive complexity of a task affect the linguistic complexity of the output?* Proceedings of the 14th EUROSLA Conference, San Sebastian, Spain.
168. Kuiken, F., & Vedder, I. (2007). Cognitive task complexity and linguistic performance in French L2 writing. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 117–135). Clevedon: Multilingual Matters.
169. Lai, C., Zhao, Y., & Wang, J. (2011). Task-Based Language Teaching in Online Ab Initio Foreign Language Classrooms. *The Modern Language Journal*, 95, Supplementary Issue, 81–103.
170. Lambert, C., & Kormos, J. (2014). Complexity, accuracy, and fluency in task-based L2 research: Toward more developmentally based measures of second language acquisition. *Applied Linguistics*, 35, 1–9.
171. Lapkin, S. (1984). How well do immersion students speak and write French? *Canadian Modern Language Review*, 40(5), 576–585.
172. Lapkin, S., & Swain, M. (1990). French immersion research agenda for the 90s. *Canadian Modern Language Review*, 46(4), 638–674.

173. Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27, 590–619.
174. Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 579–589.
175. Larsen-Freeman, D. (2010). Not so fast: A discussion of L2 morpheme processing and acquisition. *Language Learning*, 60(1), 221–230.
176. Lee, I. L. (2005). *A study of teachers' and students' perceptions of task-based EFL instruction in vocational high school*. Unpublished MA thesis. National Chung Cheng University, Taiwan.
177. Lee, J. F. (2000). *Tasks and communicating in language classrooms*. Boston: McGraw-Hill.
178. Lee, S. K. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning*, 57(1), 87–118.
179. Lee, S. K., & Huang, H. T. (2008). Visual input enhancement and grammar learning: A meta-analytic review, *Studies in Second Language Acquisition*, 30(3), 307–331.
180. Leeman, J., Arteagoitia, I., Fridman, B., & Doughty, C. (1995). Integrating attention to form with meaning: Focus on form in content-based Spanish instruction. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 217–258). Honolulu: University of Hawai'i Press.
181. Legutke, M. & Thomas, H. (1991). *Process and experience in the language classroom*. New York: Longman.
182. Lewis, M. (1996). Implications of a lexical view of language. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 10–16). Oxford: Macmillan Publishers Limited.
183. Li, D. (1998). "It's always more difficult than you plan and imagine": Teachers' perceived difficulties in introducing the communicative approach in South Korea. *TESOL Quarterly*, 32, 677–697.
184. Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60(2), 309–365.
185. Lightbown, P. M., & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching: Effects on second language learning. *Studies in Second Language Acquisition*, 12, 429–448.
186. Littlewood, W. (2007). Communicative and task-based language teaching in East Asian classrooms. *Language Teaching*, 40, 243–249.
187. Long, M. H. (1983). Does second language instruction make a difference? A review of the research. *TESOL Quarterly*, 17, 359–382.
188. Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language teaching. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 77–99). Clevedon: Multilingual Matters.

189. Long, M. H. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (pp.115–141). New York: Harper and Row.
190. Long, M. H. (1989). Task, group, and task-based interactions. *University of Hawaii Working papers in ESL*, 8(2), 1–26.
191. Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, D. Coste, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). Amsterdam: John Benjamins.
192. Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. K. Bhatia (Eds.), *The handbook of second language acquisition* (pp. 413–468). New York: Academic Press.
193. Long, M. H. (2000). Focus on form in task-based language teaching. In R. D. Lambart & E. Shohamy (Eds.), *Language policy and pedagogy* (pp. 181–194). Amsterdam: John Benjamins.
194. Long, M. H. (2005). Methodological issues in learner needs analysis. In M. H. Long (Ed.), *Second language needs analysis* (pp. 19–76). Cambridge: Cambridge University Press.
195. Long, M. H. (2007). Problem-solving and theory change in SLA. In M. H. Long (Ed.), *Problems in SLA* (pp. 21–40). Mahwah, NJ: Erlbaum.
196. Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Malden: Wiley-Blackwell.
197. Long, M. H., & Crookes, G. (1992). Three approaches to task-based language teaching. *TESOL Quarterly*, 26(1), 27–56.
198. Long, M. H., & Crookes, G. (1993). Units of analysis in syllabus design: The case for task. In G. Crookes & S. M. Gass (Eds.), *Tasks in a pedagogical context. Integrating theory and practice* (pp. 9–56). Clevedon: Multilingual Matters.
199. Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15–41). Cambridge: Cambridge University Press.
200. Long, M. H., Gor, K., & Jackson, S. (2012). Linguistic correlates of second language proficiency: Proof of concept with ILR 2–3 in Russian. *Studies in Second Language Acquisition*, 34, 99–126.
201. Long, M. H., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *The Modern Language Journal*, 82(3), 357–371.
202. Long, M. H. & Norris, J. (2000). Task-based teaching and assessment. In M. Byram (Ed.), *Encyclopedia of language teaching* (pp. 597–603). London: Routledge.
203. Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes & S. M. Gass (Eds.), *Tasks and language learning. Integrating theory and practice* (pp. 123–167). Clevedon: Multilingual Matters.
204. Lynch, T. (2001). Seeing what they meant: Transcribing as a route to noticing. *ELT Journal*, 55, 124–132.

205. Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4(3), 221–250.
206. Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 407–452). Oxford: Oxford University Press.
207. Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses and red herrings? *The Modern Language Journal*, 82, 338–356.
208. Mackey, A. (1995). Stepping up the pace: Input, interaction and interlanguage development: An empirical study of questions in ESL. Unpublished Ph.D. dissertation, University of Sydney.
209. Mackey, A. (1999). Input, interaction and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21, 557–587.
210. Mackey, A. (2012). *Input, interaction and corrective feedback in L2 learning*. Oxford: Oxford University Press.
211. Mackey, W. F. (1965). *Language teaching analysis*. London: Longman.
212. Magnani, M., & Artoni, D. (2015). Teaching learnable grammar in Russian as a second language: A syllabus proposal for case. In E. F. Quero Gervilla, B. Barros García, & T. R. Kopylova (Eds.), *Trends in Slavic studies* (pp. 57–70). Moscow: KRASAND.
213. Marsh, D., & Wolff, D. (Eds.) (2007). *Diverse contexts – converging goals: CLIL in Europe*. Frankfurt: Peter Lang.
214. Marsh, H. W. (1998). Pairwise deletion for missing data in structural equation models: Nonpositive definite matrices, parameter estimates, goodness of fit, and adjusted sample sizes. *Structural Equation Modeling: A Multidisciplinary Journal*, 5, 22–36.
215. McDonough, K., & Chaikitmongkol, W. (2007). Teachers' and learners' reactions to a task-based EFL course in Thailand. *TESOL Quarterly*, 41, 107–132.
216. McKay, S. (1980). Towards an integrated syllabus. In K. Croft (Ed.), *Readings in English as a second language* (pp. 72–84). Boston, MA: Little, Brown.
217. Meara, P., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 16(3), 5–19.
218. Mehisto, P., Marsh, D., & Frigols, M. J. (2008). *Uncovering CLIL: Content and language integrated learning in bilingual and multilingual education*. Oxford: Macmillan.
219. Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20, 52–83.
220. Mislevy, R., Steinberg, L., & Almond, R. (2002). Design and analysis in task-based language assessment. *Language Testing*, 19(4), 477–496.
221. Mohan, B. A. (1977). Toward a situational curriculum. In H. D. Brown, C. Yorio, & R. Crymes (Eds.), *On TESOL '77* (pp. 250–257). Washington, DC: TESOL.
222. Munby, J. (1978). *Communicative syllabus design*. Cambridge: Cambridge University Press.

223. Mustafa, Z. (2008). *Teachers' levels of use in the adoption of task-based language teaching in Malaysian classrooms*. Paper presented at the Third International Conference on Language Learning, University Saint Malaysia, Malaysia.
224. Newton, J. (2001). Options for vocabulary learning through communication tasks. *ELT Journal*, 55, 30–37.
225. Newton, J., & Kennedy, G. (1996). Effects of communication tasks on the grammatical relations marked by the second language learners. *System*, 24(3), 309–322.
226. Nielson, K. (2014). Evaluation of an online, task-based Chinese course. In M. González-Lloret & L. Ortega (Eds.), *Technology-mediated TBLT. Researching technology and tasks* (pp. 295–321). Amsterdam: John Benjamins.
227. Nizęgorodcew, A. (2007). *Input for instructed L2 learners: The relevance of relevance*. Clevedon: Multilingual Matters.
228. Nobuyoshi, J., & Ellis, R. (1993). Focused communication tasks and second language acquisition. *ELT journal*, 47, 203–218.
229. Norris, J. M. (2002). Interpretations, intended uses and designs in task-based language assessment. *Language Testing*, 19(4), 337–346.
230. Norris, J. M. (2009). Task-based teaching and testing. In M. H. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 578–594). Cambridge, MA: Blackwell.
231. Norris, J. M. (2014). *How do we assess task-based performance?* Paper given at LARC/CALPER testing and assessment webinar.
232. Norris, J. M. (2016). Current uses for task-based language assessment. *Annual Review of Applied Linguistics*, 36, 230–244.
233. Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50, 417–528.
234. Norris, J. M., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. In R. Ellis (Ed.), *Form-focused instruction and second language learning* (pp. 157–213). Malden: Blackwell Publishers.
235. Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555–578.
236. Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge: Cambridge University Press.
237. Nunan, D. (1993). Task-based syllabus design: Selecting, grading and sequencing task. In G. Crookes & M. Gass (Eds.), *Tasks in a pedagogical context. Integrating theory and practice* (pp. 9–54). Clevedon: Multilingual Matters.
238. Nunan, D. (2004). *Task-based language teaching*. Cambridge: Cambridge University Press.
239. Nunan, D. (2006). Task-based language teaching in the Asia context: Defining 'Task'. *The Asian EFL Journal Quarterly*, 8(3), 12–18.
240. O'Neill, R. (1970). *English in situations*. London: Oxford University Press.
241. Oh, S. (2001). Two types of input modification and EFL reading comprehension: Simplification versus elaboration. *TESOL Quarterly*, 35(1), 69–96.

242. Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21(1), 109–148.
243. Ovsienko, U. G. (2008). Russkyj jazyk dlja nachinauschikh. Moskva: Russky yazyk. Kursy. Овсиенко, Ю.Г. (2008). Русский язык для начинающих. Москва: Русский язык. Курсы.
244. Palmer, H. E., West, M. P., Faucett, L., & Thorndike, E. L. (1936). *Interim report on vocabulary selection for the teaching of English as a foreign language*. London.
245. Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30(4), 590–601.
246. Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, 31(1), 117–134.
247. Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393–419). London: Academic Press.
248. Paulston, C. B. (1981). Notional syllabuses revisited: Some comments. *Applied Linguistics*, 2(1), 93–95.
249. Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74, 525–556.
250. Pica, T. (1985). The selective impact of classroom instruction on second language acquisition. *Applied Linguistics*, 6(3), 214–222.
251. Pica, T., & Doughty, C. (1985). The role of group work in classroom second language acquisition. *Studies in Second Language Acquisition*, 7, 233–248.
252. Pica, T., & Doughty, C. (1988). Variations in classroom interaction as a function of participant pattern and task. In J. Fine (Ed.), *Second language discourse* (pp. 41–55). Norwood, NJ: Ablex.
253. Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*, 11, 63–90.
254. Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communication tasks for second language instruction. In G. Crookes & S. M. Gass (Eds.), *Tasks and language learning. Integrating theory and practice* (pp. 9–34). Clevedon: Multilingual Matters.
255. Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition*, 6, 186–214.
256. Pienemann, M. (1985). Learnability and syllabus construction. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 23–75). Clevedon: Multilingual Matters.
257. Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47, 101–143.
258. Prabhu, N. S. (1984). Procedural syllabuses. In T. E. Read (Ed.), *Trends in language syllabus design* (pp. 272–280). Singapore: Singapore University Press/RELIC.
259. Prabhu, N. S. (1987). *Second language pedagogy*. Oxford: Oxford University Press.

260. Quero Gervilla, A. (2005). Análisis de errores en la adquisición del caso en ruso por hispanohablantes. Dykinson, S.L.
261. Rahimpour, M. (1997). *Task condition, task complexity and variation in L2 discourse*. Unpublished Ph.D. dissertation, University of Queensland, Australia.
262. Reed, G. F. (1968). Skill. In E. A. Lunzer & J. F. Morris (Eds.), *Developments in human learning* (pp. 104–143). New York: Staples.
263. Ribé, R., & Vidal, N. (1993). *Project work: Step by step*. Oxford: Heinemann.
264. Richards, J. C., & Rodgers, T. S. (1986). *Approaches and methods in language teaching: A description and analysis*. Cambridge: Cambridge University Press.
265. Richterich, R. (1980). *Identifying the needs of adults learning a foreign language*. Oxford: Pergamon Press.
266. Roberts, M. A. (1995). Awareness and the efficacy of error correction. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 163–182). Honolulu: University of Hawai'i Press.
267. Robinson, P. (1994). Implicit knowledge, second language acquisition and syllabus construction. *TESOL Quarterly*, 28, 160–166.
268. Robinson, P. (1995a). Task complexity and second language narrative discourse. *Language Learning*, 45(1), 99–140.
269. Robinson, P. (1995b). Attention, memory and the 'noticing' hypothesis. *Language Learning*, 45(2), 283–331.
270. Robinson, P. (1996a). Learning simple and complex second language rules under implicit, incidental, rule-search and instructed conditions. *Studies in Second Language Acquisition*, 18(1), 27–68.
271. Robinson, P. (1996b). *Consciousness, rules and instructed second language acquisition*. New York: Peter Lang.
272. Robinson, P. (1998). State of the art: SLA theory and second language syllabus design. *The Language Teacher*, 22(4), 7–14.
273. Robinson, P. (2001a). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27–57.
274. Robinson, P. (2001b). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287–318). Cambridge: Cambridge University Press.
275. Robinson, P. (2003). The Cognition Hypothesis, task design, and adult task-based language learning. *Second Language Studies*, 21(2), 45–105.
276. Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, 43(1), 1–32.
277. Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7–26). Clevedon: Multilingual Matters.

278. Robinson, P. (2011a). Task-based language learning: A review of issues. *Language Learning*, 61, Issue Supplement s1, 1–36.
279. Robinson, P. (Ed.) (2011b). *Second language task complexity. Researching the Cognition Hypothesis of language learning and performance*. Amsterdam: John Benjamins.
280. Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching*, 45(3), 161–176.
281. Rott, S. (2007). The effect of frequency of input-enhancements on word learning and text comprehension. *Language Learning*, 57(2), 165–199.
282. Rubinstein, G. (1995). On acquisition of Russian cases by American classroom learners. *International Review of Applied Linguistics in Language Teaching*, 33(1), 10–34.
283. Ruiz-Zorrilla Cruzate, M., Kornakov, P., & Castellví Vives, J. (2001). *Curs del lengua russa. Nivel inicial*. Barcelona.
284. Ruso, N. (2007). The influence of task based learning on EFL classrooms. *EFL Journal*, 18, 1–23.
285. Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133–164). Philadelphia, PA: John Benjamins.
286. Rutherford, W. (1987). *Second language grammar: Learning and teaching*. London: Longman.
287. Rutherford, W., & Sharwood Smith, M. (1985). Consciousness raising and Universal Grammar. *Applied Linguistics*, 6, 274–282.
288. Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. London: Palgrave Macmillan.
289. Samuda, V. (2001). Guiding relationships between form and meaning during task performance: The role of the teacher. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks. Second language learning, teaching, and testing* (pp. 119–140). Harlow: Longman.
290. Scherba, L. V. (1974). *Prepodavanie inostrannykh jazykov v srednej shkole. Obschie voprosy metodiki*. Moskva: Izdatel'stvo Vysshaja shkola. Щерба Л.В. (1974). *Преподавание иностранных языков в средней школе. Общие вопросы методики*. Москва: Издательство Высшая школа.
291. Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129–158.
292. Schmidt, R. (1993). Consciousness, learning and interlanguage pragmatics. In G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 21–42). New York: Oxford University Press.
293. Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–64). Honolulu: University of Hawai'i Press.
294. Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

295. Schwartz, B. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition*, 15(2), 147–163.
296. Seedhouse, P. (1999). Task-based interaction. *ELT Journal*, 53(3), 149–156.
297. Seedhouse, P. (2005). “Task” as a research construct. *Language Learning*, 55(3), 533–570.
298. Sharwood Smith, M. (1981). Consciousness raising and the second language learner. *Applied Linguistics*, 2, 159–168.
299. Sharwood Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research*, 7, 118–132.
300. Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, 15(2), 165–179.
301. Shehadeh, A., & Coombe, C. A. (Eds.) (2012). *Task-based language teaching in foreign language contexts: Research and implementation*. Amsterdam: John Benjamins.
302. Shimura, A. (1990). *Testing the availability of universal grammar in second language acquisition: The governing category parameter and the proper antecedent parameter in Japanese as a second language*. Unpublished MA thesis, University of Hawai’i.
303. Shintani, N. (2011). A comparative study of the effects of input-based and production-based instruction on vocabulary acquisition by young EFL learners. *Language Teaching Research*, 15(2), 137–158.
304. Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 140–160). New York: Longman.
305. Skehan, P. (1996a). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17, 38–62.
306. Skehan, P. (1996b). Second language acquisition research and task-based instruction. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 17–30). Oxford: Macmillian Publishers Limited.
307. Skehan, P. (1998a). *A cognitive approach to language learning*. Oxford: Oxford University Press.
308. Skehan, P. (1998b). Task-based instruction. *Annual Review of Applied Linguistics*, 18, 268–286.
309. Skehan, P. (2001). Tasks and language performance assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks. Second language learning, teaching, and testing* (pp. 163–185). Harlow: Longman.
310. Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36, 1–14.
311. Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510–532.
312. Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1(3), 185–211.
313. Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49, 93–120.

314. Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). Cambridge: Cambridge University Press.
315. Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 193–218). Amsterdam: John Benjamins.
316. Skehan, P., & Foster, P. (2012). Complexity, accuracy, fluency and lexis in task-based performance: A synthesis of the Ealing research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 199–220). Amsterdam: John Benjamins.
317. Slimani-Rolls, A. (2005). Rethinking task-based language learning: What we can learn from the learners. *Language Teaching Research*, 9(2), 195–218.
318. Spada, N., & Lightbown, P. (1993). Instruction and the development of questions in the L2 classroom. *Studies in Second Language Acquisition*, 15(2), 205–224.
319. Swain, M. (1985). Communicative competence: Some roles for comprehensible input and comprehensible output in its development. In S. M. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Cambridge: Newbury House.
320. Swain, M. (1991). French immersion and its offshoots: Getting two for one. In B. Freed (Ed.), *Foreign language acquisition: Research and the classroom* (pp. 91–103). Lexington, MA: Heath.
321. Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review*, 50, 158–164.
322. Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in Applied Linguistics* (pp. 125–144). Oxford: Oxford University Press.
323. Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–81). Cambridge: Cambridge University Press.
324. Swain, M. (2000). French immersion research in Canada: Recent contributions to SLA and Applied Linguistics. *Annual Review of Applied Linguistics*, 20, 199–212.
325. Swain, M., & Johnson, R. K. (1997). Immersion education: A category within bilingual education. In M. Swain & R. K. Johnson (Eds.), *Immersion education: International perspectives* (pp. 1–16). Cambridge: Cambridge University Press.
326. Swain, M., & Lapkin, S. (1986). Immersion French in secondary schools: “The goods” and “the bads”. *Contact*, 5(3), 2–9.
327. Swain, M., & Lapkin, S. (1989). Canadian immersion and adult second language teaching: What is the connection? *Modern Language Journal*, 73, 150–159.
328. Swain, M., & Lapkin, S. (2000). Task-based second language learning: The uses of the first language. *Language Teaching Research*, 4(3), 251–274.
329. Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching*

- pedagogic tasks. Second language learning, teaching, and testing* (pp. 99–118). Harlow: Longman.
330. Swales, J. (1990). *Genre analysis*. Cambridge: Cambridge University Press.
 331. Swan, M. (2005). Legislation by hypothesis: The case of task-based instruction. *Applied Linguistics*, 26(3), 376–401.
 332. Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Boston, MA: Pearson Education.
 333. Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description*. Vol. 3. *Grammatical categories and the lexicon* (pp. 57–149). Cambridge: Cambridge University Press.
 334. Tomasello, M., & Herron, C. (1988). Down the garden path: Inducing and correcting overgeneralization errors in the foreign language classroom. *Applied Psycholinguistics*, 9, 237–246.
 335. Tomasello, M., & Herron, C. (1989). Feedback for language transfer errors. *Studies in Second Language Acquisition*, 11, 384–395.
 336. Trahey, M., & White, L. (1993). Positive evidence and pre-emption in the second language classroom. *Studies in Second Language Acquisition*, 15(2), 181–204.
 337. Van den Branden, K. (2006). Introduction: Task-based language teaching in a nutshell. In K. Van den Branden (Ed.), *Task-based language education: From theory to practice* (pp. 1–16). Cambridge: Cambridge University Press.
 338. Van den Branden, K. (2013). Task-based language teaching and learning. In P. Robinson (Ed.), *The Routledge encyclopedia of second language acquisition* (pp. 628–632). New York: Routledge.
 339. Van den Branden, K., Bygate, M., & Norris, J. M. (Eds.) (2009). *Task-based language teaching: A reader*. Amsterdam: John Benjamins.
 340. van Ek, J. A. (1980). *The threshold level English*. Oxford: Pergamon Press.
 341. van Hout, R., & Vermeer, A. (2007). Comparing measures of lexical richness. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 93–115). Cambridge: Cambridge University Press.
 342. VanPatten, B. (1992). Second language acquisition and foreign language teaching. Part 2. *ADFL Bulletin*, 23, 23–27.
 343. VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Norwood, NJ: Ablex.
 344. VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, 52(4), 755–803.
 345. VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, 15(2), 225–243.
 346. VanPatten, B., & Sanz, C. (1995). From input to output: Processing instruction and communicative tasks. In F. 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.

347. Vsevolodova, M. V. (2000). *Teorija funkcionalno-komunikativnogo sintaksisa*. Moskva: Izdatel'stvo Moskovskogo universiteta. Всеволодова М.В. (2000). *Теория функционально-коммуникативного синтаксиса*. Москва: Издательство Московского университета.
348. Vygotskij, L. S. (1966). Myshlenie i reč. *Izbrannye psikhologičeskie issledovanija*. Moskva: Izdatel'stvo Akademija pedagogičeskikh nauk RSFSR. Выготский Л.С. (1966). Мышление и речь // *Избранные психологические исследования*. Москва: Издательство Академия педагогических наук РСФСР.
349. Wajnryb, R. (1990). *Grammar dictation*. Oxford: Oxford University Press.
350. White, J. (1998). Getting the learner's attention. A typographical input enhancement study. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 85–114). Cambridge: Cambridge University Press.
351. White, L. (1991). Adverb placement in second language acquisition: Some positive and negative evidence in the classroom. *Second Language Research*, 7(2), 133–161.
352. White, L., Spada, N., Lightbown, P. M., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics*, 12, 416–432.
353. White, R. V. (1988). *The ELT curriculum. Design, innovation and management*. Oxford: Basil Blackwell.
354. White, R. V., & Arndt, V. (1991). *Process writing*. Harlow: Longman.
355. Widdowson, H. G. (1978a). *Teaching language as communication*. Oxford: Oxford University Press.
356. Widdowson, H. G. (1978b). Notional-functional syllabuses: 1978, part 4. In C. H. Blatchford & J. Schachter (Eds.), *On TESOL '78* (pp. 32–35). Washington, DC: TESOL.
357. Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, 14, 85–106.
358. Wilkins, D. A. (1976). *Notional syllabuses*. Oxford: Oxford University Press.
359. Williams, J., & Evans, J. (1998). What kind of focus and on which forms? In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 139–155). Cambridge: Cambridge University Press.
360. Willis, D. (1990). *The lexical syllabus: A new approach to language teaching*. London: Collins.
361. Willis, D. (2007). *Doing task-based teaching*. Oxford: Oxford University Press.
362. Willis, D., & Willis, J. (1996). Consciousness-raising activities. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 63–76). Oxford: Macmillan Publishers Limited.
363. Willis, J. (1996). *A framework for task-based learning*. Harlow: Longman.
364. Willis, J., & Willis, D. (1988). *The Collins COBUILD English course, Level 1. Teacher's Book*. London: Collins.
365. Wode, H. (1981). Language-acquisitional universals: A unified view of language acquisition. In H. Winitz (Ed.), *Native language and foreign language acquisition* (pp. 218–234). New York: New York Academy of Sciences.

366. Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). Second language development in writing: Measures of fluency, accuracy, and complexity. Honolulu: University of Hawai'i Press.
367. Yalden, J. (1983). *The communicative syllabus: Evolution, design and implementation*. Oxford: Pergamon Press.
368. Yano, Y., Long, M. H., & Ross, S. (1994). The effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning*, 44(2), 189–219.
369. Yuan F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity, and accuracy in L2 monologic oral production. *Applied Linguistics*, 24(1), 1–27.
370. Zhang, Y. (2007). TBLT innovation in primary school English language teaching in mainland China. In K. Van den Branden, K. Van Gorp, & M. Verhelst (Eds.), *Tasks in action: Task-based language education from a classroom-based perspective* (pp. 68–91). Cambridge: Cambridge Scholars Publishing.
371. Zobl, H. (1980). The formal and developmental selectivity of L1 influence on L2 acquisition. *Language Learning*, 30(1), 43–57.

APPENDICES

Appendix A. Background information about participants

TS group (Rus I)

Name	Gender	Age	Course	L1
VeAb	Female	20	Modern Languages and Literature (MLL)	Catalan, Spanish
PaCa	Female	19	MLL	Catalan, Spanish
MeFe	Female	18	MLL	Catalan, Spanish
ElGl	Male	19	MLL	Catalan, Spanish
AiLe	Female	19	Linguistics	Spanish
MaLl	Female	19	MLL	Catalan, Spanish
CaMa	Female	19	MLL	Catalan, Spanish
CarMar	Female	18	MLL	Catalan, Spanish
MaMa	Female	20	MLL	Catalan, Spanish
ClMa	Female	19	MLL	Catalan, Spanish
RoMo	Female	19	MLL	Catalan, Spanish
ErMo	Male	19	MLL	Catalan, Spanish
LaMo	Female	19	MLL	Catalan, Spanish
FrNo	Female	19	MLL	Catalan, Spanish
IrOl	Female	18	MLL	Catalan, Spanish
AnPo	Male	19	MLL	Catalan, Spanish
PaRi	Female	19	MLL	Catalan, Spanish
MiRo	Female	18	MLL	Catalan, Spanish
MaVi	Female	20	MLL	Catalan, Spanish
NaSe	Female	19	Linguistics	Spanish

TS group (Rus II)

Name	Gender	Age	Course	L1
RaCa	Female	19	Linguistics	Catalan, Spanish
JuCo	Female	19	MLL	Catalan, Spanish
GuDu	Male	19	MLL	Catalan, Spanish
ErGu	Female	22	MLL	Italian
MaGu	Female	19	MLL	Spanish
CrMe	Female	25	MLL	Catalan, Spanish
CaNa	Female	21	MLL	Catalan, Spanish
JuPa	Female	20	MLL	Catalan, Spanish
CaPa	Female	23	MLL	Catalan, Spanish
MaPe	Female	20	Spanish and Russian Philology	French
LuRa	Female	19	MLL	Spanish
AnSi	Female	21	MLL	Catalan, Spanish
AnSi	Male	19	MLL	Catalan, Spanish
JeSm	Female	19	MLL	Spanish
LaBa	Female	20	MLL	Catalan, Spanish

TB group (Rus I and Rus II)

Name	Gender	Age	Course	L1
AlCo	Female	22	MLL	Catalan, Spanish
BeGu	Female	19	MLL	Catalan, Spanish
CeCa	Female	19	MLL	Catalan, Spanish
DaPo	Male	19	Linguistics	Catalan, Spanish
FlUm	Female	19	Linguistics	Spanish
IgFa	Male	21	MLL	Spanish
IuOl	Female	19	MLL	Spanish
MaIn	Female	20	MLL	Catalan, Spanish
MiIb	Female	21	MLL	Catalan, Spanish
NeMa	Female	20	MLL	Catalan, Spanish
OiCa	Female	20	MLL	Bask, Spanish
RaRo	Male	25	MLL	Catalan, Spanish
ReSe	Female	22	MLL	Catalan, Spanish
SaMa	Female	22	MLL	Spanish
SaRo	Female	19	MLL	Spanish, French
ClGu	Female	21	MLL	Catalan, Spanish
PaSc	Male	20	Linguistics	Catalan, German
DiRa	Female	20	MLL	Catalan, Spanish
EIOl	Female	19	MLL	Catalan, Spanish

Appendix B. Grammar tests

RUS I

1. Поставьте слова в скобках в нужную форму. Там, где необходимо, напишите предлог.

Put the words in the brackets into the correct form. Write a corresponding preposition if necessary.

1. Он живёт _____ (Барселона).
2. Студенты идут _____ (университет) _____ (урок).
3. На столе два _____ (журнал) и пять _____ (газета).
4. Завтра я встречаюсь _____ (друзья).
5. Она на диете. Она пьёт чай _____ (сахар).
6. Она работает _____ (старый университет) _____ (улица) Гран Вия.
7. В университетском кафе есть меню _____ (студенты) и _____ (преподаватели).
8. Я каждый вечер звоню _____ (мама).
9. Ты знаешь его _____ (брат)?
10. Он ходил в больницу _____ (врач).
11. Я читаю _____ (интересная книга).
12. Вчера они ели салат _____ (помидоры).
13. Мы говорим _____ (политика).
14. Обычно я езжу _____ (работа) _____ (автобус), но сегодня я еду _____ (метро).
15. Банк слева _____ (французский ресторан).
16. У меня нет _____ (брат) или _____ (сестра).
17. У меня день рождения _____ (октябрь).
18. История _____ (Москва) очень интересная.
19. Она хотела быть _____ (актриса).
20. Я должна купить марки _____ (почта).
21. Это фильм _____ (английская принцесса).
22. Я люблю _____ (мои друзья).
23. Это книга _____ (сестра).
24. Мария пишет _____ (коллеги) письмо.
25. У нас нет _____ (фрукты). Купи три _____ (апельсин) и десять _____ (банан).
26. Я еду _____ (бабушка) в деревню.
27. Сколько стоит бутылка _____ (красное вино)?
28. Я ходил в кино _____ (сестра).
29. _____ (улицы) было много людей.
30. Ты любишь ездить _____ (велосипед)?
31. Дети слушают _____ (мама).
32. Я живу _____ (метро).
33. _____ (лето) люди любят отдыхать на море.
34. Преподаватель дал _____ (студенты) задание.
35. Я ем рис _____ (рыба).
36. Это подарок _____ (моя подруга).
37. Я много занимаюсь _____ (спорт).
38. Я читаю книгу _____ (Москва).
39. Мы едем _____ (Мадрид) _____ (машина).
40. Я хочу купить _____ (телефон).
41. Что ты делаешь _____ (суббота)?
42. Мой папа работает _____ (врач).
43. Кто директор _____ (эта школа)?
44. Он думает _____ (дети).
45. Экзамен будет _____ (июнь).
46. Я люблю _____ (белое вино).
47. В университете сто _____ (компьютер), но нет _____ (интернет).
48. Я подарила _____ (брат) телефон.
49. Куда вы идёте? Мы идём _____ (друзья).
50. Ты должен поговорить _____ (директор).

2. Прочитайте диалог и выберите правильный вариант ответа.

Read the dialog and choose the correct answer.

- Что ты будешь делать в _1_?
 -Я буду встречаться с _2_, рассказывать им о _3_, о моих новых _4_ и о тебе. В первые дни, в _5_ и воскресенье, я буду _6_ моей семьёй. Я увижу родителей, _7_ и его _8_...
 -А где живут твои родители?
 -Они живут в Петербурге, но очень далеко от центра. Мой город очень большой. Там живёт пять _9_ человек. От меня до моих родителей нужно ехать _10_ автобусе.
 -Твой брат живёт с ними?
 -Нет, у _11_ есть квартира, недалеко от родителей. В его районе есть всё, что нужно: большой парк, магазины, почта, неплохая библиотека, спортивный центр и два _12_. Когда ты приедешь в _13_, я тебя познакомлю с _14_ и с его _15_. И мы поедем _16_ в гости.
 -Я приготовил _17_ тебя сюрприз. Я тоже скоро приеду в _18_.
 -Ты мне _19_ этом ничего не говори! Вот это сюрприз! А какого числа ты приедешь?
 - _20_.

1. а) Петербург б) Петербурге в) Петербурга	6. а) с б) у в) без	11. а) его б) него в) ему	16. а) они б) о них в) к ним
2. а) друзья б) друзьям в) друзьями	7. а) брата б) брату в) брат	12. а) кинотеатр б) кинотеатра в) кинотеатров	17. а) для б) с в) о
3. а) Барселоне б) Барселону в) Барселона	8. а) жена б) жену в) жене	13. а) Петербург б) Петербурге в) Петербурга	18. а) Россия б) Россию в) России
4. а) друзья б) друзьях в) друзьями	9. а) миллион б) миллионы в) миллионов	14. а) брата б) брату в) братом	19. а) в б) для в) об
5. а) суббота б) субботе в) субботу	10. а) на б) в в) с	15. а) женой б) жене в) жену	20. а) Десятое июня б) Десятого июня в) Десять июнь

3. В некоторых предложениях есть ошибки. Найдите и исправьте их. Если предложение правильное, напишите рядом знак ✓. Если в предложении есть ошибка, напишите рядом правильный вариант.
Some sentences contain mistakes. Find and correct them. If a sentence is correct, write the symbol ✓ next to it. If there is a mistake, write the correct version.

1. Я еду к друг в Париж.	
2. Она работает на почте.	
3. Елена нравится Москва.	
4. Я звонил тебя вчера.	
5. У меня нет собака.	
6. Его маму зовут Мария.	
7. Маша смотрит программу о русских футболистах.	
8. Мой папа работает врач.	
9. Этот подарок для тебя.	
10. Ты должен есть много фруктов.	
11. Каникулы будут в август.	
12. Люди не могут жить без интернета.	
13. Я не люблю салат с майонез.	
14. Театр слева парк.	
15. Он написал письмо Лауре.	
16. Летом они поедут в Мадриде.	
17. Анна любит заниматься спортом.	
18. Карлосу 12 лет.	
19. Мы любим смотреть футбол на стадион.	
20. Мой дом напротив школы.	

RUS II

1. Поставьте слова в скобках в нужную форму. Там, где необходимо, напишите предлог.

Put the words in the brackets into the correct form. Write a corresponding preposition if necessary.

1. Магазин рядом _____ (итальянский ресторан).
2. Я живу _____ (парк). Я вижу его из окна.
3. Мы едем _____ (Лондон) _____ (машина).
4. _____ (диван) висит картина.
5. Ты должен прийти _____ (банк), а потом повернуть налево.
6. _____ (дом) есть красивый сад.
7. Когда я проходила _____ (магазин), я увидела в витрине красивое платье.
8. Обычно я ухожу _____ (офис) в 18 часов и потом иду в спортзал.
9. Барселона _____ (Таррагона) и _____ (Жирона).
10. Давай пройдем _____ (парк), так быстрее.

Приходить – прийти, приезжать – приехать; уходить – уйти, уезжать – уехать

11. Я вчера _____ в университет в 12 часов.
12. Я люблю, когда ко мне домой _____ друзья.
13. Вчера я была дома одна, потому что мои родители _____ в Москву.
14. Вечеринка закончилась, и все _____ домой.
15. Почему Вы вчера так поздно _____ на работу?
16. Где Лаура? – Она уже _____.
17. Я живу в Барселоне. Мои родители живут в Мадриде. Вчера моя мама _____ ко мне в Барселону. На следующей неделе _____ папа.
18. Каждые выходные мы _____ из города.

Входить – войти, выходить – выйти, доходить – дойти, проходить – пройти, заходить – зайти, переходить – перейти

19. Где Игорь? – Он _____ покурить.
20. Когда ты _____ до университета, слева ты увидишь бар. Мы будем там тебя ждать.
21. Она каждый день _____ в кафе выпить кофе.
22. Он _____ мимо её дома и увидел её машину.
23. Мне было очень плохо, и я не мог _____ до кровати.
24. Мы _____ улицу и повернули направо.
25. Аня _____ в комнату и закрыла дверь.
26. После работы она всегда _____ в магазин купить хлеба.
27. Ты никогда не должен _____ дорогу на красный свет.
28. Сегодня очень плохая погода, я не буду _____ на улицу.
29. Тебе нужно _____ через площадь Каталонии, и ты увидишь Рамблу.
30. Что вы обычно делаете, когда _____ в квартиру?

2. Выберите правильный вариант ответа.

Choose the correct answer.

1. Машина ... из гаража.	(А) выходит (Б) выезжает (В) въезжает
2. Ты должен выйти ... метро и повернуть налево.	(А) в (Б) до (В) из
3. Когда вы вчера ... домой?	(А) прошли (Б) вышли (В) пришли
4. Мы ... из автобуса и дальше пошли пешком.	(А) вышли (Б) вошли (В) зашли
5. Книжный магазин рядом ...	(А) с университетом (Б) университета

	(B) университет
6. До парка можно ... пешком за полчаса.	(A) прийти (B) войти (B) прийти
7. – Станция метро далеко? – Нет, вы должны ... улицу и увидите её.	(A) прийти (B) прийти (B) перейти
8. Я открыл дверь и ... в банк.	(A) вышел (B) вошёл (B) перешёл
9. Мы сидим в кафе ... твоего дома.	(A) перед (B) напротив (B) в центре
10. На улице было мало машин, и мы быстро ... до аэропорта.	(A) приехали (B) дошли (B) доехали
11. По дороге домой я ... в книжный магазин.	(A) пошла (B) зашла (B) пришла
12. ... Эрмитажа можно пройти пешком, он недалеко.	(A) мимо (B) в (B) до
13. Вчера в наш город ... известный артист.	(A) пришёл (B) приехал (B) проехал
14. Скоро лето. Родители ... в деревню, и я буду один дома.	(A) уйдут (B) уедут (B) уехали
15. Я каждый день прохожу ... этого парка	(A) до (B) мимо (B) через

3. В некоторых предложениях есть ошибки. Найдите и исправьте их. Если предложение правильное, напишите рядом знак ✓. Если в предложении есть ошибка, напишите рядом правильный вариант.
Some sentences contain mistakes. Find and correct them. If a sentence is correct, write the symbol ✓ next to it. If there is a mistake, write the correct version.

1. Он перешёл улицу и вошёл в бар.	
2. Она всегда пришла на работу в 9.	
3. Ко мне приехали друзья из Мадрида.	
4. Перед уроком я зашла в кафе.	
5. Я уже прошла до магазина.	
6. Кошка любит сидеть под столе.	
7. Я ушёл из офиса покурить.	
8. Концерт закончился, и все ушли домой.	
9. Вы не знаете, как приехать до ресторана «Прага»?	
10. Когда я прохожу мимо этого ресторана, я всегда читаю меню.	
11. Преподаватель вошёл в аудиторию и начал урок.	
12. Гитара стоит за дивана.	
13. Мы вышли из метро и увидели музей.	
14. Мой офис около от метро «Университет».	
15. В центре комнаты стоит большой стол.	

Appendix C. Control tasks

Written task RUS I

Вы получили письмо от своего русского друга. Прочитайте письмо и напишите ответ (в письме ответьте на ВСЕ вопросы друга).

You received a letter from your Russian friend. Read it and write the answer (you should answer all your friend's questions).

Привет!

Как твои дела? У меня всё хорошо. У меня были зимние каникулы две недели назад. Я отдыхал 10 дней. А у тебя были зимние каникулы? Сколько дней? Что ты делал? Ты ездил отдыхать? Куда и с кем ты ездил? Что ты видел? Где ты жил? Что тебе понравилось в твоей поездке?

Я был в Москве. Всё было очень хорошо. Было много снега, температура была около –15С°. Я люблю эту погоду, потому что я могу кататься на лыжах. А что ты любишь делать зимой?

Я много сделал во время каникул. Я посмотрел фильмы «Шрек» и «Тринадцать друзей Оушена». Ты видел эти фильмы? Ты видел какой-нибудь интересный фильм? О чём он?

В свободное время я слушал музыку. Я был на концерте «Coldplay». Ты знаешь эту группу? Тебе нравится их музыка? Какая твоя любимая группа?

Жду твой ответ.

Саша

Hi,

How are you? I'm fine. I had my winter holidays two weeks ago. I had 10 days of holidays. What about you? Have you had winter holidays? How many days? What did you do? Did you go anywhere? Where and with whom did you go? What did you see? Where did you stay? What did you like on your trip?

I was in Moscow. It was great. There was a lot of snow, and the temperature was about –15C°. I like this weather because I can ski. And what do you like to do in winter?

I did a lot of things during my holidays. I saw the films "Shrek" and "Ocean's thirteen". Have you seen them? Have you seen any interesting film? What is it about?

In my free time I listened to music. I was at the "Coldplay" concert. Do you know this band? Do you like their music? What is your favorite band?

Hope to hear from you soon.

Sasha

Written task RUS II

Вы хотите пригласить русских друзей к себе домой на день рождения. Они никогда не были у вас дома и не знают, как доехать. **Напишите им емейл**, где вы должны сказать:

- (1) в какой день и во сколько они должны прийти,
- (2) ваш адрес (улица, дом), станция метро,
- (3) как дойти от метро до вашего дома,
- (4) что вы хотите приготовить на ужин,
- (5) куда вы хотите пойти потом.

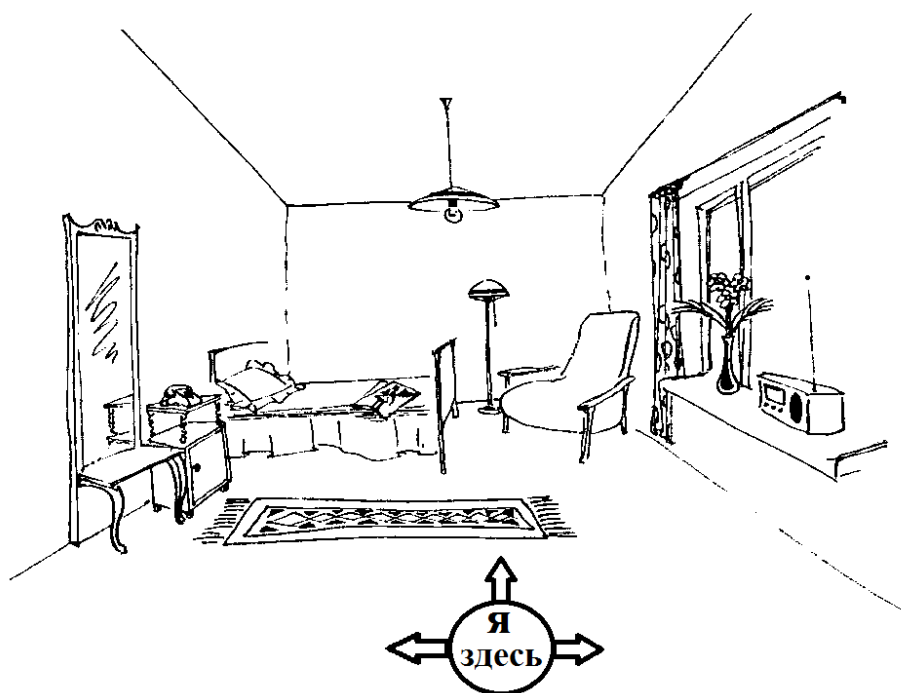
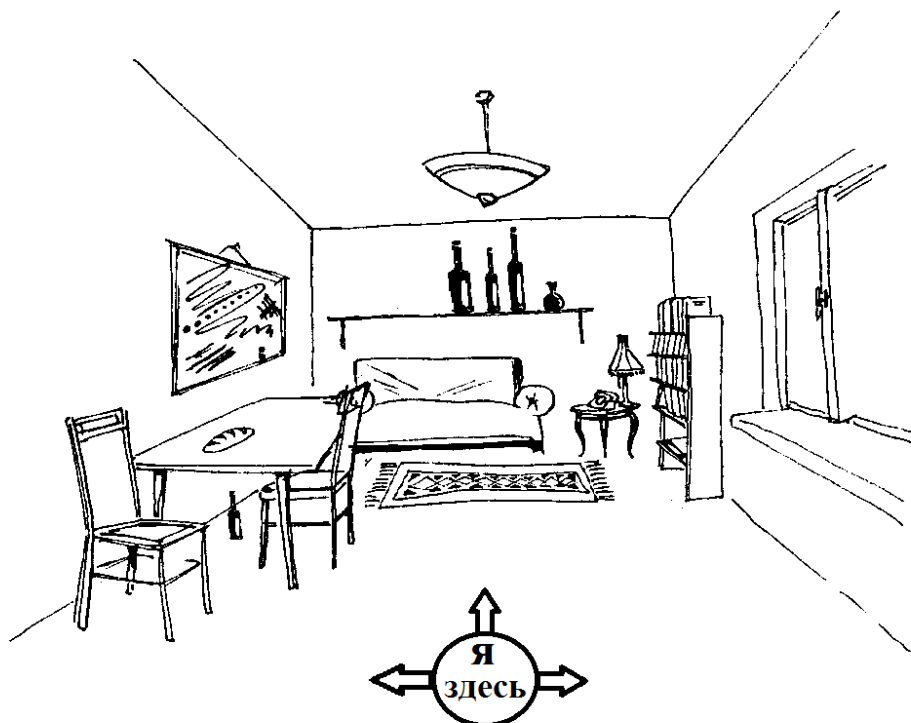
You want to invite your Russian friends to your birthday party. They have never been at your house and they don't know how to get there. **Write them an email** where you should explain:

- (1) what day and what time they should come
- (2) your home address (street, house) and the nearest metro station
- (3) how to get from the metro to your home
- (4) what you want to cook for the dinner
- (5) where you want to go to after the dinner.

Oral task 1. Room Description

Вы сняли двухкомнатную квартиру. Ваши родители её не видели и спрашивают вас, что есть в комнатах, какие они. Опишите две комнаты.

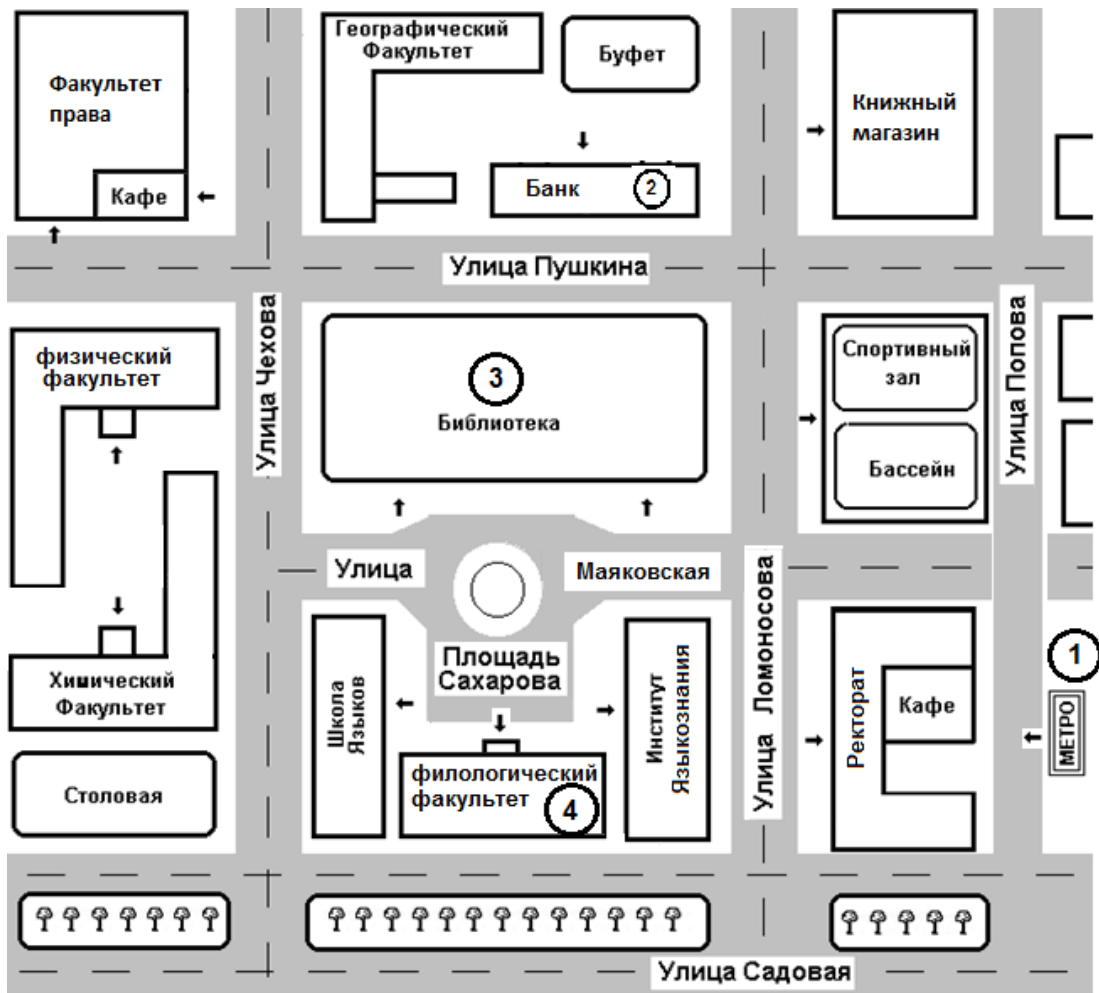
You have rented a flat with two rooms. Your parents have not seen it, and they are asking you what it is like and what you have there. Describe both rooms.



Oral task 2. Map Task

Вы учитесь в университете в Москве. Вчера в вашу группу пришёл новый студент. Сегодня вы встречаете его около метро. Ему нужно зайти в банк, в библиотеку, а потом на филологический факультет. Он не знает дороги и просит вас помочь. **Объясните, как идти.**

You study at university in Moscow. Yesterday a new student joined your group. Today you meet him near the metro station. He has to go to the bank, to the library and then to the Philological Faculty. He doesn't know how to get there and asks you for help. Explain to him how to go.



Appendix D. Descriptive statistics on CAF and grammar scores: Means, medians, standard deviations, skewness, and kurtosis

Written task and grammar tests (RUS I)

Dependent variable		Time	TS_Rus I						TB_Rus I					
			<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>Sk</i>	<i>K</i>	<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>Sk</i>	<i>K</i>
Accuracy	Errors per Words	Pre	17	23.20 (2.44)	18.54	10.06	.50 (.55)	-.66 (1.06)	14	22.87 (2.55)	23.11	9.54	.82 (.60)	1.14 (1.15)
		Post	17	14.81 (1.49)	14.17	6.15	.58 (.55)	-.39 (1.06)	14	15.12 (1.78)	13.73	6.72	.52 (.60)	-.93 (1.15)
		Delayed post	17	16.28 (1.99)	16.90	8.21	.39 (.55)	-.79 (1.06)	14	18.85 (2.04)	14.93	7.65	.72 (.60)	-.96 (1.15)
	TLU of Prepositions	Pre	16	81.19 (4.87)	87.30	19.48	-1.19 (.56)	1.75 (1.09)	13	86.76 (2.74)	86.36	9.87	-.14 (.62)	-.85 (1.19)
		Post	16	89.57 (2.17)	90.83	8.68	-.67 (.56)	.44 (1.09)	13	93.16 (1.74)	92.86	6.26	-.76 (.62)	.15 (1.19)
		Delayed post	16	91.34 (1.98)	90.83	7.91	-.16 (.56)	-1.49(1.09)	13	93.62 (2.60)	100.00	9.38	-1.31 (.62)	.97 (1.19)
	TLU of Case Forms	Pre	17	82.90 (2.23)	85.00	9.20	-.57 (.55)	-.01 (1.06)	14	84.24 (2.51)	85.80	9.39	-1.60 (.60)	3.30 (1.15)
		Post	17	87.07 (1.47)	87.04	6.07	-.38 (.55)	-.32 (1.06)	14	90.62 (1.51)	91.75	5.64	-.23 (.60)	-.90 (1.15)
		Delayed post	17	84.09 (2.45)	86.67	10.10	-.54 (.55)	-.87 (1.06)	14	82.53 (2.71)	86.36	10.14	-.73 (.60)	-.95 (1.15)
Complexity	Mean Length of Clause	Pre	17	5.74 (.19)	5.62	.79	.79 (.55)	.01 (1.06)	14	5.94 (.18)	5.97	.69	-.26 (.60)	-1.31 (1.15)
		Post	17	5.30 (.11)	5.34	.46	-.04 (.55)	.19 (1.06)	14	4.94 (.16)	4.84	.60	.80 (.60)	.98 (1.15)
		Delayed post	17	5.68 (.13)	5.67	.54	.68 (.55)	-.46 (1.06)	14	5.47 (.16)	5.67	.60	-1.16 (.60)	.43 (1.15)
	Clauses per T-Unit	Pre	17	1.11 (.03)	1.09	.12	1.13 (.55)	1.00 (1.06)	14	1.09 (.02)	1.07	.09	.76 (.60)	-.36 (1.15)
		Post	17	1.11 (.02)	1.09	.08	.35 (.55)	-.77 (1.06)	14	1.13 (.03)	1.13	.10	.62 (.60)	.85 (1.15)
		Delayed post	17	1.08 (.01)	1.07	.06	.01 (.55)	-1.01(1.06)	14	1.07 (.03)	1.03	.10	1.56 (.60)	1.86 (1.15)
Grammar	Fill in the Blanks	Pre	19	71.28 (3.35)	74.76	14.60	-.67 (.52)	-.06 (1.01)	16	63.65 (4.69)	63.59	18.78	-.15 (.56)	-1.67 (1.09)
		Post	19	79.56 (3.11)	82.52	13.57	-.43 (.52)	-.65 (1.01)	16	74.39 (5.71)	84.95	22.82	-1.16 (.56)	.95 (1.09)
		Delayed post	19	68.32 (3.22)	66.99	14.02	-.14 (.52)	-.19 (1.01)	16	59.69 (5.32)	60.00	21.27	.10 (.56)	-1.61 (1.09)
	Multiple Choice	Pre	19	68.68 (2.98)	70.00	13.00	-.19 (.52)	-.77 (1.01)	16	71.25 (4.58)	75.00	18.30	-.39 (.56)	-.46 (1.09)
		Post	19	79.21 (3.69)	85.00	16.09	-1.20 (.52)	1.90 (1.01)	16	80.94 (4.06)	87.50	16.25	-.68 (.56)	-.89 (1.09)
		Delayed post	19	72.89 (4.38)	75.00	19.10	-.39 (.52)	-.97 (1.01)	16	74.22 (3.74)	71.87	14.94	.13 (.56)	-.98 (1.09)
	Grammaticality Judgment	Pre	19	68.16 (2.84)	70.00	12.38	-.31 (.52)	-.34 (1.01)	16	68.12 (4.76)	70.00	19.05	-.65 (.56)	.06 (1.09)
		Post	19	73.16 (3.55)	75.00	15.47	-.03 (.52)	-.87 (1.01)	16	75.62 (4.35)	80.00	17.40	-.66 (.56)	-.74 (1.09)
		Delayed post	19	69.47 (3.84)	77.54	16.74	-.37 (.52)	-1.08(1.01)	16	60.55 (4.76)	62.50	19.05	-.27 (.56)	-.60 (1.09)

M= Mean; Mdn= Median; SD= Standard Deviation; Sk= Skewness; K= Kurtosis. Standard Error is in brackets.

Written task and grammar tests (RUS II)

Dependent variable			TS_Rus II					TB_Rus II						
			<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>Sk</i>	<i>K</i>	<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>Sk</i>	<i>K</i>
Accuracy	Errors per Words	Pre	14	19.55 (2.21)	21.73	8.26	.56 (.60)	.83 (1.15)	16	22.35 (1.53)	23.00	6.14	-.10 (.56)	-.86 (1.09)
		Post	14	18.57 (1.58)	15.71	5.90	.46 (.60)	-.04 (1.15)	16	15.94 (1.98)	14.42	7.95	1.19 (.56)	1.30 (1.09)
		Delayed post	14	17.63 (2.03)	17.01	7.61	.96 (.60)	.68 (1.15)	16	15.50 (1.89)	15.93	7.56	.32 (.56)	-.49 (1.09)
	Total number of Verbs of Motion	Pre	13	1.54 (.24)	1.00	.88	.30 (.62)	-.33 (1.19)	17	1.53 (.35)	1.00	1.46	.82 (.55)	.15 (1.06)
		Post	13	4.38 (.60)	4.00	2.18	.61 (.62)	-.89 (1.19)	17	5.53 (.45)	6.00	1.87	-.18 (.55)	-.38 (1.06)
		Delayed post	13	5.00 (.64)	5.00	2.31	.09 (.62)	-.94 (1.19)	17	6.12 (.50)	6.00	2.09	.57 (.55)	-.29 (1.06)
	TLU of Verbs of Motion	Pre	14	1.21 (.28)	1.00	1.05	1.36 (.60)	2.97 (1.15)	17	.71 (.32)	.00	1.31	2.50 (.55)	7.06 (1.06)
		Post	14	3.00 (.55)	2.50	2.07	.48 (.60)	-.62 (1.15)	17	4.12 (.516)	4.00	2.09	.01 (.55)	-.36 (1.06)
		Delayed post	14	3.14 (.61)	2.50	2.28	1.38 (.60)	2.39 (1.15)	17	4.65 (.53)	5.00	2.18	.35 (.55)	-.27 (1.06)
Complexity	Mean Length of Clause	Pre	13	6.20 (.20)	6.23	.74	.13 (.62)	-1.21(1.19)	16	6.34 (.24)	6.40	.95	.23 (.56)	-.95 (1.09)
		Post	13	6.04 (.26)	5.69	.96	.73 (.62)	-.17 (1.19)	16	6.43 (.19)	6.47	.78	-.48 (.56)	-.52 (1.09)
		Delayed post	13	5.87 (.21)	5.89	.77	.59 (.62)	.26 (1.19)	16	6.17 (.21)	6.04	.86	.69 (.56)	-.16 (1.09)
	Clauses per T-Unit	Pre	14	1.18 (.03)	1.14	.12	.65 (.60)	-.67 (1.15)	17	1.19 (.04)	1.14	.17	.72 (.55)	-.38 (1.06)
		Post	14	1.17 (.04)	1.19	.14	1.11 (.60)	2.53 (1.15)	17	1.23 (.05)	1.15	.20	.85 (.55)	-.26 (1.06)
		Delayed post	14	1.28 (.05)	1.23	.20	.69 (.60)	-.33 (1.15)	17	1.30 (.04)	1.23	.17	.16 (.55)	-1.73 (1.06)
	Guiraud's Index	Pre	14	2.46 (.10)	2.48	.37	-.14 (.60)	-1.47(1.15)	17	2.02 (.14)	2.00	.60	-.08 (.55)	1.50 (1.06)
		Post	14	2.92 (.09)	2.84	.34	.88 (.60)	-.16 (1.15)	17	3.08 (.12)	3.10	.51	.53 (.55)	.61 (1.06)
		Delayed post	14	3.05 (.13)	3.18	.48	-.54 (.60)	-1.13(1.14)	17	3.07 (.09)	3.00	.36	.36 (.55)	1.15 (1.06)
Grammar	Fill in the Blanks	Pre	14	38.25 (2.63)	37.78	9.83	-.28 (.60)	-1.18(1.15)	19	32.63 (3.47)	37.78	15.14	-.39 (.52)	-.44 (1.01)
		Post	14	59.68 (4.76)	60.00	17.82	-.67 (.60)	-.15 (1.15)	19	61.29 (4.16)	66.67	18.12	-.74 (.52)	-.46 (1.01)
		Delayed post	14	64.92 (3.94)	68.89	14.74	-.47 (.60)	-.58 (1.15)	19	72.51 (4.38)	77.78	19.09	-.85 (.52)	.28 (1.01)
	Multiple Choice	Pre	14	56.67 (4.83)	63.33	18.07	-.49 (.60)	-1.41(1.15)	16	37.08 (3.16)	40.00	12.64	-.21 (.56)	-.31 (1.09)
		Post	14	78.09 (3.00)	80.00	11.22	-.60 (.60)	.29 (1.15)	16	85.83 (2.97)	86.67	11.89	-.51 (.56)	-.28 (1.09)
		Delayed post	14	80.48 (2.37)	80.00	8.85	-.15 (.60)	-1.03(1.15)	16	89.17 (2.27)	90.00	9.07	-.42 (.56)	-.81 (1.09)
	Grammaticality Judgment	Pre	14	45.24 (5.19)	50.00	19.42	-1.15 (.60)	1.34 (1.15)	16	29.58 (7.55)	26.67	30.21	.09 (.56)	-2.13 (1.09)
		Post	14	63.81 (3.47)	66.67	13.00	-.31 (.60)	-.98 (1.15)	16	71.25 (2.90)	73.33	11.60	-.14 (.56)	.28 (1.09)
		Delayed post	14	69.05 (4.11)	66.67	15.38	.54 (.60)	-.49 (1.15)	16	81.67 (3.19)	83.33	12.76	-.21 (.56)	-.55 (1.09)

Oral task 1 (RUS II)

Dependent variable			TS_Rus II						TB_Rus II					
			<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>Sk</i>	<i>K</i>	<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>Sk</i>	<i>K</i>
Fluency	Speech Rate	Pre	12	49.42 (3.56)	47.20	12.35	.76 (.64)	-.19 (1.23)	17	45.24 (2.90)	45.98	11.94	1.06 (.55)	1.69 (1.06)
		Post	12	58.33 (2.44)	55.71	8.45	1.22 (.64)	1.14 (1.23)	17	54.22 (3.22)	52.50	13.28	-.68 (.55)	1.41 (1.06)
		Delayed post	12	62.21 (2.91)	65.40	10.10	-.83 (.64)	.58 (1.23)	17	57.09 (3.29)	56.83	13.55	-.70 (.55)	1.86 (1.06)
Accuracy	Errors per Words	Pre	13	21.95 (2.73)	21.43	9.83	-.05 (.61)	-.97 (1.19)	18	24.31 (1.24)	24.11	5.28	-.35 (.53)	-.49 (1.04)
		Post	13	23.57 (2.45)	23.93	8.83	.39 (.61)	-1.16 (1.19)	18	25.77 (2.61)	24.18	11.06	.96 (.53)	.50 (1.04)
		Delayed post	13	20.89 (2.42)	21.78	8.72	-.18 (.61)	-.47 (1.19)	18	21.70 (2.30)	20.79	9.76	.23 (.53)	-1.24 (1.04)
	TLU of Prepositions	Pre	13	69.51 (5.20)	72.73	18.77	-.42 (.61)	-.59 (1.19)	16	71.13 (4.45)	66.82	17.82	.16 (.56)	-.81 (1.09)
		Post	13	69.23 (6.04)	72.73	21.77	-.45 (.61)	-.88 (1.19)	16	72.86 (3.60)	77.75	14.38	-1.02 (.56)	1.17 (1.09)
		Delayed post	13	72.13 (4.09)	73.68	14.76	-.04 (.61)	-1.40 (1.19)	16	78.31 (3.46)	82.73	13.83	-1.06 (.56)	.50 (1.09)
	TLU of Case Forms	Pre	13	58.25 (5.67)	57.14	20.43	.28 (.61)	.53 (1.19)	16	51.01 (4.76)	55.90	19.03	-.91 (.56)	-.23 (1.09)
		Post	13	49.48 (4.09)	50.00	14.74	.34 (.61)	-.46 (1.19)	16	53.45 (4.50)	51.92	18.00	-.09 (.56)	-1.18 (1.09)
		Delayed post	13	60.05 (4.54)	56.25	16.36	-.31 (.61)	-.05 (1.19)	16	62.61 (5.37)	64.29	21.46	-.87 (.56)	.41 (1.09)
Complexity	Mean Length of Clause	Pre	13	5.51 (.17)	5.44	.61	.82 (.61)	-.28 (1.19)	17	5.51 (.19)	5.36	.79	.72 (.55)	.63 (1.06)
		Post	13	5.52 (.16)	5.40	.58	.14 (.61)	-1.03 (1.19)	17	5.41 (.13)	5.40	.53	.58 (.55)	.57 (1.06)
		Delayed post	13	5.27 (.07)	5.30	.25	.31 (.61)	-.36 (1.19)	17	5.45 (.11)	5.41	.44	.57 (.55)	-.61 (1.06)
	Guiraud's Index	Pre	13	1.51 (.12)	1.60	.44	-.29 (.61)	-.88 (1.19)	18	1.44 (.09)	1.50	.37	-.64 (.53)	.30 (1.04)
		Post	13	1.89 (.06)	1.87	.23	.34 (.61)	-.14 (1.19)	18	1.68 (.09)	1.63	.39	.06 (.53)	.77 (1.04)
		Delayed post	13	1.94 (.07)	2.06	.27	-.18 (.61)	-1.65 (1.19)	18	1.82 (.08)	1.72	.37	-.26 (.53)	.53 (1.04)

Oral task 2 (RUS II)

Dependent variable			TS_Rus II						TB_Rus III					
			<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>Sk</i>	<i>K</i>	<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>Sk</i>	<i>K</i>
Fluency	Speech Rate	Pre	12	66.36 (4.91)	62.42	17.01	.39 (.64)	-.64 (1.23)	18	65.41 (4.88)	65.01	20.72	.33 (.53)	-.55 (1.04)
		Post	12	75.11 (4.06)	72.91	14.06	1.17 (.64)	2.42 (1.23)	18	70.19 (4.01)	76.47	17.01	-.60 (.53)	-.94 (1.04)
		Delayed post	12	79.29 (2.06)	79.87	7.13	.22 (.64)	-.82 (1.23)	18	79.43 (5.73)	85.55	24.33	-.06 (.53)	-1.23 (1.04)
Accuracy	Errors per Words	Pre	13	28.06 (2.28)	26.00	8.24	.92 (.61)	.35 (1.19)	16	28.29 (2.30)	25.61	9.22	.79 (.56)	-.52 (1.09)
		Post	13	17.43 (1.40)	18.18	5.05	-.83 (.61)	.89 (1.19)	16	17.95 (2.17)	17.44	8.70	-.14 (.56)	-1.52 (1.09)
		Delayed post	13	18.13 (2.15)	18.46	7.74	.29 (.61)	-.25 (1.19)	16	13.98 (1.64)	14.26	6.55	.30 (.56)	-.07 (1.09)
	Total number of Verbs of Motion	Pre	13	.15 (.15)	.00	.55	3.60 (.61)	13.00 (1.19)	18	.61 (.24)	.00	1.04	1.62 (.53)	1.51 (1.04)
		Post	13	4.23 (.71)	5.00	2.55	-.66 (.61)	-.85 (1.19)	18	9.17 (.78)	9.50	3.29	.23 (.53)	-.21 (1.04)
		Delayed post	13	5.92 (.62)	5.00	2.25	.32 (.61)	-.34 (1.19)	18	8.28 (.59)	9.00	2.52	.69 (.53)	2.05 (1.04)
	TLU of Verbs of Motion	Pre	13	.08 (.08)	.00	.28	3.60 (.61)	13.00 (1.19)	18	.11 (.07)	.00	.32	2.70 (.53)	5.98 (1.04)
		Post	13	3.23 (.75)	3.00	2.71	.21 (.61)	-1.40 (1.19)	18	6.61 (.72)	6.50	3.07	.06 (.53)	-.28 (1.04)
		Delayed post	13	4.23 (.66)	4.00	2.39	-.15 (.61)	-.58 (1.19)	18	6.17 (.49)	6.50	2.06	-.79 (.53)	1.42 (1.04)
Complexity	Mean Length of Clause	Pre	13	2.58 (.22)	5.65	.79	.96 (.61)	1.94 (1.19)	16	6.37 (.25)	6.16	.99	.90 (.56)	.26 (1.09)
		Post	13	6.14 (.26)	6.30	.95	-.19 (.61)	-1.63 (1.19)	16	7.30 (.52)	6.82	2.08	1.63 (.56)	2.50 (1.09)
		Delayed post	13	6.25 (.31)	5.90	1.13	.89 (.61)	-.13 (1.19)	16	6.72 (.23)	6.57	.94	.26 (.56)	-.63 (1.09)
	Guiraud's Index	Pre	13	1.76 (.13)	1.66	.46	.64 (.61)	-.60 (1.19)	18	1.36 (.11)	1.34	.48	1.31 (.53)	2.85 (1.04)
		Post	13	2.32 (.13)	2.40	.47	-.27 (.61)	.12 (1.19)	18	2.68 (.10)	2.74	.41	-.74 (.53)	1.85 (1.04)
		Delayed post	13	2.60 (.13)	2.62	.48	.20 (.61)	-.60 (1.19)	18	2.79 (.09)	2.72	.39	.07 (.53)	-.87 (1.04)

