Input-based tasks for beginner-level learners: An approximate replication and extension of Erlam and Ellis (2018)

R. Erlam

University of Auckland, New Zealand

r.erlam@auckland.ac.nz

R. Ellis

Curtin University, Perth, Australia

rod.ellis@curtin.edu.au

Erlam & Ellis (2018) published, in *Canadian Modern Language Review*, an experimental study that investigated the effect of input-based tasks on the acquisition of vocabulary and markers of plurality by adolescent near-beginner learners of L2 French. The present paper reports an approximate replication of the original study with the aim of confirming or disconfirming the results. The research questions of both studies addressed the receptive acquisition of new vocabulary and the receptive and productive acquisition of markers of plurality resulting from instruction using input-based tasks. Both studies investigated near-beginner adolescent learners of French. The teacher, the students' usual classroom teacher, was the same in both studies. In the replication study, a new, larger group of students were investigated, the length of the instruction was increased, involving the development of additional tasks, and productive as well as the receptive knowledge of the vocabulary items was assessed. The results of the replication study confirm and extend those of the original study. The teachers' views about the role of input-based tasks with near-beginner learners remained constant in the two studies. The paper concludes with a discussion of the contribution that approximate replications can make to instructed second language acquisition research.

Rosemary Erlam is Senior Lecturer in the School of Curriculum and Pedagogy at the Faculty of Education, the University of Auckland. Her teaching experience and research interests include second language acquisition, task-based language teaching and language teacher education. She has a particular interest in the teaching of foreign languages to adolescents in the school context. Rosemary has published research in these areas in a range of international journals.

Rosemary Erlam

School of Curriculum and Pedagogy

The University of Auckland

Epsom Campus

74 Epsom Ave

Epsom

Auckland 1023

New Zealand

Rod Ellis is currently a Research Professor in the School of Education, Curtin University in Perth Australia. He is also a visiting professor at Shanghai International Studies University and an Emeritus Distinguished Professor of the University of Auckland. He has recently been elected as a fellow of the Royal Society of New Zealand. He has written extensively on second language acquisition and task-based language teaching. His most recent book is *Reflections on Task-Based Language Teaching* (2018) published by Multilingual Matters.

Rod Ellis

47/1 Sheen St

Subiaco

Australia

WA6008

---

1. Introduction

In Erlam & Ellis (2018) we reported an exploratory study of the effects of input-based language instruction on the incidental acquisition of L2 French vocabulary and grammar by beginner-level adolescent learners enrolled in a high school in Auckland, New Zealand. The study was motivated by two concerns, one theoretical and one pedagogic. We wanted to investigate whether tasks that were designed to create a functional need for beginner-level learners to attend to specific words and grammatical structures in the input that the tasks exposed them to, enabled them to acquire these linguistic elements incidentally. From a pedagogic perspective, we wanted to investigate the viability of an instructional innovation involving input-based tasks for teaching French in a state high school [2]. In this article we report a replication of this study with the same two aims as the original study.

The replication is of the approximate kind. Porte (2012) defined an approximate replication as 'repeating the original study in most respects, but changing non major variables (in a way that allows for comparability between the original and replication studies' (p. 8). One difference between the replication study reported here and previously published replication studies in *Language Teaching* is that we were the researchers responsible for both studies. In other words, we replicated our own original study. An advantage of this is that we were equally familiar with both studies and thus able to specify the similarities and differences between them. This also enables us to provide an informed account of the changes we made in the replication study. It is perhaps important to acknowledge that there are potential disadvantages in replication studies involving one or more of the same authors as the original study. Marsden et al. (in press) point out that there is a higher rate of replications which support initial findings when there is author overlap. However, an advantage with replications involving original authors, and our replication is a case in point,

is that there are chances of closer similarity to the initial study's materials and design (Marsden et al. in press). We were, nevertheless, not only concerned with demonstrating the reliability of results of the original study, we also wanted to see if we could improve the quality of the instructional materials and procedures based on what we had learned from their use in the original study. In other words, we wanted to see if we could enhance the pedagogical validity of the materials. The changes we made therefore were not just theoretically motivated, another difference between our study and other replication research.

1.1 Background to the two studies

The case for task-based language teaching has been persuasively presented in Ellis (2003), Van den Branden (2006), Samuda & Bygate (2008), Long (2015) and has been supported by numerous studies. However, the vast bulk of the research to date has focused on post-beginner level learners and has investigated output-based studies. With some notable exceptions (e.g. Ellis 2001; Shintani 2016) there have been few studies that have investigated beginner-level learners and even fewer that have examined the effects of input-based tasks – the type of task best suited to beginner level learners. This reflects the belief that task-based instruction is not well-suited to beginner-level learners and that it is necessary to first teach language directly to build up the linguistic resources that will enable learners to participate effectively in performing tasks. That is, there is a hidden assumption that task-based instruction necessarily involves production-based tasks (Swan 2005). In their review of a subset of research in TBLT, Plonsky & Kim (2016) focus uniquely on studies of production and include 85 studies of task-based production from 2006 to 2015! Thus, little is known about whether input-based tasks are effective in enabling learners to make a start on learning a new language without prior direct instruction in the target language. In other words, we do not know whether task-based instruction, realised through input-based tasks, constitutes a viable way of starting a language programme.

Reflecting the hidden assumption referred to above, beginner-level learners are typically taught through direct instruction by presenting and practising linguistic items deemed appropriate for this level. This was certainly true of the instructional context in Erlam & Ellis (2018). Thus, the input-based approach we investigated constituted an innovation. One of the purposes of the original study, then, was to investigate the practicality of using input-based tasks as an innovation and also, by observing the lessons and interviewing the teacher at the end of the study, to identify problems and elicit suggestions for using input-based foreign language teaching in a beginner-level classroom.

The tasks used in both the original and the replication study were input-based and focused. Input-based tasks are well-suited to beginner-level learners. Production-based tasks are not viable for this kind of learner as they have not yet developed the linguistic resources for speaking. Input-based tasks constitute one way of implementing comprehension-based instruction (Winitz 1981). They are premised on the assumption that learners will be able to comprehend the input they are exposed to and, as a result, will PICK UP specific linguistic features embedded in the input. In other words, they aim to develop learners' ability to both listen to and comprehend the target language and thereby to acquire new language. To facilitate both comprehension and acquisition the input materials need to be designed to ensure that learners will only be successful in comprehending if they are able to process the target linguistic items and features.

The tasks we designed were focused – that is, they were designed to create a context for the use and potential acquisition of a set of words and plural forms. Focused tasks are not favoured by all task-based advocates. Skehan (1998, p122-123), for example, rejects what he calls 'structure-trapping tasks'. But they do enable the researcher to investigate what effect the performance of a task has on the acquisition of pre-determined linguistic features when pre- and post-tests are incorporated into the design of an experimental study as in Ellis, Tanaka & Yamazaki (1994), De la Fuente (2006) & Shintani (2016). As Loschky & Bley-Vroman (1993) observed, it is very difficult to make the processing of predetermined linguistic features essential in production-based focused tasks. It is, however, possible to achieve this in input-based tasks. Focused tasks also have an important pedagogical advantage as they enable the teacher to see whether any learning of the target language is taking place by observing the learners' responses as the tasks are performed. It is very difficult for teachers to ascertain whether learning has occurred when the tasks are unfocused. Focused tasks, therefore, meet the needs of both researchers and teachers.

1.2 Theoretical background

Both the original and the replication study drew on the same theoretical background:

1. Receptive knowledge precedes productive knowledge of an L2.

   In the case of vocabulary it is generally accepted that receptive knowledge of new words is established before productive knowledge (Nation 2001). Research on the relationship between the production and perception of second and foreign language sounds supports the claim that perception precedes production (Flege 1991). In the case of grammar, there are also strong theoretical grounds for claiming that acquisition is initially receptive. Input-processing, however, assists the development of both receptive and productive knowledge of linguistic forms. Thus, learners can develop productive knowledge solely from performing input-based tasks.

2. The necessity of attention

   Whereas Krashen (1981) considered that acquisition was a subconscious process, it is now widely accepted following Schmidt's (1994; 2001) seminal work that conscious attention to linguistic forms, if not absolutely necessary, is likely to occur and accelerates acquisition. The aim of the input-based tasks we used was to induce learners' repeated attention to the target words and markers of plurality that were embedded in the input.

3. Overcoming default processing strategies

   The study drew on Van Patten's (1996, p. 17; 2007) Input Processing theory. This states that learners make of use default processing strategies (e.g. 'Learners process input for meaning before they process it for form') which prevent them from attending to linguistic features in the input. For example, learners are unlikely to attend to the plural markers in a sentence like *Les deux chats sont sous la table* 'The two cats are under the table' because *deux* 'two' signals that more than one cat is being referred to, removing the need to process the plural markers. Input-processing instruction aims to help learners overcome these default processing strategies by

forcing attention onto the key linguistic markers. Our studies drew on input-processing instruction but differed from it in two key respects; there was no a-priori explicit instruction and the materials took the form of tasks (as defined in Ellis & Shintani 2014) rather than discrete, decontextualized sentences. In other words, our studies were framed to investigate task-based language instruction.

4. Implicit/ incidental acquisition

The rationale for task-based language teaching rests on the claim that language learning proceeds naturally and most effectively through the incidental/ implicit learning that takes place as learners process input and output when engaged in attempts to use the language (Ellis 2003; Long 2015). Incidental acquisition refers to the picking-up and remembering of linguistic material without any deliberate intention on the part of the learners. It differs from implicit language learning in that incidental acquisition does not preclude the possibility that consciousness is involved. That is, learners may acquire incidentally with or without any awareness of what they are learning. According to the Critical Period, learners lose the ability to acquire language implicitly (DeKeyser 2000). Learners who have reached puberty, therefore, may no longer be able to acquire language implicitly. This does not preclude the possibility of incidental acquisition, however. We framed our study as a study of incidental acquisition.

Webb (2016) noted that 'research on incidental vocabulary learning through listening is an under-researched area' (p. 135). He refers to a number of studies (e.g. Elley 1989; Vidal 2003; Brown, Waring, & Donkaewbua 2008) that investigated vocabulary learning through listening to stories and noted that, while these studies did show that some learning took place, it was relatively little. In these studies, the target items were not task-essential; that is, the learners did not need to process the items in order to understand the stories. In contrast, in Ellis, Tanaka & Yamazaki (1994) and Ellis & He (1999) the LISTEN-AND-DO tasks used did make the target items task-essential. That is, learners could only successfully complete the task if they were able to understand the target items. These studies reported clear evidence of vocabulary learning.

There is almost a complete lack of research that has investigated the incidental learning of grammar through listening. Husltijn (2003) in his review of research investigating the incidental/ intentional distinction focused exclusively on studies of the incidental acquisition of vocabulary. The situation has not changed since. Shintani & Ellis (2010) and Shintani (2016) were the only studies of incidental grammar learning through instruction that we could locate. Task-essentialness is arguably even more crucial for grammar than for vocabulary. As noted above learners rely on default processing strategies. Also, many grammatical features are redundant in context and thus will not be attended to unless the task makes this necessary. In Erlam & Ellis (2018) and in the replication study reported below we designed tasks that aimed to make form-function mapping of the target words and grammar features essential for achieving the task outcomes.

1.3 The importance of ecological validity

In designing both the original and replication study we were concerned to achieve a high level of ecological validity. This was important given that one of our aims was to investigate

the practicality of input-based tasks as an alternative way to the teaching of French to beginner-level learners. For this reason, we elected to carry out the research in the learners' normal classrooms at the scheduled times for French and also to ask the students' normal teacher to teach the lessons. We wanted to see if the input-tasks WORKED pedagogically, if they resulted in acquisition, and the extent to which the teacher felt that they constituted a valid way of teaching learners at this level. The same teacher was involved in the original and replication study. Of course, this does not mean that her teaching was the same. She now had some experience of using input-based tasks and could benefit from the changes we made to the materials by incorporating the ideas that she had offered at the end of the original study. We need to acknowledge that a potential threat to the ecological validity of both studies was the presence of a researcher (one of the authors) in the classroom to document to what extent the instruction was carried out as intended and to provide support to the teacher, when requested, in the implementation of the tasks (e.g., in handing out worksheets).

2. The original study

The focus of the study was beginner-level learners' acquisition of a set of words (mainly nouns) that had been embedded in the input of the tasks and also their acquisition of French markers of plurality. To this end we formulated the following research questions:

1. Did the learners acquire a) receptive and b) productive knowledge of the target words as a result of performing the input-based tasks?
2. Did the learners acquire a) receptive and b) productive knowledge of the French markers of plurality as a result of performing the tasks?
3. What views did the a) students and b) teacher express about their experience of task-based teaching?

The first two questions addressed issues of theoretical interest as explained above. The third question was motivated by our wish to examine the extent to which the innovation was pedagogically acceptable to the teacher who had agreed to participate in the study.

The learners in the study were 34 Year 9 students in a Catholic girls' high school. They were aged approximately 13 years old. Some of the learners had had limited experience of learning French in a 'taster' programme while others were complete beginners. Two classes were involved, one was taught the lessons and served as the Experimental Group and the other was the test Control Group. The teachers of both groups were the classes' regular teachers. The lessons for the Experimental group took place at their normally scheduled times.

Both the Experimental and Control groups completed a pre-test, an immediate post-test and a delayed post-test following a 2 week holiday break. The Experimental group received 1½ hours of teaching spread out over two lessons on two separate days. The instruction took the form of five input-based tasks. The tasks were designed to provide exposure to 12 words (the target vocabulary) and to various markers of plurality in French (*SONT, DES* and *LES*). The testing regime consisted of a vocabulary listening test, a grammar listening test, and an oral elicited imitation test, with the last of these administered to only a sub-set of the students in both groups.

One of the reasons for replicating the study was to address what we came to see as a number of limitations in the original study. These limitations, which were explicitly stated in the concluding section of the report of the original study were as follows: (1) the sample size was quite small with the result that a Type II error might have occurred (i.e. significant

effects may have been lost due to the lack of statistical power), (2) the length of the instruction was very short, (3) there was no vocabulary production test so we had no knowledge of whether the input-based tasks assisted the development of productive and as well as receptive knowledge (4) we failed to make it clear to the teacher that she needed to provide feedback on the learners' responses to individual input stimuli and observations of her teaching indicated that she did not consistently do this, and (5) as the teacher pointed out in the final interview the tasks , the tasks could be made more interesting if they had incorporated an element of competition among the students. Acknowledging these limitations enabled us to identify the changes needed for the replication study.

3. The replication study

3.1 Participants

This study took place in one of New Zealand's major cities, in the same girls' school as the study reported in Erlam & Ellis (2018). The participants (n = 50) were in two French classes in Year 9, the first year of secondary school education. They were aged approximately 13 years. The school operates on a 10 day timetable and students in Year 9 have 5 lessons of French over each 10 day period, each lasting one hour. The same teacher as in the original study agreed to conduct the teaching with her Year 9 class in this study; this class is hereafter referred to as the Experimental Group. The n size was 23 compared with 19 in the original study. The teacher of another Year 9 class agreed to allow her class to participate as the Control group (n = 27; compared to n = 15 in the original study). Thus, we were able to meet one of the aims of the replication study, namely increasing the sample size of both the Experimental and Control Group.

Students completed a Background Information questionnaire. A summary of the information taken from this questionnaire is presented in Table 1. The background details of the original study are shown in brackets. Table 1 demonstrates that the majority of students in each group and in both studies had done some French at the start but for the vast majority this consisted of a 'taster' course completed during the previous two years. The aim of a 'taster' course is to give students an initial experience of the languages that will be offered at Year 9 and of which they will choose one. Accordingly in Years 7 and 8, they study each of four languages, French, Maori, Mandarin and Spanish, during a period of 20 weeks. However, their exposure to the language during this time is limited to 3 (1 hour) periods every 2 weeks (in year 9 they have 5 hour long periods every two weeks). It is likely that students in both studies had not had any exposure to French for nearly 18 months before Year 9 while others may have taken French in the 6 month period prior to beginning Year 9.

The composition of both the Experimental and Control groups in the two studies was broadly similar, although as Table 1 shows, the learners in the replication study had slightly less prior experience of learning French (i.e. they were 'beginner' like) overall.

Table 1: Background information about participants

3.2 Design

Table 2 sets out the schedule of testing and teaching for each study. A key difference is that the replication study commenced in Week 12 rather than in Week 20 as in the original study. This meant that the students had had less potential exposure to the target structures prior to the start of the replication study. The school had only 5 French lessons in every 10 day

period, therefore, teaching and testing sessions were typically spread over a time period as dictated by the school's timetabling. The testing time frames are longer in the replication study for a number of reasons. These are: more students in the replication study completed the oral production tests, there were two of these rather than one, and each test was administered one-on-one. (Ethics requirements necessitated that signed consent be obtained from both parents and students for participation in these tests). In each study the Control group sat all tests over the same timeframe but continued with their scheduled instruction. Their teacher was asked to avoid an explicit focus on the target structures during this period.

In both studies, the teacher of each class administered the receptive tests after receiving instruction from the researcher. The researcher administered the production tests. The researcher was present in the classroom to ensure that all tests were administered as planned.

Table 2: Schedule of teaching and testing in both studies

3.3 Target structures

3.3.1 Grammatical items

The target structures were markers of plurality in L2 French as in the original study. Plurality in French is marked on determiners (e.g. *les* and *des*), nouns, adjectives and verbs. However marking on nouns/adjectives and on regular verbs for plurality is not always aurally salient. As the instruction in this study was oral in nature, the study had to focus on plural marking on determiners (both the definite article *les* and the partitive article *des*) and the common regular verb *sont* (from *être*).

e.g. *Je voudrais DES pommes* 'I would like SOME apples'. *LES* pommes *SONT* rouges 'The apples ARE red'.

Markers of plurality were contrasted with the equivalent singular forms during the instructional treatments of both studies. These singular forms were included in testing in order to ascertain if students could distinguish between plural and singular referents. These singular forms were: *LE/LA, UN/UNE* and *EST* (from *être*).

3.3.2 Lexical items

With the addition of a set of tasks around the new theme of ANIMALS, there were 21 target lexical items, 8 more than in Erlam & Ellis (2018). The target lexical items are listed in Table 3 below. The items in the CLOTHES set were largely the same in both studies except that some of the cognates included in the original study were omitted from the replication study (i.e. *bikini, tee shirt, cool*). All items in the animals lexical set were new in the replication study.

Table 3: Target lexical items in the replication study

3.4 Instructional treatments

The same instructional tasks used in Erlam & Ellis (2018) were used in the present study except that an attempt was made to take up the teacher's suggestion that the tasks could be made more engaging by incorporating a competitive element. The modified task was called the FLYSWAT task, which is described below in Table 4. Additional tasks around the theme of animals were also incorporated into this study, leading to the instructional period being extended from 1 ½ to 2 ½ hours.

In both studies all tasks were given to the teacher before the treatment along with detailed instructions. She was reminded that she was not to give any explicit/rule explanations about the target forms. Both teachers were asked and agreed to avoid a focus on the target structures during the period of the study. This definitely occurred in the Experimental group as the researcher was present for all lessons. The Control group teacher said that she had given individual students some lexical items in the plural with the determiner *des* to enable them to complete activities but that there had been no explicit class focus on this feature. In the original study the Control group teacher (a different teacher for this group than in the replication study) said that there had been no focus on the target structures at all.

Before the instruction began, the students completed the same pre task in the original study. They listened to descriptions of three adolescents describing their wardrobes, hearing 22 statements in all. For each statement they saw 2 pictures and chose the picture that best matched the statement. English translations were not available to the students but are given here to assist the reader.

e.g. *J'ai des jupes* 'I have several skirts' picture of (a) one skirt (b) several skirts

In selecting one of these 2 pictures students were making a choice that demonstrated whether they had understood that the statement was singular or plural. For each choice made, in both studies, the class received immediate feedback about what the correct answer should have been but without any explanation.

After the pre task had been completed students worked through the 11 tasks depicted in Table 2 over a period of 3 lessons. Table 4 shows those tasks that were the same in both studies along with those modified or designed specifically for the replication study.

Table 4: Tasks completed during the study

During the completion of Task 1 the teacher roamed about the room and gave individual students feedback about the correctness of their choices. This was also the case the first time Task 5 was played. During Task 2 students received feedback as their classmates were given a point if they made a correct choice. This was the case for all the Flyswat games, which were played on a total of 5 occasions. It was also the case for Task 10. Otherwise students were required to make choices without receiving immediate feedback although successful completion of task outcomes indicated the extent to which they had correctly comprehended the stimuli. In the original study students received immediate feedback on their performance in the pre-task only.

3.5 Testing instruments

3.5.1 Vocabulary test

The Vocabulary test was designed to assess both comprehension and production of the target items. This represented a change from the original study as only the students in the replication study completed the production part. The comprehension part of the test was administered as a pen and paper test to all participants in both studies. The students listened as each stimulus was read out once by the teacher and circled the picture from a choice of three which best matched the word they heard. Students were scored 1 mark for each correct answer giving a possible total of 20[3]. Those students in the replication study who agreed to participate in the one-on-one oral production test (n = 40) were asked to name those items they had correctly identified on the comprehension test (it was assumed that they would be unable to correctly name items they had no receptive knowledge of). They were given a score out of 2 for each item. Two points were awarded for the correct word with correct pronunciation, a score of 1 for either an approximation of the word (e.g. *fero* for *feroce*) or incorrect pronunciation of the word (e.g. pronunciation of final 'c' in *blanc)* and 0 if they were unable to even approximate production of a word. The maximum score on this test was therefore 42 (21 items). Reliability for this test was calculated using Cronbach's alpha on post test scores of the production test and was: a = .815.

3.5.2 Grammar listening test

The Grammar listening test was designed to assess receptive knowledge of markers of plurality. As in Erlam & Ellis (2018) students listened to 30 statements repeated once only and selected from a choice of two the picture that corresponded to what they had heard. Five of the statements used in the original study that had performed poorly on an item analysis were modified in the replication study to fit with the new lexical focus. For example, item 2, which had originally been *Regardez les chats* 'look at the cats' was changed to:

Regardez les zèbres 'look at the zebras'.

Students heard, but did not see in written form, this stimulus. They then saw 2 pictures: a picture of one zebra and a picture of two zebras and circled their choice.

In both studies there were 30 stimuli in the test, 6 assessed understanding of *sont, des* and *les* and 4 assessed understanding of the corresponding singular forms (*est, un/une, le/la*).

Reliability for this test was calculated using Cronbach's alpha on post test scores: α = .500. Reliability for the post test of the original study was: α = .545. These lower than ideal reliability estimates may have been because students had a 50% chance of getting each item right by guessing. It is difficult to see how this could be avoided given that for each stimulus a binary choice between a singular and plural form needed to be made.

3.5.3 The Elicited Imitation test

This test was unchanged from the original study. A smaller subset of students (n = 40 in the replication study and n = 11 in the original study) agreed to take part. It was designed to test students' productive knowledge of markers of plurality. Students listened to 24 statements and first selected the picture out of a choice of 2 which corresponded to what they had heard. They then repeated the statement they had heard in correct French. Students were given 4 statements to practice with before they began the test. Some statements contained more than one target structure and scoring gave credit for the correct production of each. There were in total 33 items, nine testing for *sont*, and seven each for *les* and *des*. For the singular forms, 5 assessed *est* and 3 assessed *un/une* and *le/la*. Students only received credit for the correct

pronunciation of those items for which they demonstrated receptive knowledge. This was to reduce the likelihood that students were just repeating verbatim what they had heard rather than processing the stimulus (Erlam 2006).

Reliability for this test, estimated on the post test, was: a = .740. In the original study reliability for the post test was a = .710.

3.5.4 Questionnaire

In the replication study, but not in the original study, the students completed a questionnaire at the end of the study asking them their opinions about the lessons that they had received. Twenty one of the twenty three students in the Experimental group completed this questionnaire. In the questionnaire they were asked to specify if the lessons they had received during the period of the study differed from their normal class lessons. If they indicated yes they were asked to specify why. They were then asked to indicate something that they liked about these lessons and something that they did not like.

3.5.5 Interview

The teacher was also interviewed after the study had been completed. There were a range of questions that asked about her experience of taking part in this study and that gauged her attitude to task-based language teaching. Some of these questions were the same as in the original study, but others were new to the replication study. In particular, she was asked whether the changes that had been made were, in her opinion, improvements.

3.6 Analysis

Descriptive statistics were calculated for both groups on all tests used in the replication study. Tests of normality were conducted on descriptive statistics and, if assumptions of normality were not violated, parametric tests (i.e. Independent t-tests) were computed to compare the Experimental and Control groups' scores. Where assumptions of normality were violated, non-parametric Independent samples Mann-Whitney U tests were used. The alpha level for statistical significance was set at .05. Effect sizes were also calculated, to establish between-group differences, i.e., gains for the Experimental group in comparison with the Control group. The effect size was interpreted as being either small (d = >0.4), medium (d = >0.7) or large (d = >1.0) according to recommendations for L2 research. A different scale was adopted for within-group comparisons, where a d value of .60 is considered small, 1.00 as medium and 1.40 as large (Plonsky & Oswald 2014).

4. Results

4.1 Vocabulary listening test

Table 5: Descriptive statistics for the Vocabulary listening test

Tests of normality were violated for this test, therefore non--parametric tests were performed on descriptive scores to test for statistical significance. Independent samples Mann-Whitney U tests established that there were no difference in Experimental and Control group scores on

the pre-test (p = .116). However, the descriptive statistics show a difference between the two groups so t-tests were conducted on gain scores instead. Gain scores were normally distributed. Independent samples t tests showed that the experimental group made statistically significant gains over the control group from pre- to post-test, $t(48) = 2.905$, $p = .006$, $d = 0.81$ and from pre- to delayed post-test, $t(46) = 4.557$, $p = .000$, $d = 1.28$.

Table 6 compares the within and between group effect sizes of the original and replication study. These results are very similar except that the within-group effect sizes were notably larger in the replication study and the between-group effect size smaller for the immediate post-test. (see Table 6).

Table 6: Effect sizes (d) for the Vocabulary listening test in the original and replication studies

4.2 Vocabulary production test

A smaller proportion of students in each group (Experimental, $n = 21$; Control, $n = 19$) took the Vocabulary production.

Table 7: Descriptive statistics for Vocabulary production test

Tests of normality were not violated for this data. Independent samples t-tests showed no significant different between groups on the pre-test, $t(38) = 1.023$, $p = .313$. There were significant differences between the Experimental and Control group for both the post-test, $t(38) = 5.591$, $p = .000$ and the delayed post-test, $t(38) = 5.535$, $p = .000$. Between-group effect sizes were large - $d = 1.780$ for the immediate post-test and 1.802 for the delayed post-test. This test was not administered in the original study.

4.3 Grammar listening test

Table 8: Descriptive statistics for Grammar listening test

Tests of normality were not violated for the pre-test scores. Independent samples t-tests established that there was no difference in Experimental and Control group scores on the pre-test, $t(48) = .834$, $p = .408$. Tests of normality were violated for both the post and delayed post-test versions of this test, therefore non--parametric tests were performed on descriptive scores to test for statistical significance. There were statistically significant differences between Experimental and Control group scores on both the immediate post-test ($p = .000$) and delayed post-test ($p = .000$).

A direct comparison with the original study is not possible as there was a near significant difference between the Experimental and Control Groups in the pre-test so gain scores rather than actual post-test scores were used to compare groups. A comparison of effect sizes in the two studies, however, is revealing. Whereas the between group effect size decreased from immediate to delayed post-test in the case of the original study it increased in the replication study. A similar difference is evident in the change in the within group effect

sizes over time. That is, whereas the effect sizes for the two within-group comparisons decreased in the original study, they increased for the replication study.

Table 9: Effect sizes (d) for the Grammar listening test in the original and replication studies

4.4 Elicited imitation test

A subset of students took this test. However, this was much larger than the subset of students (n = 6) who took this test in the original study. This test was designed to measure learners' productive knowledge of markers of plurality.

Table 10: Descriptive statistics for the Elicited imitation test

Tests of normality were not violated for the pre-test and immediate post-test scores. Independent samples t-tests showed no significant different between groups on either the pre or immediate post-test, t (38) = -.847, p = .402; t (38) = 1.358, p = .183. Tests of normality were violated for the delayed post test scores. A Mann-Whitney U test showed no significant difference between the Experimental and Control groups on the delayed post-test ($p > .05$). For the Experimental group, the within-group effect sizes were respectively d = 0.456 and d = 0.589 for pre- to immediate post-test and for pre- to delayed post-test.

In the original study the Elicited imitation test was only administered as an immediate post-test. As in the replication study, the group difference was not statistically significant. The within-group effect size (pre-test to post-test) for the Experimental Group was d = 3.176.

4. 5 Retrospective reports

4.5.1 Student questionnaire

The students were given a questionnaire at the end of the study asking them their opinion about the lessons that they had received during the study. Twenty-one of the 23 students who participated in the study completed this questionnaire.

Nineteen students thought that the lessons were different, only two felt that they were not. The most common reason given for why the lessons were different was the fact that they had to complete so many tests (11 students). Six students said that playing games was different and four mentioned that they did not work from their text books. Two students in each case mentioned the following: the lessons were fun/ there was no computer work/ the lessons were focused on listening/ they felt they learnt more/they didn't feel they learnt much. One student stated that the lessons were more interactive and another that they were easier.

In response to the question asking what they liked about the lessons, they mentioned the games (8 students), fun lessons (4), lessons that were more interactive (3), learning new words (2), repetition (2). One student in each case mentioned liking the following: different ways of learning/ seeing pictures that related to words/ learning a lot. In response to the question about what they did not like about the lessons two students in each case mentioned that they did not like the following: repetitive lessons/ not knowing what they were learning

about/ not feeling that they learnt a lot. One student felt that she did not learn anything and did not like this.

4.5.2 Teacher interview

In the interview, the teacher commented that the teaching was quite different from what she was used to as normally she would just give a little bit of input and then have the students practise production. She suggested that one problem with the input-based tasks was that the students could just guess, whereas having them produce the target words and items would eliminate this and prompt them to ask questions about what they were learning. She thought that grammar is hard to learn unless you are explicitly taught it.

She noted that the changes made in the replication study helped to reinforce learning – a view not entirely borne out by the results, however. Overall, though, she found the tasks quite repetitive and thought that the students probably got quite bored with them but admitted that that might have been because she got bored herself. She commented 'for me it probably just went on too long – doing the same thing over and over again'. Her suggestion for a new task was a shop task or restaurant task involving student production (e.g. ordering from a menu).

The teachers' comments are similar to those she made at the end of the original study where she also commented on the repetitive nature of the tasks and the need for production activities.

5. Discussion

A number of changes were made to the original study with the purpose of improving the quality of the instruction and enhancing the internal validity of the replication study. The main changes were:

- The sample size of both the Experimental and Control Groups was increased.
- Additional tasks were added and also some of the tasks introduced a competitive element.
- A Vocabulary production test was included.
- More learners completed the Elicited imitation Test.
- The learners completed a post-study questionnaire to elicit their views about the instruction they had received.

Nevertheless, both studies involved learners taken from the same student population, the instructional context and the design of the two studies was the same, they covered a very similar period of time, and they involved the same teacher for the Experimental Group (but not for the Control Group). The testing regime was also identical with the exception of the addition of the Vocabulary production test. On a continuum stretching from exact to approximate to conceptual replication (Polio & Gass 1997), the replication study lay between exact and approximate. The fact that the researchers exerted control over environmental and other variables in this study, may account in part for the results obtained. Plonsky & Oswald (2014) suggest that gains may be considerably greater in contexts where there is control over those variables that might impact on results.

The replication study, like the original study, had two aims. The first concerned whether (and to what extent) incidental acquisition occurred as a result of performing the input-based tasks. This was addressed by considering both vocabulary (Research Question 1) and

grammar (Research Question 2). The second aim was to evaluate to what extent a pedagogic innovation involving the use of input-based tasks was successful.

By and large the results of the replication study mirrored those of the original study. In both studies the learners demonstrated significant gains in receptive knowledge of the target words in the immediate test and maintained these in the delayed post-test (two weeks later). The within-group effect sizes for the Experimental Group were medium in the original study and large in the replication study. The between-group effect sizes were medium for the immediate post-test in the replication study but large for the delayed post-test in both studies[4]. Thus, we can confidently claim that incidental acquisition of receptive knowledge of words occurs when performing input-based tasks if the target words are made task-essential. This result mirrors those of other studies of incidental acquisition from input-based tasks (e.g. Ellis, Tanaka & Yamazaki 1994; Ellis & He 1999; Shintani 2016). Table 5 shows that on average the learners in the Experimental Group acquired receptive knowledge of 5 new words as a result of performing the tasks in 2 ½ lessons lasting 150 minutes – a rate of acquisition that compares favourably with that of other input-based studies.

The replication study extended the finding of the original study by showing that the input-tasks also facilitated the incidental acquisition of productive knowledge of the target words even though the tasks did not actually require production of the words. The learning that occurred was also maintained over time. Shintani (2016), for a very different population of learners (6 year old Japanese children learning English), also reported that listen-and-do tasks fostered the acquisition of productive knowledge of new words.

It could be argued that intentional learning of the target words would have resulted in more words being learned – both receptively and productively - in a much shorter time. Input-based tasks, however, provide a context for much more than learning words. As in our study they also offer opportunities for learning grammar and, more importantly, for experiencing the processing of input in real time – an aspect of learning crucial for developing communicative competence in a language.

The results for the incidental acquisition of grammar in both the original and replication study showed that the learners made significant gains in their receptive knowledge of French markers of plurality. There was, however, an interesting difference in the results of the two studies. In the original study, the gains evident in the immediate post-test were largely lost in the delayed post-test with the result that there was no statistically significant difference between the Experimental and Control Group. In the replication study, in contrast, not only was the beneficial effect of the instruction maintained in the delayed post-test but also enhanced as can be seen in the effect sizes shown in Table 9. The likely explanation for this is the greater exposure to task-essential exemplars of the target grammatical features as a result of the increase in number of tasks and the length of instruction. It is interesting to note that the effect sizes in the replication study for both between-group comparisons are similar to those presented in Shintani, Li & Ellis's (2013) meta-analysis of research involving comprehension-based grammar instruction. They also reported large effect sizes for between group immediate (d = 1.96) and delayed post-test comparisons (d = 1.58). On the other hand, they also reported large effect sizes for within group comparisons, which we did not find in either of our studies.

We had anticipated that the greater exposure would also result in productive acquisition of the target structure. However, the results for the Elicited imitation test showed that, as in the original study, there was no statistically significant difference between the scores for the

Experimental and Control Groups. Gains from pre- to post-test were largely the result of a test practice effect. This was somewhat disappointing. Shintani & Ellis (2010) and Shintani (2016) reported that exposure to English plural-s in listen-and-do tasks resulted in durable statistically significant gains in an oral production task by young Japanese children. It is possible that difference in results in these and our studies was due to the difference in age of the students involved (i.e. 6 years as opposed to 13). But we consider a more likely explanation lies is the difference in instructional time - 2 ½ lessons in our replication study as opposed to 6 lessons in Shintani & Ellis (2010) and 9 in Shintani (2016). In other words the incidental acquisition of productive knowledge of grammatical features such as those investigated in these studies will only occur if learners have very substantial task-essential exposure to them or have already started to acquire them at the beginning of the study.

To evaluate the extent to which the input-based instruction – an innovation in the particular instructional context we investigated – was successful, the lessons were observed by a researcher (one of the authors) and we administered a questionnaire to the students and interviewed the teacher. As in the original study, observations of the classes showed that the teacher was able to execute the input-based tasks as intended, including the small change involving providing immediate feedback on the students' responses in tasks where this was possible. The students' responses to the questionnaire also indicated that most of them enjoyed the tasks and the interactive nature of the classes. However, some of the students were less positive, querying whether they actually learned anything and indicating a preference for an instructional approach that catered to intentional learning. The teacher also felt that explicit instruction would have helped and also clearly favoured an approach that involved student production. We were left with the impression that once the study was over the teacher would return to her usual way of teaching if only because to continue would mean the teacher having to develop her own input-based tasks, which she said she did not have time for. In short, it was clear that the innovation was successfully implemented but would not lead to long-term use of input-based tasks.

6. Conclusion

Mackey (2012) commented 'to qualify as a candidate for replication, a study should address appropriate, theoretically interesting, and currently relevant research questions' (p. 28). Our replication study addressed questions about incidental acquisition that are of both theoretical and pedagogic importance as well. While there is ample evidence that vocabulary can be acquired incidentally (Hulstijn 2003), little is currently known about the incidental acquisition of grammar, especially in a classroom context. Do learners PICK UP grammar from the input they are exposed to? In our replication study we sought to not just confirm the results of the original study (and many other studies) regarding the incidental acquisition of vocabulary but to also probe whether grammar can also be acquired incidentally by beginner-level learners, which has been little investigated to date. Our replication study also addressed an important pedagogical question – how can task-based language teaching be made appropriate for beginner-level learners with no speaking ability?

Mackey (2012) also commented that there needs to be a clear rationale for a replication study. The rationale for the present replication study was the wish to establish the reliability of the findings of the original study and also, by introducing a few changes designed to improve the quality of the instruction, to see if the results could be enhanced. The replication study provides evidence of the robustness of the original findings, namely that a

short period of instruction involving input-based tasks results in receptive knowledge of target words and grammatical structures. The changes we made showed also that this type of instruction results in productive knowledge of the target words. In the case of the target grammatical structures, however, the greater exposure in the replication study did not lead to productive knowledge. This finding is important both theoretically for second language acquisition researchers and pedagogically for proponents of task-based teaching. It raises questions about how much input is needed for the incidental acquisition of grammar and whether a task-based approach will work more efficiently for grammar learning if it is complemented by some explicit instruction – as the teacher involved in the study thought necessary.

A further aim of replicating the original study was to accumulate evidence about the feasibility of using input-based tasks to teach a foreign language in a state high school. The replication study confirmed the feasibility of this approach but also suggested that it was unlikely to be adopted in the long term by the teacher involved. An implication is that input-based tasks have a role to play with beginner learners but that, in a context where there is limited instructional time available, they may prove acceptable (and possibly more effective) if accompanied with some direct teaching.

Finally a comment about conducting replication studies. Mackey (2012) observed that one problem that arises is when there are differences in the findings between the original and the replication study and trying to decide which set of results are reliable and how to explain the differences. We did not experience this problem in our replication study. In part this was because there were few differences in findings, perhaps, because it lay quite close to the exact end of the replication continuum. But also it was because we were replicating our own study and so had complete and detailed knowledge of the changes we had made to the original study. Other researchers might also like to consider replicating their own studies.

Notes

1. We have followed the guidelines provided by Brown (2012) for reporting a replication study and also benefitted from a model replication study supplied by the editor of *Language Teaching*.
2. The teacher who taught the lessons had been introduced to task-based language teaching as part of an in-service professional development programme, however had not used any input-based tasks in her teaching.
3. The colours ‘noir’ and ‘blanc’ constituted one item only in the comprehension test but 2 in the production test.
4. The likely explanation for the larger effect sizes in the replication study is the additional words included in this study. A comparison of the pre-test scores for the two studies shows that the learners knew a larger proportion of the target words in the original study (a mean of 9.58 out of 12) than in the replication study (a mean of 11.26 out of 20).

References

Brown, J.D. (2012). Writing up a replication report. In G. Porte (ed.) *Replication in applied linguistics: A practical guide.* Cambridge: Cambridge University Press, 173–197.

Brown, R., R. Waring & S. Donakaewbua (2008). Incidental vocabulary acquisition from reading, reading-while-listening and listening to stories. *Reading in a Foreign Language* 20.2, 136–163.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22, 499–533.

De la Fuente, M. (2006). Classroom L2 vocabulary acquisition: Investigating the role of pedagogical tasks and form-focused instruction. *Language Teaching Research* 10, 263–295.

Elley, W. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly* 24, 174–187.

Ellis, R. (2001). Non-reciprocal tasks, comprehension and second language acquisition. In M. Bygate, P. Skehan & M. Swain (eds.), *Researching pedagogic tasks, second language learning, teaching and testing*. Harlow: Pearson Education, 49–74.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. & X. He. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition* 21.2, 285–301.

Ellis, R. & N. Shintani. (2014). *Exploring language pedagogy through second language acquisition research.* London: Routledge.

Ellis, R., Y. Tanaka & A. Yamazaki. (1994). Classroom interaction, comprehension and the acquisition of word meanings. *Language Learning* 44.3, 449–491.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics* 27.3, 464–491.

Erlam, R. & R. Ellis. (2018). Task-based language teaching for beginner-level learners of L2 French: An exploratory study. *Canadian Modern Language Review* 74.1, *1–26.*

Flege, J.E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In Th. Huebner & ChA. Ferguson (eds.), *Crosscurrents in Second language Acquisition and Linguistic Theories.* Philadelphia: John Benjamins, 249–289.

Hulstijn, J. (2003). Incidental and intentional learning. In C. Doughty & M. Long (eds.), *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell Publishing, 349–381.

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamaon.

Long, M. (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley Blackwell.

Loschky, L. & R. Bley-Vroman. (1993). Grammar and task-based methodology. In G. Crookes & S. Gass (eds.), *Tasks and language learning: integrating theory and practice*. Clevedon: Multilingual Matters, 123–167.

Mackey, A. (2012). Why (or why not), when, and how to replicate research. In G. Porte (ed.) *Replication in applied linguistics: A practical guide.* Cambridge: Cambridge University Press, 21–46

Marsden, E., K. Morgan-Short, S. Thompson, & D. Abugaber. (in press). Replication in Second Language Research: Narrative and Systematic Reviews and Recommendations for the Field. *Language Learning.*

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Plonksy, L., & Y. Kim. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics,* 36, 73–97.

Plonsky, L., & F. Oswald. (2014). How Big is 'Big'? Interpreting Effect Sizes in L2 Research. *Language Learning* 64.4, 878–912.

Polio, C. & S. Gass. (1997). Replication and reporting: A commentary. *Studies in Second Language Acquisition* 19, 499–508

Porte, G. (2012). *Replication **i**n applied linguistics: A practical guide*. Cambridge: Cambridge University Press.

Porte, G. (2013). Who needs replication? CALICO Journal 30.1, 10–15.

Samuda, V. & M. Bygate. (2008). *Tasks in second language learning*. Basingstoke: Palgrave MacMillan

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review* 11, 11–26.

Schmidt, R. (2001). Attention. In P. Robinson (ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press, 3–32.

Shintani, N. (2016). *Input-based tasks in foreign language instruction for young learners*. Amsterdam, Netherlands.

Shintani, N. & R. Ellis. (2010). The incidental acquisition of English plural *–s* by Japanese children in comprehension-based lessons: A process-product study. *Studies in Second Language Acquisition* 32.4, 607–637.

Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language Learning* 63, 296–329.

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

Swan, M. (2005). Legislating by hypothesis: The case of task-based instruction. *Applied Linguistics* 26, 376–401.

Van den Branden, K. (ed.) (2006). *Task-based language education: From theory to practice*. Cambridge: Cambridge University Press.

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Norwood, N.J.: Ablex.

VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten & J. Williams (eds.), *Theories in second language acquisition: An introduction.* Mahwah, NJ: Lawrence Erlbaum, 115–135.

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics* 24.1, 56–89.

Webb, S. (2016). Learning vocabulary through meaning-focused input: Replication of Elley (1989) and Lie & Nation (1985). *Language Teaching* 49.1, 129–140.

Winitz, H. (1981). *The Comprehension Approach to Foreign Language Instruction*. New York: Newbury House.

Table 1: Background information about participants

	Prior experience learning French			Speak another language at home	
	no	**Yes as a 'taster' course only**	**More sustained learning**	**yes**	**no**
Experimental group (n =23)* (n = 19)	5 (3)	15 (12)	1 [for 2 yrs] (3)	3 (3)	18 (15)
Control group (n =27) (n = 15)	8 (4)	17 (9)	2 [for 3 or more yrs] (1)	7 (2)	20 (12)

*Background questionnaire data was available for 21 participants only in the Experimental group

Table 2: Schedule of teaching and testing in both studies

Event	**Replication study**	**Original study**
Pre-testing	March 22nd -24th	June 23rd
	Easter break [1 week]	
Teaching	April 1 – 6th [2 ½ lessons]	June 26th – 29th [1 ½ lessons]
Post testing	April 6th – 11th	June 29th – July 3rd
	2 week break	
Delayed post testing	May 4th – 9th	July 21st

Table 3: Target lexical items in the replication study

lexical set	**nouns**	**adjectives**
clothes	*casquette, pull, jupe, jean, short, tennis, pyjama, bottes*	*vieux*
animals	*enclos, singe, phoque, ours, serpent*	*petit, grand, feroce, jaune, noir, blanc, gris*

Table 4: Tasks completed during the study

Task no & name, times completed	Description	Replication study	Original study
1. Bingo, 2X	Students cross out items on a bingo sheet as teacher reads out stimuli. e.g. *UN pull/DES pulls* 'a pullover/some pullovers'	X	X
2. Flyswat (clothing)	A picture of clothing is projected on the wall. Class is divided into 2 teams. Students take turns to play for their team and be the first to hit the picture corresponding to the stimulus read out by the teacher. e.g. *DES jupes* 'skirts'	X	
3. Shopping	Students listen to descriptions of what French adolescents buy and enter the letter corresponding to each purchase (from a table of pictures) and money spent to a table. They then work out who spend the most. E.g. *Pascal achète DES tee shirts*.	X	X
4. Cool or not	Students decide whether the items in task 3 are *cool* or *moche* 'ugly' as the teacher reads out questions. E.g. *Est-ce que LES pulls SONT cools ou moche?* 'are the pullovers cool or ugly ?'	X	X
5. Bingo moche 2X	As in Task 1 but students listen to statements about whether clothes are *cool* or *moche*, cross off the item and put ☺ or ☹ next to it. E.g. *LES pyjamas SONT moches* 'the pyjamas are ugly'	X	X
6. Flyswat (cool or not) Played on 2 separate occasions.	As in Task 2, but pictures depict plural or singular clothing items with emoticons indicating whether they are *cool* or *moche*. E.g. *ils SONT cools*	X	
7. Shopping in the sales	As in Task 3 but the items are reduced in price. Students have to establish which of the shoppers saves the most.	X	X
8. Flyswat (animals)	As in Task 2 but this time pictures depict either one or more than one animal.	X	
9. Building a zoo	Students have a picture of a zoo and envelopes of cut-out animals. As teacher reads out stimuli they place animals in correct position. e.g. *Dans l'enclos A il y a DES lions* 'In enclosure A there are some lions'	X	
10. Who is it?	Students have zoo picture. They listen to descriptions of animals and say who is being referred to. E.g. *Ils SONT gris. Ils SONT grands* 'They are grey. They are big'.	X	
11. Flyswat (animals)	As in task 2 but students are given descriptions of animals. E.g. *Ils SONT feroces* 'They are fierce'.	X	

Table 5: Descriptive statistics for the Vocabulary listening test

Test	**Experimental group**			**Control group**		
(Max = /20)	*M*	*SD*	*n*	*M*	*SD*	*n*
Pre test	11.26	2.20	23	10.26	2.55	27
Post test 1	16.70	1.19	23	11.30	2.54	27
Post test 2	16.62	1.20	21	12.26	2.33	27

Table 6: Effect sizes (d) for the Vocabulary listening test in the original and replication studies

	Within group		Between group	
	Pre-immediate PT	Pre- delayed PT	Immediate PT	delayed PT
Original study	1.150	1.293	2.034	1.685
Replication study	2.462	2. 508	0.81	1.28

Table 7: Descriptive statistics for Vocabulary production test

Test	**Experimental group**			**Control group**		
(Max = /42)	M	SD	n	M	SD	n
Pre-test	9.71	4.22	21	8.37	4.09	19
post-test 1	23.43	4.39	21	14.05	6.15	19
post-test 2	25.58	4.43	19	16.67	5.35	18

Table 8: Descriptive statistics for Grammar listening test

Test	**Experimental group**			**Control group**		
(Max = /30)	M	SD	n	M	SD	n
Pre-test	16.35	2.39	23	15.82	2.13	27
post-test 1	19.91	3.30	23	16.15	2.80	27
post-test 2	20.86	2.46	21	16.19	2.37	27

Table 9: Effect sizes (d) for the Grammar listening test in the original and replication studies

	Within group		Between group	
	Pre-immediate PT	Pre- delayed PT	Immediate PT	delayed PT
Original study	1.091	0.746	1.392	0.621
Replication study	0.760	1.107	1.233	1. 934

Table 10: Descriptive statistics for the Elicited imitation test

Test	**Experimental group**			**Control group**		
(Max = /33)	*M*	*SD*	*n*	*M*	*SD*	*n*
Pre test	12.24	3.27	21	13.05	2.76	19
Post-test 1	15.10	6.15	21	12.95	3.27	19
Post-test 2	16.05	7.00	19	12.89	3.74	15