Investigating explicit and implicit L2 knowledge and learning: Replications of Erlam (2005) and Roehr-Brackin & Tellier (2019)

Karen Roehr-Brackin (University of Essex)

Abstract

This paper makes the case for close and approximate replications of Erlam (2005) and a conceptual replication of Roehr-Brackin & Tellier (2019). The two studies recommended for replication are informed by research on explicit and implicit knowledge, learning and teaching. They are ecologically valid classroom studies with either adolescent or child learners as participants and thus worked with as yet relatively underrepresented populations in the field of instructed second language acquisition. Erlam (2005) identified a levelling effect of a particular method of explicit instruction, while Roehr-Brackin & Tellier (2019) showed that language-analytic ability has a role to play even in younger children's language learning. The researchers' approaches duly reflect the need to take into account cognitive individual learner differences when working in intact classrooms. As the findings of each of the original studies have potentially profound implications for theory and practice in the field, replication is deemed both timely and desirable. In order to facilitate this endeavour, the key features of the original studies are summarised, and specific proposals on the methodological characteristics of suitable replication studies are put forward.

Biographical note

Karen Roehr-Brackin is a Reader in the Department of Language and Linguistics at the University of Essex in the UK and director of the Centre for Research in Language Development throughout the lifespan. She has published widely in second language acquisition as well as the role of individual learner differences in instructed second/additional language learning at different ages. Her work has been published in international peer-reviewed journals and edited volumes, and she has authored a monograph on metalinguistic awareness and second language acquisition, published in 2018.

Introduction

The role of explicit and implicit knowledge and learning in the context of second language acquisition (SLA) has attracted considerable interest over the past decades (N. C. Ellis, 1994; Rebuschat, 2015). There is a large body of research that has investigated to what extent explicit learning takes place in different settings and in learners of different ages, as well as when and how explicit knowledge may be beneficial for second language (L2) development (R. Ellis, 2004; Roehr-Brackin, 2018). By the same token, researchers have sought to establish to what extent and in which contexts L2 learners of different ages may learn implicitly and construct implicit knowledge (DeKeyser, 2003; N. C. Ellis, 2005, 2015). This work has been complemented by studies investigating the effectiveness of different types of language instruction, and specifically the potential impact of explicit and form-focused vs. implicit and meaning-focused teaching (Goo, Granena, Yilmaz, & Novella, 2015; Norris & Ortega, 2001).

The two original studies that are the focus of this paper and are recommended for replication are situated at the core of the triangle of explicit and implicit knowledge, learning and teaching: Erlam (2005) and Roehr-Brackin & Tellier (2019). In the remainder of this section, I outline the impact of the two studies on their field, and I make the case for replication studies to be undertaken. In the next section, a summary of each of the original studies is provided. Subsequently, I discuss in detail what kind of replication study (close, approximate, conceptual) is recommended, and what methodological approach each recommended replication study should take. The final section offers concluding remarks that are applicable to all of the replications proposed here.

The specific impact of Erlam (2005) in the area of explicit and implicit L2 learning and teaching is attributable to the researcher's argument that a particular explicit instructional approach may have a levelling effect, in the sense that all learners, regardless of individual differences in certain cognitive abilities, may benefit in equal measure if they are exposed to such teaching. Needless to say, the identification of an instructional approach where one size really does fit all would be a revelation to the language teaching profession, language learners and the publishing industry alike. Subsequent studies have taken up the idea of a potential levelling effect through learners'

habituation in order to retrospectively explain obtained patterns of results (Rodríguez Silva & Roehr-Brackin, 2016; Tellier & Roehr-Brackin, 2013; Tomak, 2019), but there is currently no other study that has taken the bull by the horns, as it were, and set out to directly and deliberately replicate Erlam's (2005) key finding.

Roehr-Brackin & Tellier (2019) argue that language-analytic ability, a component of language learning aptitude that is typically assumed to be of importance in adult L2 learning, may be a predictor in child L2 learning as well. This implies that it may actually be possible even for quite young children to learn explicitly, provided that they are exposed to age-appropriate form-focused instruction. The pattern of findings obtained in this study is in direct contrast with the results obtained in the most directly comparable study (Muñoz, 2014), which is all the more remarkable as the children in Roehr-Brackin & Tellier's (2019) study were, at ages 8-9, at the younger end of the scale. The conclusions drawn challenge the long-held assumption that younger children cannot (yet) be analytic learners (Harley & Hart, 1997; Sawyer & Ranta, 2001), while they are compatible with recent work that has set out to directly compare children and adults exposed to the same explicit or implicit instructional approach in a controlled experimental setting (Lichtman, 2013, 2016). Thus, although the study is a very recent addition to the field, there is a clear need to establish whether the results that may seem counter-intuitive to some can indeed be replicated.

In summary, the conclusions put forward by each of the two studies in focus represent a potential step-change. At the same time, it must be acknowledged that these conclusions have arisen from data from a single, though sizeable, sample of participants. Therefore, replication of the original studies is an important endeavour to give researchers and research users such as practitioners and policy makers full confidence in the implications that can be derived from these studies.

The two studies share a number of characteristics which help us situate them in the field of explicit and implicit L2 knowledge and learning. First and foremost, both studies exemplify applied empirical research in that they work with participants in language classrooms. This emphasis on ecological validity is desirable, but still in comparatively short supply in the field. The need to control independent variables in research with an experimental or quasi-experimental design has led to a

number of laboratory-based studies (e.g. Lichtman, 2016; Sanz & Morgan-Short, 2004; Suzuki & DeKeyser, 2017; Tolentino & Tokowicz, 2014) that have yielded very valuable findings. However, such studies have necessarily relied on relatively short treatments and/or the use of artificial or semi-artificial mini-languages. To what extent results arising from laboratory studies can be directly applied to the L2 classroom is debatable, so applied research carried out in actual classrooms is needed.

Second, both original studies work with school-age learners, i.e. adolescents (Erlam, 2005) and children (Roehr-Brackin & Tellier, 2019). The vast majority of research in the area of explicit and implicit L2 knowledge and learning, just like the vast majority of research in all other areas of SLA, has relied on educated adult participants at college or university level. Younger learners are comparatively under-represented, and this is no doubt at least partly due to the practical difficulties of conducting research in school settings. Parental consent must be obtained, school leaders and teachers must be willing to allow researchers into their schools despite potential disruption to a busy curriculum, and the participating young learners themselves should ideally be motivated and willing to do their best. Although these conditions can be daunting for researchers, it is worth bearing in mind that much L2 learning and teaching across the world takes place in school settings, and if we wish to understand language development in such settings, we must include them in our research agenda, practical challenges notwithstanding.

Third, and related to the previous point, research into explicit and implicit L2 knowledge and learning has often focused on English as the target language. No doubt this has been due to the ready availability of English language learners in educational institutions across the world. Thus, similarly to the age range of typical research participants, practical circumstances may have dictated their typical L2. The two studies that are recommended for replication both focus on L2 French – also a high-prestige European language and thus not rarely investigated, but a step removed from the extremely popular target language that English has become.

Last but certainly not least, both original studies have included moderating variables that are likely to interact with explicit and implicit L2 knowledge and learning, thus appropriately reflecting

the increased sophistication of research in the field. Language learning aptitude, or the ability to learn new languages quickly and with ease, has a long and distinguished history in SLA, and its predictive power in the context of SLA has been amply demonstrated (Li, 2015, 2016; Wen, Biedroń, & Skehan, 2017). However, classroom-based studies concerned with explicit and implicit knowledge and learning in children or adolescents have rarely investigated the role of aptitude, although the cumulative findings of aptitude research on the one hand and research on explicit and implicit knowledge and learning on the other hand are sufficiently robust to allow for specific predictions to be made. Erlam (2005) and Roehr-Brackin & Tellier (2019) both included measures of participants' language learning aptitude; Erlam (2005) additionally measured her adolescent participants' working memory, while Roehr-Brackin & Tellier (2019) assessed their child participants' metalinguistic awareness.

Taken together, the points discussed so far can be consolidated into a general argument for replication: the studies in focus worked with less frequently represented populations, a less investigated target language and included important predictors. Therefore, it is particularly crucial that the potentially very impactful findings obtained in the original studies be substantiated. Beyond this general argument, there are a number of specific methodological reasons which further strengthen the case for replication.

Each of the two studies in focus has methodological limitations which are mostly acknowledged by the authors and can be addressed in replication work. In brief, Erlam (2005) used a single target structure, French direct object pronouns. It would be desirable to use a different structure as a target or, ideally, include several target structures to ensure that any results extend beyond one grammar point. Moreover, two further limitations are noted by the researcher herself (Erlam, 2005, p. 167): some of the tests used in the study had low reliability (Cronbach's alpha = .595 for the aptitude measure of language-analytic ability, .432 for the listening comprehension test, .654 for the reading comprehension test), and the test used to measure participants' working memory was potentially problematic, given that it was group-administered and relied on written responses. Finally, the order of test administration in the study was less than ideal, with aptitude measures administered several

months after the experimental intervention. As aptitude served as a predictor, it should ideally be measured before any experimental treatment commences.

In Roehr-Brackin & Tellier (2019), the child participants experienced the target language in a single instructional condition, i.e. age-appropriate French classes with a focus-on-form element. Thus, the reported findings are restricted to this type of instruction and cannot be generalised to or compared with other types of instruction children in particular may well be exposed to. A replication study could include two or more instructional conditions as well as a control condition.

In order to contextualise the specific points made in the previous paragraphs and prepare the ground for more detailed recommendations for approaches to replication, the next section provides brief summaries of the two original studies.

The original studies: Erlam (2005) and Roehr-Brackin & Tellier (2019)

Erlam's (2005) study can be loosely situated in the aptitude-treatment interaction (ATI) paradigm within L2 learning and teaching research (for a brief overview, see Roehr, 2012). Broadly speaking, ATI research seeks to establish whether matching type of instruction with learners' aptitude profiles will lead to improved results both in terms of L2 achievement and learner satisfaction (de Graaff, 1997; Ranta, 2002; Robinson, 1997; Wesche, 1981). Erlam (2005) states that she did not fully follow the ATI paradigm, since she did not allocate study participants to specific instructional conditions in accordance with their cognitive ability profiles. However, in line with ATI research, she aims to draw conclusions as to whether certain approaches to instruction are especially effective with learners exhibiting particular aptitude profiles.

A total of 92 English-speaking learners of L2 French at a high school in New Zealand participated in the study, which had a quasi-experimental pre-test, immediate post-test, two-month delayed post-test design. The participants were approximately 14 years old and in their second year of French study. As noted above, the targeted structure was French direct object pronouns. The participants were assigned to one of three instructional conditions (deductive, inductive, structured input) or a control group. The learners in the experimental groups were exposed to three 45-minute

sessions of instruction. The learners' L2 achievement was assessed by means of listening and reading comprehension, written and oral production tests covering the four skills of speaking, listening, reading and writing. Six months after the instructional intervention, the 60 learners from the experimental conditions completed a test battery consisting of aptitude subtests measuring phonetic coding ability and language-analytic ability and a working memory task.

The main findings arising from the study can be summarised as follows. No statistically significant effect of experimental group was found for any of the aptitude subtests used in the study. The measure of language-analytic ability accounted for 24% of the variance in the written production immediate post-test, and the measure of language-analytic ability and the working memory test accounted for 48% of variance in the written production delayed post-test, with the working memory test making a greater contribution.

Erlam (2005) reports little overall evidence that deductive instruction was particularly beneficial to learners who did well on the aptitude measures. Indeed, quite to the contrary, it appears that this type of instruction benefited all learners. In other words, with regard to L2 achievement, effects of individual differences in aptitude and working memory were minimised in the deductive condition. In this context, the researcher notes that the deductive condition was closest to the type of instruction the learners were used to from their regular language classes, which may have resulted in greater motivation and thus more effort being put in.

Learners with higher levels of language-analytic ability benefited more from the inductive instruction, as suggested by their performance on the written production delayed post-test. Overall, it appears that the inductive group needed time to consolidate their learning of the target structure. The structured input group showed various correlations between the language-analytic ability subtest, the test of working memory and the written production test. Overall, learners with higher levels of language-analytic ability and working memory capacity benefited more from structured input instruction.

Overall, working memory scores correlated more strongly with delayed post-test performance than with immediate post-test performance, which suggests that learners with higher levels of working memory were better able to maintain long-term representations of the target structure. The researcher suggests that such learners processed the input more deeply and may therefore be better able to successfully maintain long-term representations.

In conclusion, the most interesting and impactful finding arising from Erlam's (2005) study is the apparent absence of a role for individual learner differences in aptitude or working memory in the deductive instructional condition. In other words, it appears to be the case that deductive instruction has a levelling effect, that is, it works equally well with learners of differing cognitive ability profiles.

Roehr-Brackin & Tellier's (2019) study is situated in the domain of children's instructed L2 learning. As language instruction in primary schools has become the norm in many countries across the world, the question of an ideal starting age and children's associated capacity for language learning has attracted renewed attention. Contrary to popular belief, previous research has shown conclusively that in a classroom setting, young children are relatively slow and inefficient learners compared with adolescents and adults (Jaekel, Schurig, Florian, & Ritter, 2017; Larson-Hall, 2008; Muñoz, 2006). This is typically attributed to children's reliance on implicit learning, which is a gradual process requiring large amounts of input over a long period of time.

Clearly, neither intensive nor extensive input is on offer in a typical instructed setting. In fact, input is limited to a few hours a week, and sometimes to as little as 30-45 minutes a week, as is the case in many primary schools in England. In such a scenario, fast and resource-intensive explicit learning is more effective. Adolescents and adults rely on explicit learning to a much greater extent, which explains their comparative advantage over younger children in instructed settings. But can children draw on explicit processes as well? Recent research has shown that the success of explicit learning is associated with language learning aptitude and metalinguistic awareness (Roehr-Brackin, 2018; Tellier & Roehr-Brackin, 2017), which are still developing in young children. So what is the role of these factors in children exposed to explicit teaching that may foster explicit learning?

In their study, Roehr-Brackin & Tellier (2019) investigated the role of language-analytic ability in primary-school children's foreign language learning. Language-analytic ability is the ability to treat language as an object of analysis and arrive at linguistic generalisations, and it can thus be linked with both language learning aptitude and metalinguistic awareness. Researchers have suggested that as young children tend to learn implicitly, they will mostly rely on the memory component of aptitude, while language-analytic ability will only begin to play a role in learners from around age 12 onwards. Unlike previous research, Roehr-Brackin & Tellier (2019) looked at the relationship and development of aptitude and metalinguistic awareness over the long term in order to examine the role of these two constructs in children's L2 achievement.

Intact Year 4 classes (ages 8-9) at five primary schools in England participated in the study, and data were analysed from a total of 111 monolingual English-speaking children. The study was conducted over one school year during which the children had 75 minutes of language input per week. The children were divided into four groups and consecutively taught two languages. In Phase 1 (16 weeks), each group was taught, respectively, German, Italian, Esperanto, or Esperanto with a focus-on-form element. In Phase 2 (16 weeks), all four groups were taught French with a focus-on-form element. The children completed tests of language learning aptitude and metalinguistic awareness at the beginning and end of Phase 1 and a test of French proficiency comprising listening and reading comprehension, writing and grammar sections at the beginning and end of Phase 2.

Roehr-Brackin & Tellier (2019) report that language learning aptitude and metalinguistic awareness were significantly correlated in the children. The association grew from medium to strong over Phase 1 of the study. The children made significant improvements with a medium effect size on the tests of aptitude and metalinguistic awareness over Phase 1 of the study, and scores at the two testing times (beginning and end of Phase 1) were strongly correlated. Aptitude test scores (at the beginning of Phase 2) and gains in French proficiency were significantly correlated at a medium level of strength. Furthermore, aptitude scores significantly correlated with gains in French reading, grammar and listening, but not writing. Finally, children's scores on aptitude subtests assessing

language-analytic ability and phonetic coding ability significantly predicted achievement in French, with language-analytic ability the strongest predictor.

The findings show that aptitude and metalinguistic awareness were still developing in the participating 8 to 9-year-old children, which is in line with expectations. As children have not yet reached cognitive maturity, their language-related abilities are still dynamic. The correlations between scores at the two testing times indicate that children performing strongly on the first occasion also performed strongly on the second occasion. Hence, although aptitude and metalinguistic awareness were still developing, they can still serve as potential predictors of L2 success.

This was confirmed by the finding that two aptitude components significantly predicted children's progress in L2 French. Interestingly, and contrary to theoretical argumentation and empirical results from previous research, language-analytic ability was most important in this context. In this sense, the participants showed an 'adult' pattern rather than a typical 'child' pattern. This result is not in keeping with the argument that young learners primarily draw on memory ability and/or implicit learning processes. Consequently, Roehr-Brackin & Tellier (2019) argue that it may not be relative cognitive maturity alone that determines children's approach to L2 learning. Experiencing explicit, form-focused instruction may foster the role of language-analytic ability even in children as young as 8 to 9 years. Hence, the type of instruction young learners experience may be as important as chronological age in determining the use of primarily implicit vs. primarily explicit learning. The more form-focused the instruction the learners experience, the more relevant language-analytic ability seems to become. Thus, explicit learning can be important even in childhood. If substantiated in a replication, this finding has potentially powerful implications for L2 instruction at primary school.

Approaches to replication

Replication studies may be close, approximate or conceptual in nature (Porte & McManus, 2019). A close replication stays as true to the original as possible by making only a single change. It is acknowledged that an exact replication cannot be achieved in a research field such as SLA, since amongst other things, participants' characteristics will necessarily be different. Even if the same

learners were to be recruited for a replication that is carried out shortly after the original study, the learners would no longer be at the exact same level of L2 proficiency.

This factor is compounded in research with children or adolescents, since young learners continually mature and are still developing in terms of their cognitive capacity; the younger the participant, the more profound the impact, i.e. the same child will have very different abilities at the age of 8 and at the age of 10, for instance. Hence, the single change deliberately introduced in a close replication will necessarily be accompanied by inevitable changes in participant characteristics. An approximate replication will still be as close to the original study as possible to ensure comparability, yet not just one, but two variables may be adjusted. Finally, in a conceptual replication a number of variables may be changed in order to broaden the generalisability and implications of the findings with a view to achieving greater theoretical relevance.

In what follows, two replication studies of Erlam (2005) are recommended, the first a close replication, and the second an approximate replication. Furthermore, a conceptual replication of Roehr-Brackin & Tellier (2019) is proposed.

Replications of Erlam (2005)

Erlam's (2005) study draws on both early research and more recent developments in the field of language learning aptitude. She acknowledges that the Pimsleur Language Aptitude Battery (PLAB; Pimsleur, 1966) is generally considered suitable for adolescent learners, compared with the classic Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959; Carroll & Sapon, 2002b), which was designed for adult learners. Following proposals that conceptualise language learning aptitude in terms of a three-component model comprising phonetic coding ability, language-analytic ability and memory (Skehan, 2002, 2016), the researcher takes care to include the construct of working memory in her operationalisation of aptitude, in line with recent developments in the field (Doughty, 2019; Miyake & Friedman, 1998; Wen et al., 2017).

While two different approaches to replication are detailed below, both should address two acknowledged methodological weaknesses of the original study. Specifically, all measures of

participants' cognitive ability need to be administered prior to any experimental treatment, since they serve as predictors for any observed improvements or lack thereof in language skills following instructional intervention. This is particularly important in the case of younger participants whose cognitive abilities are still developing. Erlam's (2005) 14-year-old learners were arguably close to cognitive maturity and their cognitive capacities had therefore reached a potentially more stable state, but nonetheless the six-month delay between experimental intervention and aptitude measurement in the case of the original study could have had a confounding impact.

Furthermore, and dictated by practical circumstances, the original study relied on a somewhat unconventional approach to measuring participants' working memory. The researcher used a group-administered multi-syllabic word test with a written response format. While an argument for the validity of this approach is put forward (Erlam, 2005, p. 154), the measure arguably fails to meet the usual standards required of tasks aimed at assessing online storage and processing abilities. A wide range of working memory tests is available, and they are all administered individually to allow for oral responses, regardless of whether stimuli are presented auditorily or visually (Juffs & Harrington, 2011; Linck, Osthus, Koeth, & Bunting, 2014; Williams, 2012).

By contrast, Erlam (2005) employed a written response format to facilitate group administration of her test. The group-administered presentation of stimuli was timed, but individuals' reading times are known to vary. This means that the faster readers could potentially rehearse the stimuli or employ memory strategies to improve subsequent recall performance. These are confounding factors that working memory tests should seek to prevent. By the same token, the timed written response format likewise could not take individual differences in processing speed into account, again allowing for a potential confound of conscious, strategic thinking by (some) learners. Erlam (2005, p. 153) states that she would have preferred to use a non-word repetition task. This, or any other accepted measure of phonological loop capacity such as a forward digit span task, would be suitable for a replication study, provided that the measure is administered individually and any opportunity to take a strategic approach to the task is thus curtailed.

A final remark is in order on a point that Erlam (2005, p. 167) presents as a limitation of her study. The researcher notes that her design does not fully match an ATI paradigm, according to which participants should be allocated to specific instructional conditions based on their aptitude profiles. This was not possible due to the fact that the researcher worked with intact classes. However, it could be argued that Erlam's (2005) ecologically valid research design was actually the better approach. Working with intact groups may not be as theoretically tidy as a strict ATI paradigm, but the findings arising from such research are more informative for subsequent practical applications. In school settings, it is rarely (if ever) the case that learners are screened for aptitude and streamed precisely in accordance with their individual profiles, so they can be matched with a particularly favourable instructional approach. Quite to the contrary, it is much more likely that groups of learners with different ability levels find themselves in the same language classroom. Therefore, it is of particular value to both teachers and learners themselves to know if a specific instructional approach might be equally suitable for all because it minimises the impact of individual learner differences.

Close replication of Erlam (2005)

Given the potentially highly impactful finding of the original study with regard to the apparent neutralising effect of deductive instruction, it is important to establish that this result is not restricted to the grammar point targeted, that is, the form and appropriate use of French direct object pronouns. A close replication study should therefore target a different structure while keeping all other variables the same, with the exception of the adjustments recommended above which are required to ensure validity, i.e. administration of the aptitude measures prior to treatment and the use of an individually administered working memory test with an oral response format.

The choice of a new target structure should be informed by the participants' language course syllabus. In other words, the target should be either a grammar point which the learners have not encountered before, but which is coming up in their syllabus, or a grammar point which the learners have studied before, but which has not been mastered yet, as evidenced by repeated and persistent inaccuracy or variability in its use. Suitable structures of L2 French might be, for instance, subject-verb agreement, the form and use of different tenses, or the choice of the appropriate auxiliary verb in

compound tenses. Selecting a new target structure means that the researcher will need to redesign the language tests that are to be used as outcome measures. The test format can be retained for each of the measures of the four skills, but content and scoring would need to be adjusted.

Approximate replication of Erlam (2005)

Over and above a close replication as outlined in the preceding paragraphs, an approximate replication is also recommended. Such a study could, but does not have to, follow on from a preceding close replication. The purpose of the approximate replication would be similar to the purpose of the close replication, that is, to ensure applicability of findings beyond the structure targeted in the original study. However, instead of choosing a different L2 French target structure, in an approximate replication the researcher could opt for a different L2 and thus necessarily also for a different target structure or structures, or they could retain the focus on L2 French, but would include more than one new target structure.

If more than one target structure is included, a principled decision as to which grammar points to select will need to be made. As in the case of a close replication with a single new target structure, the researcher's choice must be informed by the participants' language course syllabus. However, it is likely that there will be a number of potential targets. In that case, learning difficulty of the grammar points would be an appropriate determining factor. The field offers detailed discussions of how to determine subjective and objective learning difficulty of different linguistic features (Collins, Trofimovich, White, Cardoso, & Horst, 2009; DeKeyser, 2005; R. Ellis, 2006; Housen, Pierrard, & Van Daele, 2005; Roehr & Gánem-Gutiérrez, 2009; Spada & Tomita, 2010). Drawing on this research, the choice should include one or more 'easy' and 'difficult' targets. This would ensure a useful level of comparability and generalisability, as each target could be considered representative of other structures of a similar difficulty level.

An approximate replication carried out along these lines would require the researcher to redesign the language tests to be employed as outcome measures. In addition, the time of the experimental intervention would have to be adjusted to do justice to the increased number of targeted

grammar points. In other words, instead of a treatment of 3 x 45 minutes as used in the original study for a single target feature, 6 x 45 minutes may be required in a replication study with two target features, for instance. However, it should be noted that it will probably not be necessary to automatically double the treatment time with each additional feature, especially if the targets appear in similar communicative contexts.

A further option for an approximate replication would be to retain the focus on L2 French and include new target structures while working with participants who speak a first language (L1) other than English. Such an adjustment would serve to rebalance the scales somewhat in favour of speakers of languages other than English, given that existing work in the field is heavily biased towards English, primarily spoken as an L2 (Norris & Ortega, 2001), to be sure, but even so it is desirable to have a broader range of language combinations represented in SLA research, since this will increase the generalisability of findings. All the methodological adjustments pointed out so far would need to be made, but additionally, further changes to the instruments would be required. Specifically, the aptitude measures used in the original study were drawn from the PLAB and MLAT batteries, both of which are designed for L1 speakers of English. The MLAT has been translated into and validated for a small number of other languages, so these alternative test versions could be drawn on if they are available for the target L1. If suitable aptitude measures that are not translations of the MLAT exist for the L1 in question, then this would be another option. If no such measures are available, the researcher could draw on the language-neutral LLAMA aptitude test battery (Meara, 2005; Rogers et al., 2016), which is freely available online. Whereas a change in aptitude measures would remove the replication study one step further from the original and thereby make subsequent comparisons more problematic, it should also be borne in mind that the adult MLAT sub-test used in the original study showed low reliability, conceivably because it was very difficult for the adolescent participants (Erlam, 2005, p. 153). For this reason, a change in aptitude measures could be seen as an advantage, although naturally reliability would have to be established in the replication study as well.

Theoretical considerations

Two remarks are in order in relation to both types of replication as outlined in the previous paragraphs. The first point concerns the operationalisation of instructional treatments in Erlam's (2005) original study. The researcher compares three approaches which are labelled "deductive", "inductive" and "structured-input". The deductive condition comprises explicit instruction on the target structure with rule explanation, form-focused activities, output practice and corrective feedback. In the inductive condition, no explicit metalinguistic explanations are given, but participants engage in practice activities designed to encourage hypothesis-testing about the target structure; participants also engage in output practice and consciousness-raising activities, but they are not told to search for rules or underlying patterns. In the structured input condition, which is loosely modelled on a processing instruction paradigm (DeKeyser & Prieto Botana, 2015; VanPatten, 2004), explicit metalinguistic information is provided, learners participate in aural and written input-based activities, though there is no output practice; they take part in an error identification activity and corrective feedback is provided.

While the activities used in the deductive and structured-input conditions are in line with the characteristics conventionally associated with these approaches, the inductive condition was operationalised slightly differently from what is often understood by such an approach in current research. The researcher points out that learners were not actively encouraged to search for underlying rules or patterns in the input (Erlam, 2005, p. 165), thus treating the approach as what might be termed an implicit teaching condition.

Conversely, any teaching approach which does not present learners with pedagogical grammar rules and then asks them to apply these to exemplars, but instead proceeds by presenting language in use and then asks learners to try and identify patterns or formulate rules that capture any observed regularities is normally termed 'inductive'. In other words, deductive and inductive approaches operationalised in this manner, while ordering activities in contrasting ways, can both be described as explicit approaches to teaching (Goo et al., 2015; Norris & Ortega, 2001).

A researcher undertaking a replication study will need to be aware of the slight diversion in Erlam's (2005) 'implicit' operationalisation from what might be seen as the more theoretically precise

and perhaps nowadays more conventional 'explicit' operationalisation of the inductive approach. Accordingly, the researcher will need to arrive at a decision as to whether they should follow Erlam's approach or adapt their inductive condition to include explicit instructions to search for underlying patterns, thus placing the approach squarely in the domain of explicit teaching.

The second point to bear in mind relates to the statistical analyses employed in the original study. The analyses seem entirely suitable, are reported in detail, and should thus be replicable. However, it is of course possible that a replication study will yield a data set with different characteristics. For instance, distributions of scores may diverge from normality, or differences on pre-test scores may emerge between the experimental groups. Statistical tests would have to be adapted accordingly. Moreover, it is recommended that exact p-values are reported in all instances – something that is omitted in the original study. It would also be desirable to report effect sizes.

Finally, it is to be hoped that any replication study will identify weak-to-moderate positive correlations between the aptitude measures that are employed. This would be different from the findings reported by Erlam (2005, p. 158), who found no significant correlations, but it would contribute to the validity argument for the instruments used, since they are meant to measure sub-components of the same construct, rather than completely independent variables.

Conceptual replication of Roehr-Brackin & Tellier (2019)

Roehr-Brackin & Tellier (2019) is a recent study which took into account the most up-to-date developments in research on children's instructed language learning. The resulting findings, however, were different from what comparable research had uncovered, since the young participants' performance tended to be more in line with 'adult' patterns than 'child' patterns. Accordingly, the researchers argue that, contrary to what is often assumed, children may in fact be able to learn explicitly, provided that they are exposed to relevant teaching approaches that facilitate such learning. This argument is based on a data set obtained from a group of children who were all exposed to French teaching with a focus-on-form element in Phase 2 of the study, i.e. the second half of the school year during which data were collected. In order to establish with certainty whether a specific

instructional approach can indeed facilitate explicit learning earlier than perhaps anticipated, a comparison is required. Thus, the recommended approach in the case of this study is a conceptual replication. In other words, the research design needs to be adjusted, and as a consequence a number of variables will be changed.

The recommended research design is a quasi-experimental one which, while still working with intact groups and thus preserving ecological validity, compares the effects of a form-focused instructional treatment that is the same as or highly similar to the one taken in the original study with a meaning-focused instructional treatment that does not incorporate any planned focus on form and thus does not actively encourage the use of explicit knowledge and learning. The latter approach would mirror what commonly happens in primary-level classrooms, i.e. children encounter the target language mostly in an oral/aural mode and mostly learn playfully, e.g. via songs, games, and similar meaning-focused activities. The researcher undertaking such a replication study would need to design a suitable teaching sequence which matches the L2 French teaching materials used in the original study in terms of linguistic content, but without the metalinguistic activities that constitute the focus-on-form element. The extra time made available by omitting these activities should be devoted to further communicative language practice in terms of speaking, listening, reading and writing.

In order to allow for a fair comparison between the experimental groups experiencing a form-focused vs. a meaning-focused approach, the L2 proficiency measures should be extended to include a speaking sub-test, thus complementing the listening, reading, writing and grammar sub-tests used by Roehr-Brackin & Tellier (2019). The researcher undertaking the replication would have to bear in mind that speaking can only be assessed individually or in small groups, so additional time for testing the children needs to be planned for.

Any experimental or quasi-experimental research design should ideally include a control group. If this can be put into place in a conceptual replication study, the research design would be strengthened considerably. However, it is acknowledged that the practical constraints of conducting research in a school setting may make this impossible because it would require the school(s) involved as well as the children's parents or guardians to agree to a class or several classes of children not

having any language input during the data collection period while still participating in all the tests. This presents an ethical dilemma that is not easy to resolve, since the approach would not only expose the children allocated to the control group to a fairly onerous testing schedule, but also arguably put them at a (temporary) educational disadvantage because of the lack of language instruction that would be offered to their peers in the experimental groups.

Bearing in mind these issues, running the conceptual replication in the same way as the original study, i.e. without a control group, would seem acceptable. Just as in Roehr-Brackin & Tellier (2019), it would not allow the researcher to investigate age effects in isolation, but the comparison of two experimental groups would still enable them to disentangle instructional effects from age effects in the young participants. Put differently, without a control group, one would not be able to establish whether maturation alone can result in improved performance on measures of language learning aptitude or metalinguistic awareness. However, one would be able to establish whether and to what extent these variables predict success in either of the instructional conditions, whether developments in aptitude and metalinguistic awareness are associated with exposure to either type of instruction, and which type of instruction results in greater improvements in L2 proficiency.

If the conceptual replication study is carried out with children whose L1 is not English, the measures of language learning aptitude and metalinguistic awareness would have to be adapted accordingly. Roehr-Brackin & Tellier (2019) measured aptitude by means of the MLAT-E(UK) (Carroll & Sapon, 2002a), which is the child version of the adult MLAT and relies on the same model of aptitude. The test was designed for L1 speakers of English, although there are validated versions available in a small number of other languages, such as Spanish, Catalan and Hungarian. If the participating children have a different language background, the researcher would have to identify an alternative test.

In the original study, metalinguistic awareness was assessed by means of a specifically developed test for English-speaking children aged 8-11 which includes items on lexical semantics, morphosyntax, linguistic ambiguity and basic metalinguistic terminology. The test draws on a number of European languages including English as well as a specially constructed artificial mini-language

(Tellier, 2013). This test does not exist in other versions but could potentially be adapted for speakers of European languages other than English. Such test development would require careful preparation and piloting, of course.

Conclusion

In conclusion, a few remarks are in order which apply to all the recommended approaches to replication detailed in the preceding sections. First, researchers need to bear in mind that the proficiency level of the participants is an important variable that will impact not only on the design of the instructional materials, but also on the nature of the linguistic outcome measures. In both original studies the participants were at a low level of proficiency. The adolescents in Erlam (2005) were in their second year of French instruction and could therefore be considered elementary-level learners. The primary-school children in Roehr-Brackin & Tellier (2019) were beginners in French. In view of the small number of hours in the children's syllabus dedicated to language instruction during the school year, progress was very slow. This meant that the French proficiency test had to be designed accordingly, with writing, for instance, assessed by drawing heavily on L1 prompts and including accurate copying of French words and phrases as a task. In other words, testing options with younger, low-proficiency learners are severely restricted, and a researcher conducting a replication with similar groups of participants would have to be prepared to develop tailor-made measures that match exactly the limited input their learners have received, so the participants have a realistic chance of coping with the tests and a floor effect is avoided.

Second, both original studies under discussion in this paper relied on classic statistical analysis methods involving significance testing. While these analyses would be equally suitable for replication studies, researchers undertaking such replications may wish to use alternative methods, e.g. robust statistics drawing on R, which essentially do not require a normal distribution in the data sets to be analysed. A final decision would most likely depend on the nature of the data set as well the guidelines and recommendations of the publication outlet to which a researcher intends to submit the report on their replication study.

References

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

Carroll, J. B., & Sapon, S. M. (2002a). *Modern Language Aptitude Test - Elementary: MLAT-E*. Rockville, MD: Second Language Testing Foundation.

Carroll, J. B., & Sapon, S. M. (2002b). *Modern Language Aptitude Test: MLAT*. N. Bethesda, MD: Second Language Testing Inc.

Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some input on the easy/difficult grammar question: An empirical study. *Modern Language Journal, 93*(3), 336-353.

de Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition, 19*, 249-276.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313-348). Malden, MA: Blackwell.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning, 55*(s1), 1-25.

DeKeyser, R. M., & Prieto Botana, G. (2015). The effectiveness of processing instruction in L2 grammar acquisition: A narrative review. *Applied Linguistics, 36*(3), 290-305. doi:10.1093/applin/amu071

Doughty, C. J. (2019). Cognitive language aptitude. *Language Learning, 69*(S2), 101-126. doi:10.1111/lang.12322

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*(2), 305-352.

Ellis, N. C. (2015). Implicit AND explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 3-23). Amsterdam: John Benjamins.

Ellis, N. C. (Ed.) (1994). *Implicit and explicit learning of languages*. London: Academic Press.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning, 54*(2), 227-275.

Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics, 27*(3), 431-463.

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research, 9*(2), 147-171. doi:10.1191/1362168805lr161oa

Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443-482). Amsterdam: John Benjamins.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition, 19*, 379-400.

Housen, A., Pierrard, M., & Van Daele, S. (2005). Structure complexity and the efficacy of explicit grammar instruction. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 235-269). Berlin: Mouton de Gruyter.

Jaekel, N., Schurig, M., Florian, M., & Ritter, M. (2017). From early starters to late finishers? A longitudinal study of early foreign language learning in school. *Language Learning, 67*(3), 631-664.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(2), 137-166. doi:10.1017/S0261444810000509

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research, 24*(1), 35-63.

Li, S. (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics, 36*(3), 385-408. doi:10.1093/applin/amu054

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition, 38*(4), 801-842. doi:10.017/S027226311500042X
Lichtman, K. (2013). Developmental comparisons of implicit and explicit language learning. *Language Acquisition, 20*(2), 93-108.
Lichtman, K. (2016). Age and learning environment: Are children implicit second language learners? *Journal of Child Language, 43*, 707-730.
Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21*, 861-883. doi:10.3758/s13423-013-0565-2
Meara, P. (2005). *Llama Language Aptitude Tests*. University of Wales Swansea.
Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, NJ: Erlbaum.
Muñoz, C. (2014). The association between aptitude components and language skills in young learners. In M. Pawlak & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism: Studies in honour of David Singleton* (pp. 51-68). Cham: Springer.
Muñoz, C. (Ed.) (2006). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.
Norris, J. M., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning, 51*(1), 157-213.
Pimsleur, P. (1966). *The Pimsleur Language Aptitude Battery*. New York: Harcourt.
Porte, G. K., & McManus, K. (2019). *Doing replication research in applied linguistics*. New York, NY: Taylor & Francis/Routledge.
Ranta, L. (2002). The role of learners' language analytic ability in the communicative classroom. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 159-180). Amsterdam: John Benjamins.
Rebuschat, P. (Ed.) (2015). *Implicit and explicit learning of languages*. Amsterdam: John Benjamins.
Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47*(1), 45-99. doi:10.1111/0023-8333.21997002
Rodríguez Silva, L. H., & Roehr-Brackin, K. (2016). *Language learning aptitude and working memory as predictors of instructed adult L2 learners' explicit and implicit L2 knowledge*. Paper presented at the Cognitive Approaches to Language Pedagogy 2 conference, Basel, Switzerland, 10-11 June 2016.
Roehr, K. (2012). Aptitude treatment interaction (ATI) research. In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Acquisition* (pp. 31-35). London: Routledge.
Roehr, K., & Gánem-Gutiérrez, G. A. (2009). Metalinguistic knowledge: A stepping stone towards L2 proficiency? In A. Benati (Ed.), *Issues in second language proficiency* (pp. 79-94). London: Continuum.
Roehr-Brackin, K. (2018). *Metalinguistic awareness and second language acquisition*. London: Routledge.
Roehr-Brackin, K., & Tellier, A. (2019). The role of language-analytic ability in children's instructed second language learning. *Studies in Second Language Acquisition, 41*, 1111-1131. doi:10.1017/S0272263119000214
Rogers, V., Meara, P., Aspinall, R., Fallon, L., Goss, T., Keey, E., & Thomas, R. (2016). Testing aptitude: Investigating Meara's (2005) LLAMA tests. *EuroSLA Yearbook, 16*, 179-210. doi:10.1075/eurosla.16.07.rog
Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning, 54*(1), 35-78.
Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 319-353). Cambridge: Cambridge University Press.
Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69-94). Amsterdam: John Benjamins.

Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 17-40). Amsterdam: John Benjamins.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*(2), 263-308.

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning, 67*(4), 747-790.

Tellier, A. (2013). Developing a measure of metalinguistic awareness for children aged 8-11. In K. Roehr & G. A. Gánem-Gutiérrez (Eds.), *The metalinguistic dimension in instructed second language learning* (pp. 15-43). London: Bloomsbury.

Tellier, A., & Roehr-Brackin, K. (2013). Metalinguistic awareness in children with differing language learning experience. *EuroSLA Yearbook, 13*, 81-108.

Tellier, A., & Roehr-Brackin, K. (2017). Raising children's metalinguistic awareness to enhance classroom second language learning. In M. d. P. García Mayo (Ed.), *Learning foreign languages in primary school: Research insights* (pp. 22-48). Bristol: Multilingual Matters.

Tolentino, L. C., & Tokowicz, N. (2014). Cross-language similarity modulates effectiveness of second language grammar instruction. *Language Learning, 64*(2), 279-309.

Tomak, I. (2019). *Contributions of type of instruction, individual differences in cognitive ability and age to the development of explicit and implicit knowledge of L2 English articles: A study with Russian learners of English.* Unpublished PhD thesis, University of Essex.

VanPatten, B. (Ed.) (2004). *Processing instruction: Theory, research, and commentary*. Mahwah, NJ: Erlbaum.

Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching, 50*(1), 1-31. doi:10.1017/S0261444816000276

Wesche, M. B. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 119-154). Rowley, MA: Newbury House.

Williams, J. N. (2012). Working memory and SLA. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 427-441). London: Routledge.