Language Teaching (2008), **41:1**, 1–14 © Cambridge University Press
doi:10.1017/S0261444807004727

Replication studies in language learning and teaching: Questions and answers

Language Teaching Review Panel

Language Teaching is now encouraging original papers which replicate previous significant experimental studies in the field of language learning and teaching (http://assets.cambridge.org/LTA/LTA_ifc.pdf; this issue, pages iii–iv). While historically the journal has been a source of reference, particularly through its state-of-the-art reviews and other research surveys, within the new content (and format) from 2008, there is space for a novel, direct and unique contribution to current research agendas. With their heightened need for the critical review of previous work on the aspect in question, replication studies also accord with the fundamental 'survey' nature and objectives which *Language Teaching* has always sought to present to its readers.

The commitment to publishing replication studies reflects the editors' belief that such research should play a more significant role in the field than it has up to now and that it is both useful and necessary. The potential for replicating studies in order to validate results is a requirement of scientific inquiry and should become more prominent in establishing and confirming the outcomes of research in L2 learning and teaching. At the very least, replications can improve the way we interpret empirical research because they provide a 'second opinion' on the hypotheses, methods and/or results presented in the original work. In the following pages, members of the journal's Review Panel answer questions on the nature, application and significance of replicating studies in our field.

The editors would like to thank the following reviewers for their contributions to this paper: Rebekha Abbuhl, Rebecca Adams, Melissa Bowles, Jeff Connor-Linton, Paul Grahame Doyle, Mike Grenfell, Alison Mackey, Hossein Nassaji, Lourdes Ortega, Peter Robinson and Latricia Trites.

Why has replication not been more seriously promoted as a research approach in the past in our field and in scholarly publications?

As a field of scientific inquiry, second language acquisition (SLA) is young, having been systematically researched only since the 1970s. Therefore, much territory is still unexplored and many research questions are open for investigation. This in itself gives researchers freedom to explore and to break new ground more easily than in other, more established fields. In addition, as Albert Valdman indicates, researchers are attracted to 'original' research because of the 'aura of glamour and the exhilaration of innovation' that accompany it (Valdman 1993: 505).

Culturally, both the social sciences and society at large value the 'ground-breaking', 'innovative' and 'original'. Researchers may thus spend their limited time, energy and resources engaged in research activities which they believe will lead to the most reward

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press

in terms of, for example, tenure, promotions, grants and other accolades. Indeed, there are institutional deterrents to replication as well. In research universities, where originality is valued, faculty members whose tenure files contain many replications may be questioned by their colleagues and even considered to be uncreative (van der Veer, van IJzendoorn & Valsiner 1994). In turn, faculty may pass on this belief – that replication studies are somehow inferior – to their graduate students. As Polio & Gass (1997) suggested a decade ago, faculty should play a role in changing the tide by discussing the importance of replication in graduate seminars and Ph.D. courses, and by encouraging students to read the existing literature carefully and to conduct replication studies.

Journal editors, aware of these values, are also faced with an economic question: What to do with the scarce number of pages at their disposal? They will understandably want to devote their pages to publishing articles they believe will gain them the largest readership – original research. However, page limitations may indirectly contribute to the lack of replication studies in another way. Since, due to space limitations, article authors are not able to reproduce or discuss at length the key elements of their studies (e.g., the exact questionnaires, pre- or post-tests used), interested readers may not be given sufficient detail to be able to replicate the study anyway.

Finally, there are methodological issues. If a replication study obtains the same result as the previous original study (e.g., that some form of treatment led to significant interlanguage development), the results may be seen as trivial or as not contributing to the growth of the field. However, if the replication study does NOT obtain the same result (e.g., it found that the form of treatment used in the original study did NOT lead to significant interlanguage development), we are left with several thorny questions, including whether the diverging results were due to a methodological flaw in the original study or in the replication. Given the lack of prestige (and thus rewards) associated with replication studies, researchers may feel that their limited resources are better spent in original research than in addressing these questions, despite the fact that it is precisely the potential for problematizing (and perhaps questioning) an original study's results (and conclusions) that makes replication studies so valuable (and even necessary). Recognizing the fallibility of any research – but especially of small subject number (*n* or N) studies and still developing construct operationalizations and measurement techniques – demands replications to ensure that misleading or erroneous results and conclusions from a study do not become the nebulous foundation of subsequent theory, research and practice.

What are the relative advantages to the field of replications termed 'exact' (or 'literal'), 'approximate' (or 'systematic'), and 'constructive' (or 'conceptual')?

EXACT replications, when they obtain equivalent results to an original study, allow researchers to have a high degree of confidence in the internal and external validity of the original. However, since exact replications, by definition, involve examining the same subjects employed in the original study, this kind of replication is difficult, if not impossible, in the field of SLA (see below).

To address this problem, APPROXIMATE (or SYSTEMATIC) replications have been suggested as an alternative. Here, one key variable (such as the learners' proficiency, L1 background or

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press

learning context) is changed. If the replication yields equivalent results to the original study, we can have more confidence that the results of the original study can be generalized to, for example, a new population of learners, learning context or linguistic feature.

Like exact and approximate replications, CONSTRUCTIVE (or CONCEPTUAL) replications can be designed to assess the validity or generalizability of a study. However, they can also be used to assess the explanations that the original researcher gave to account for the results obtained in his or her study. Thus, for example, in a situation where the original researcher offered three explanations for his or her results, the researcher undertaking a conceptual replication can manipulate non-major variables, operationalizations or design features of the original study in order to evaluate (and possibly rule out one or more of) those explanations.

Assessing validity leads the researcher to question whether the findings indicate actual relationships among variables, or whether they simply reflect a particular operationalization of the constructs involved. By examining the same theoretical relationship in a different way, conceptual replications allow researchers to establish whether prior research results revealed actual relationships among constructs. However, as Hendrik (1991: 46) states, 'a conceptual replication is a high risk endeavor. When results from the two experiments are equivalent, the payoff in increased confidence in the original experiment is rich indeed. However, when the conceptual replication fails to replicate the results of the original study, it is practically worthless . . . [since] the reason for the failure cannot be known'.

What are the particular difficulties that might be encountered when attempting 'exact' replications of a previous study in our field?

If we want to repeat any investigation of language acquisition with the same learners, they will almost certainly have learned more and thus are not in the same 'state' as they were in the original experiment. This is, of course, assuming the possibility of gathering the same learners for a repeat experiment, which is far from certain. Thus, we are constrained by having to select another set of subjects. In many cases, the conditions for learning that occurred in the first experiment are unlikely to be exactly reproducible either, especially if the research was classroom-based. All these need not be viewed as constraints if we wish to REPEAT the original study with different subjects or different conditions so as to add to the dataset for the phenomenon under study; but in what sense can this then be said to be an 'exact' replication?

Ironically, one of the greatest problems is that the information provided in the original published study is incomplete and does not lend itself to adequate replication. While researchers are taught that responsible research entails full disclosure of methods, procedures and analysis, many published articles are lacking in one area or another. For example, authors often do not include in their published work questionnaires, surveys or assessment instruments that were used in the research. Other articles are limited in their reporting of accurate methodological procedures and decision-making processes. Many articles give general information about participants, omitting critical background information that would enable more accurate replication. Often when authors are faced with page or word limits in a publication, they tend to cut the text in the procedural section and often do not include relevant appendices.

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press

Since no subject population involved in a study can be exactly replicated, particularly in our field, what are the important parameters for classifying and sampling subject populations when conducting replication research?

It is important to specify what the important parameters for classifying and sampling subject populations are. However, these parameters will vary in order of significance according to the sub-field, such as language learning strategies, pedagogic grammars, interlanguage, etc. The most obvious parameters would be the ability level in second language learning and teaching, the bilingual context, age and gender. It is also important to keep the languages distinct. The temporal and spatial context, that is, over what time the study was carried out and the geographical source of the data, also need to be defined. Most of all, in replication studies, it is important to argue for and establish difference rather than aim at similarity. In other words, what we want to know in replication studies is not how similar the various subject populations were but how different, and according to what parameters. This should be judged as a positive attribute of the study, not a negative one.

How does one go about finding a suitable study to replicate in their area of interest?

Finding a suitable study to replicate is closely related to the identification of appropriate and theoretically interesting research questions. As in original studies, the research questions in replication studies need to be current; that is, they might typically be incompletely or insufficiently answered in the previous literature and thus warrant further or additional investigation. One source of research questions is often the suggestions made by other researchers at the end of their work. Despite the fact that researchers often indicate how their studies should be replicated and/or extended, these are rarely carried out. Another source of ideas is the critical analysis of existing research, which can lead to the identification of gaps that may strike a reader as important. Often, when reading an article, one might recognize that a potentially important variable has not been controlled for or that different languages might function differently on a certain important dimension. This can lead the reader to wonder what might happen in a replication. Alternatively, some controversy may have been left unresolved, and this can form the basis of a follow-up study. Such considerations often affect decisions regarding the type of replication that should be carried out.

How does one identify that a study might usefully be replicated?

In theory, of course, ANY study can be replicated; the question is whether or not the field NEEDS the replication and, of course, why. And that might require one to consider how significant the original study has turned out to be since it was conducted and published. If, for example, for some years there has been no critical address in the literature of a study one might wish to replicate, there will be a need to argue that this really is a significant study, worthy of replication. The fact that it was perhaps published in a major international journal does not in itself indicate this. What has to be stated is the projected significance for the field, apart from simply repeating the experiment, in terms of what we are likely to gain from the new (replication) research.

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press

It is fair to say that in a replica study paper the focus should always remain on the original study or studies being replicated, rather than on the present study itself. This is because we are to assume that the objective in the new study was closely to test the findings of that original in some way, rather than simply transfer the original study's methodology and procedures to produce a new study in a different context and with different subjects. In general, the more variables are changed from the original study, the more distant the new study is from what was done originally, and therefore the two become less comparable. Consequently, the new study becomes a development from the original rather than an investigation into the latter's validity, reliability and generalizability.

There are several other factors which researchers should consider when selecting a study to replicate and deciding how to conduct the replication. Clearly, the need is greatest for exact replications when important theories are supported by only a small number of studies. Approximate and conceptual replications are particularly needed when the body of research on a topic has not been carried out in multiple settings or when similar operationalizations of research constructs have been used in different studies.

A researcher might also want to consider whether there were weaknesses in a previous study that could be overcome in the design or procedure. For example, there might be a need for replication in small N and small *n* studies with a low power/Type 2 risk and/or significant, but small, effect size results. There may be variables that influenced the findings inadvertently that could be controlled or moderated in a replication. Researchers should also evaluate the generalizabilty of the study, and in particular its relevance to their context. An approximate replication can help determine whether or not the relationships found apply in other contexts.

For example, Wong (2001) conducted a partial replication of a study by VanPatten (1990) which had been interpreted as showing that L2 learners of Spanish have difficulty attending simultaneously to the form and meaning of linguistic input in the oral mode. Wong decided to investigate whether the results would hold for learners of a different L2 and whether the modality of the input would make a difference. Her study was carried out with learners of L2 English from the same institution. She looked at the participants' performance in both written and oral modes using reading and listening tasks. Although the results lent support to VanPatten's findings in the oral mode, it turned out that attending to linguistic form while reading did not negatively affect comprehension. The replication provided evidence that different results may occur with different target languages and in different modalities, thereby suggesting the importance of additional research in the area. This example illustrates one of the ways in which replication studies can contribute to the further development of theories and hypotheses about L2 learning, and it highlights the usefulness of revisiting ideas following an approach that has been modified in principled ways.

Additionally, researchers should consider how the study relates to other research reported in the literature. In particular, studies that form the basis for theoretical claims need to be replicated. Similarly, when empirical findings are not predicted by current theory or are at odds with the findings of other, similar research, replications should be considered. An exact replication can strengthen reliability claims for unanticipated findings that might not otherwise receive attention, and repeated approximate and conceptual replications can lead to the re-conceptualization of current theory. In this way, replication studies not only verify previous findings, but actually make a contribution to theory-building. Having said this,

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press

researchers should not presume that only studies with design weaknesses merit replication; all studies have weaknesses or otherwise make compromises, and replication inevitably helps to expose and remedy any weaknesses.

Should the researchers who carried out the original study be encouraged to replicate their own work or should independent replication be the norm?

While we often think of replications in terms of replicating others' work, it is equally – if not more important – for researchers to replicate their own studies. Each researcher is responsible for establishing the reliability and validity of their own research, and carrying out replication studies is an important part of that. The opportunity to replicate or improve on one's own study would encourage laudable behaviour: researchers would not feel as much need to move too soon from one topic to another, but could learn from their experience or mistakes.

The opportunity to publish self-replication studies could motivate researchers to higher standards of replicability, since they would have to 'prove' that they did things the same way (to an extent that would probably not be required of other-replications). Furthermore, why should the original researcher be the one who is not allowed to learn from his/her own mistakes? Another good reason for self-replication might be substantive revision (even reconceptualization) of one's own methods – data collection, coding, and/or analysis – resulting from responses to the original study.

It is true that replications are sometimes undertaken by researchers outside the original research team, to independently test the boundaries of reliability or generalization of the original interpretations or claims. But it is also desirable for a research team to stay with a given research program and explore its limits and ramifications. This can very appropriately be done by conducting replications and partial replications of their own previous research.

If this were one of the 'hard sciences' (let us say, medicine), and a team of researchers – who had studied and examined a particular area in detail, were familiar with it, and had developed a good methodology for analyzing it – decided to take one of their (well respected) studies and replicate it, it would probably be hailed as a good idea. Such a study might show that a drug has a significant effect on another ailment, or helps another population. The same argument can be made for one of the 'soft sciences', including SLA: a group of researchers have studied a certain area in detail, are experts in it, and have devised a useful methodology for studying it. They acknowledge that their previous study has limitations, so the next logical step would be to address those limitations.

What areas are currently most in need of replication studies in our field, and why?

Many opportunities can be seen in researching the effects of implicit and explicit conditions on language learning and the relationship between explicitness and awareness. Since relatively few SLA studies have been replicated, there are many strands of research that could benefit from replication. Many SLA studies whose findings have potential pedagogical implications are laboratory-based. Such a carefully controlled environment has the advantage of limiting intervening and/or moderating variables that could affect results, but is obviously not an

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press

authentic classroom environment. It is therefore crucial to replicate laboratory studies on the effects of implicit and explicit conditions and feedback in classrooms of various types (e.g., foreign language vs. second language) and in various settings for ecological validity. Researchers in the area of conversational interaction have recently begun conducting studies in the classroom to determine whether the laboratory-based findings for recasts and other forms of oral feedback hold true there as well (e.g., Gass, Mackey & Ross-Feldman 2005; Ellis, Loewen & Erlam 2006).

Within the same area, most SLA studies on the effects of instruction have had a classic pre-post-test design, leaving open the question of HOW instructional interventions affected learning outcomes. A review by Leow (1999) revealed that very few published studies included any online measures (that is, any measurement of what was going on while learners underwent the instruction). Since then, the use of online process measures such as concurrent verbal protocols ('think alouds') and retrospective protocols ('stimulated recalls') has increased, and recent studies have examined the qualitative data obtained from the protocols to gain insight into learners' thought processes as they interacted with L2 input. Among the results uncovered by these protocols is that learners do not seem to perceive corrective feedback on phonological features, lexis and morphology equally (Mackey, Gass & McDonough 2000) and that despite being placed in an implicit learning condition, not all learners process input equally (or implicitly) (Bowles 2003). Introspective measures are one way of tapping language learners' internal processes and, if added to the research designs of pre-post-test instruction studies, can access information that post-test scores alone cannot provide. Using mixed methods, researchers replicating instruction studies could shed light on many more such issues: learner strategies, awareness and input processing techniques among others.

Replications in the field of second language writing are virtually nonexistent. The *Journal of Second Language Writing*, for example, has published only one replication (Allison, Varghese & Mei 1999) since its inception in 1992. Replications of L2 writing studies are rare in other journals too; a recent search of *LLBA* (*Linguistics and Language Behavior Abstracts* database), for example, resulted in only four other articles in peer-reviewed journals (Devenney 1989; Adams 2003; Cobb 2003; Gascoigne 2004). This is problematic, given the degree of controversy surrounding most of the key questions in the field of L2 writing today.

For example, while many researchers would agree that some kind of intervention is necessary for the development of grammatical accuracy in L2 writing, there is considerable current debate as to whether feedback leads to significant gains in either the short-term or the long-term, and if it does, what kind of feedback is best (see e.g., Lalande 1982; Semke 1984; Robb, Ross & Shortreed 1986; Fathman & Whalley 1990; Kepner 1991; Sheppard 1992; Ferris et al. 1997; Lee 1997; Conrad & Goldstein 1999; Ashwell 2000; Fazio 2001; Ferris & Roberts 2001; Yates & Kenkel 2002; Chandler 2003; Hyland 2003; Bitchener, Young & Cameron 2005). Commenting on this debate, Ferris (2004: 49) notes that 'we are virtually at Square One, as the existing research base is incomplete and inconsistent'. The existing research base is also nearly devoid of replications, making it difficult to (a) determine to what extent the results of past studies can be generalized to new contexts or learners and (b) fully understand the extent to which feedback impacts the development of L2 writing skills.

It would be good to establish a list of the subject areas and issues within them most suitable for replication studies. Conventionally, it is often felt that statistical approaches are most

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press

readily replicated, but this is not necessarily so. In many ways what needs to be replicated to begin with is the construction of the research object (e.g. L2 writing, L2 attention, L2 anxiety). Often this is redefined in order to take account of local contexts. Furthermore, there is a lack of consensus on basic terms of analysis. It would help to take original concepts (e.g. language acquisition) and test them further rather than seeking to redefine them from personal perspectives. This search for novelty is endemic in the academic field at the moment: in order to be original, researchers often simply try to become novel.

How much information should reasonably be presented in a replication paper to permit satisfactory interpretation and further replication?

In reporting original studies, it is generally agreed that researchers should provide sufficient detail for others to be able to replicate their studies. In writing up replications, however, researchers can refer to the original studies and compare and contrast them with the replications without feeling the need to repeat information that has already been provided. Given the usual space limitations in journals, researchers carrying out replications may be best advised to focus on describing what changes they have made or how their studies differ in terms of methodology or participant population, while providing other details in footnotes and appendices. The results of the original study should also be compared to those of the replication in the discussion section, focusing on how and why they are similar or different.

A few journals are already taking steps that promote this kind of reporting without compromising space. *Language Learning & Technology*, for example, encourages 'articles that take advantage of the electronic format by including hypermedia links to multimedia material both within and outside the article' (see http://llt.msu.edu/contrib.html). *Applied Linguistics*, which is published both in print and online, offers another compromise, adhering to a strict 9,000 word limit in the print version, but indicating that '[a]dditional material can be made available in the online version of the article' (http://www.oxfordjournals.org/applij/for_authors/index.html).

Polio & Gass (1997) recommend that researchers should consider submitting detailed appendices for publishers to keep online or in hard copies. These appendices may include

> (a) Detailed guidelines and examples of coding categories, (b) A listing of examples that were excluded from consideration, (c) Measures of proficiency (descriptions of tests where test security is a problem), (d) Instruments for data elicitation, including pre-tests and post-tests, (e) Experimental protocols and instructions to subjects, and (f) Demographic background of subjects. (Polio & Gass 1997: 506)

If publishers have not yet embraced this idea, authors may post such information on their personal web pages or mention in the papers what information is available to readers through personal contact. CHILDES (the Child Language Data Exchange System, available at http://childes.psy.cmu.edu) provides linguists with opportunities to utilize digital archives and allows access to 'a database of transcripts, programs for computer analysis of transcripts, methods for linguistic coding, and systems for linking transcripts to digitized audio and video'. Through this type of data-sharing system, the validity of a replication study as well as the issues dealt with in the study can be discussed further, encouraging researchers to implement methodologically sound and theoretically important research in the years to come.

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press

Language Teaching is able to store details of replication studies on the journal's web pages and make them accessible to subscribers.

What range of subject populations and stimulus domains would an initial finding have to be replicated across in order to be characterized as 'established'?

In general, the more varied and numerous the subject populations and stimulus domains are in replications repeating the findings of an original study, the more the findings gain in terms of the reliability and generalizability that allow them to be considered 'established.' With respect to a particular area of research, it is important to consider carefully which specific factors should be both controlled and varied in replication studies. In the field of SLA, the list of factors to control and vary systematically should probably include (but not be limited to):

- the age of the participants (e.g., children of a range of different ages, young adults, older adults, elderly adults)
- language learning contexts and the learners' purposes within those contexts (e.g., second- and foreign-language classrooms, immersion situations, study abroad contexts, university settings, private language institutions, etc.)
- research settings (e.g., classroom vs. laboratory, a combination of the two)
- learners' L2 proficiency and developmental levels
- type of linguistic target (e.g., lexical, phonological, morphological, syntactic, or pragmatic; perceptually salient or non-salient, structurally more or less complex, pedagogically more or less difficult to explain)
- type of modality or task (e.g., aural vs. written, comprehension vs. production, relatively constrained/prompted vs. free/spontaneous).

An initial finding from an (internally valid) original study can be considered 'established' when consistent findings are gained from a series of replication studies investigating, for example, AT LEAST one other language, AT LEAST one different subject population (selected with appropriate measures of L2 proficiency across studies), and AT LEAST one different modality. Since language grammars or structures may differ greatly with respect to the explicitness, frequency, and importance of their functions, it would be ideal for the field if relatively high numbers of replication studies with different linguistic forms were required before an initial finding focusing on one particular target could be characterized as 'established'. Replications are also important in that different modalities activate different cognitive processes in L2 learners, and linguistic behaviours may differ across learner populations due in part to their prior linguistic experiences.

If a study is replicated several times, employing different groups of participants under different conditions with systematic changes to one or more of the factors listed above, and if the results are basically the same across situations, that increases our confidence in the reliability of the original findings and supports their external validity. On the other hand, consistent failure to replicate or confirm an initial finding can raise serious questions as to its value and can lead researchers to abandon it eventually. This not only helps to make science

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press

a 'self-correcting discipline', but can also help to uncover scientific problems (Goodwin 1998: 10). On the other hand, from a practical and realistic standpoint, in many areas of current SLA research, statements of findings are often qualified as being 'established WITH RESPECT TO' one or more of the factors above, as opposed to being considered established across the board. Indeed, while it is interesting and useful to find the same results across multiple studies with different populations, it is also interesting to find one set of consistently repeated results with one narrowly defined population and another set of repeated results with another population. This can also help to move researchers toward explanations of their findings.

How does the issue of replication relate to the broader issue of generalizability of findings to different populations, and types or levels of a type of stimulus domain?

The issue of replication is closely connected with that of generalizability, or what has been known as external validity. Generalizability refers to the degree to which the findings of a previous study can be extended to similar or different populations, settings, tasks, domains, etc. One way of evaluating the generalizability of a study, and hence establishing its external validity, is through replication. If the findings of a study can be replicated and repeated in similar or different contexts, that study can be taken to have high generalizability; otherwise, the results may have only limited or no generalizability. Judgements about generalizability of research results are sometimes based on results obtained from the use of inferential statistics or tests of significance in quantitative research. However, although significance tests, if appropriately conducted, may provide information about the status of the research hypothesis, a number of social scientists have argued that this approach does not tell us much about generalizability of research findings (e.g., Carver 1978; Neale & Liebert 1986; Rosnow & Rosenthal 1989; Thompson 1992, 1994). Therefore, they suggest that the best way of determining generalizability of research results is through replication. It may also be argued that generalizability is not the goal of all research. For example, it may be said that qualitative research and case studies do not intend to generalize their findings beyond their specific cases or contexts. However, some researchers believe that even qualitative research should meet the criterion of generalizability although the concept may apply differently to such research. Firestone (1993), for example, examined three arguments for generalization of findings: generalization from sample to population, generalization to theory (utility of research findings for building or expanding theories), and generalization from case to case (transferability of findings). He argued that while the first one may be applicable only to quantitative research, the other two can apply to qualitative research. McDonough & McDonough (1997) also discussed three senses of generalizability: one relating to the degree of applicability of findings in other contexts, another relating to the degree of accuracy of statements made about specific problems under investigation, and the third – the utility of findings for practical purposes. If that is the case, it can be argued that the notion of replication is also relevant to qualitative research including case studies because replication can be used to show whether the findings of a case study have any broader utility by comparing the findings to those of other similar or different cases or situations.

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press

Why should replication become a more accepted way of proceeding in empirical research in our field?

In a cross-disciplinary view of text and discourse analysis, Titscher et al. (2000) observed that:

> A fundamental rule for any scientific work says that the manner in which the results were obtained must be verifiable. This requirement derives, in essence, from the postulate that scientific discovery is not merely self-discovery: RESEARCH MUST BE GENERALIZABLE AND TRANSPARENT, AND (WHERE POSSIBLE) CAPABLE OF BEING REPLICATED AND REPEATED. (Titscher et al. 2000: 11; emphasis added)

Replication is a method of validating research in the sciences, and is considered a criterion for the acceptance of new theories and knowledge. In the context of testing hypotheses through empirical research, replication requires experiments to be repeated in such a way that investigators can be reasonably certain that the results of two experiments are comparable and, therefore, that they are measuring the same phenomenon. To ensure that the original experiment has been accurately reproduced, the same methodological procedures must be used and the same control of variables applied, otherwise investigators are vulnerable to the contention that they have measured something different in the second experiment. In addition, there is the requirement that investigators measure the same aspects of the phenomena under investigation. If a set of measurements differs from an earlier set, further experiments may be required to establish whether a hypothesis has been disproved or a particular experiment was insufficiently rigorous.

This approach to investigation is usually referred to as the SCIENTIFIC METHOD and, while it has come under attack as a misconception of what scientists actually do by philosophers such as Lakatos, Kuhn and Feyerbrand, nevertheless within scientific communities it is still deemed to be the acceptable way of doing empirical research. If applied linguistic research is to be considered 'scientific work' in the sense used here, rather than 'pre-scientific' (Brumfit 2005: 40), it has to demonstrate greater concern for issues such as replication.

Replication can also improve upon previous research designs, measures and tasks. Polio & Gass (1997: 500) have noted that the lack of attention to replication in the field of SLA is 'particularly striking when one considers that most of the published studies base generalizations on relatively few subjects'. It is often assumed that if a study has investigated a question before, there is no need to investigate it further. Therefore, researchers usually look for new questions that have not been previously investigated. However, as Cohen (1990) notes, no single study, no matter how successfully it is carried out, can provide any definitive answer to a research question, and it is through replication that any research problem can be satisfactorily dealt with or resolved. Replication is even more needed today than before given that our field is becoming increasingly diverse in scope and investigation of topics, resulting in divergent and at times fragmented research results. Replication studies can verify and consolidate previous findings, helping results to be converged and extended. Thus, they must be more valued, encouraged and carried out in our field.

What are 'meta-analyses' of replications, and what areas of research in our field might currently benefit from such information?

A META-ANALYSIS is a study that collects data from previous, related research studies and analyzes these compiled data as one. A major benefit from this undertaking is that it alleviates

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press

problems that often plague researchers when analyzing their data from smaller studies. As many students struggle to make sense of the data they have collected for class projects, they may need to be constantly reassured that their inability to accept or reject a research hypothesis has less to do with their study's design and more to do with the limitations of time, a small number of participants, and those participants' lack of engagement. In addition, they need to be told that many researchers encounter these same frustrations.

While a literature review may simply be a compilation of research related to the current topic of research in order to find a niche for original research, a quality meta-analysis carefully examines the methodologies, construct definitions, results and findings to determine a 'goodness of fit' for these previous studies as they relate to the current research question, which then serves as a validation measure. While attempting to merge studies, creating robust numbers and homogeneous data samples, a meta-analysis serves as an additional filter for analyzing and interpreting the results. By pooling data from a collection of replication studies, a meta-analysis assists in the generalizability of the findings to a broader population.

There are several areas of research that would benefit from meta-analyses, particularly action research studies, case studies, diary studies and other longitudinal observation studies. Each of these modes of research requires extensive time and in-depth analysis of a very small sample and seeks to address issues inherent in language learning, which is clearly not unique to only one learning environment, case, or subject. Replication studies of these sorts of research are often conducted in an effort to prove or disprove findings. Conducting meta-analyses of these replication studies not only aids such generalizability, but also goes one step further in that these replication studies are compiled into one, allowing them, and their results, to be discussed in terms of a single study.

How can more replication research be encouraged in our field?

It is necessary to change the mindset of the profession. While students are sometimes allowed to do replication studies, most professionals are discouraged from doing so. In a discussion with an international student, this reviewer was informed that such a devaluation of replication studies is globally pervasive. Teachers are encouraged to do research; however, they are quickly discouraged from any sort of replication.

Our field trains researchers to find a niche, looking for that unique angle that will set their research apart. Instead, we need to see that verification of research and actually questioning the results of previous research will only strengthen the quality of research that is eventually produced. In the field of language teaching, we tend to take research such as classroom-based research and action research at face value, never questioning the validity of the questions being asked or the reliability and generalizability of the findings.

As the field of SLA matures and the base of research findings grows, the need for replication is becoming increasingly apparent. The stage is being set for replication to flourish in years to come, with some authors already heralding its importance (e.g., Mackey & Gass 2005). It is hoped that in this positive environment, new and established SLA researchers alike will appreciate the inherent value of replication and contribute their research accordingly.

One has to say that replication studies have not been given sufficient recognition in language teaching and learning research, and this is partly a product of the prevailing

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press

academic discourse and culture. In other words, everyone feels that they have to collect their own data rather than use another's. It would be good if key journals hosted Special Issues on replications studies, possibly dealing with specific fields. There might also be 'Replication strands' to the Special Interest Groups of various conferences. Indeed, it might be possible to have explicit 'replication' seminars, if not conferences. Web-based networks lend themselves to establishing 'work-in-progress' sites where replication studies could be reported. *Language Teaching*'s contribution to this agenda is a regular section designated to the presentation of such work. Finally, one senses that research-funding bodies do not assign sufficient value to replication studies. It is hoped that their policy statements will in due course express explicit support for replication studies in the field of language learning and teaching.

REFERENCES

Adams, R. (2003). L2 output, reformulation and noticing: Implications for IL development. *Language Teaching Research* 7.3, 347–376.

Allison, D., S. Varghese & W. S. Mei (1999). Local coherence and its limits: A second look at second sentences. *Journal of Second Language Writing* 8.1, 77–97.

Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing* 9.3, 227–257.

Bitchener, J., S. Young & D. Cameron (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing* 12.3, 191–205.

Bowles, M. (2003). The effects of textual input enhancement on language learning: An online/offline study of fourth-semester Spanish students. In P. Kempchinsky & C. E. Piñeros (eds.), *Theory, practice, and acquisition: The 6th Hispanic Linguistics Symposium and the 5th Conference on the Acquisition of Spanish and Portuguese*. Somerville, MA: Cascadilla Press, 395–411.

Brumfit, C. (2005). Teacher professionalism and research. In G. Cook & B. Seidlhofer (eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson*. Oxford: Oxford University Press, 27–42.

Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review* 48, 378–399.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing* 12, 267–296.

Cobb, T. (2003). Analyzing late interlanguage with learner corpora: Quebec replications of three European studies. *The Canadian Modern Language Review* 59.3, 393–423.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist* 45, 1304–1312.

Conrad, S. M. & L. M. Goldstein (1999). ESL student revision after teacher-written comments: Text, contexts, and individuals. *Journal of Second Language Writing* 8.2, 147–179.

Devenney, R. (1989). How ESL teachers and peers evaluate and respond to student writing. *RELC Journal* 20.1, 77–90.

Ellis, R., S. Loewen & R. Erlam (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition* 28, 339–368.

Fathman, A. K. & E. Whalley (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (ed.), *Second language writing: Research insights for the classroom*. Cambridge: Cambridge University Press, 178–190.

Fazio, J. L. (2001). The effect of corrections and commentaries on the journal writing accuracy of minority- and majority-language students. *Journal of Second Language Writing* 10, 235–249.

Ferris, D. R. (2004). The 'grammar correction' debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime . . . ?). *Journal of Second Language Writing* 13.1, 49–62.

Ferris, D. R., S. Pezone, C. R. Tade & S. Tinti (1997). Teacher commentary on student writing: Descriptions and implications. *Journal of Second Language Writing* 6.2, 155–182.

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press

Ferris, D. R. & B. Roberts (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing* 10.3, 161–184.
Firestone, W. A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher* 22, 16–23.
Gascoigne, C. (2004). Examining the effect of feedback in beginning L2 composition. *Foreign Language Annals* 37.1, 71–76.
Gass, S., A. Mackey & L. Ross-Feldman (2005). Task-based interactions in classroom and laboratory settings. *Language Learning* 55.4, 575–611.
Goodwin, C. J. (1998). *Research in psychology: Methods and design* (2nd edn.). New York: John Wiley & Sons Inc.
Hendrik, C. (1991). Replications, strict replications, and conceptual replications: Are they important? In J. W. Neuliep (ed.), *Replication research in the social sciences*. Newbury Park, CA: Sage Publications, 41–49.
Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System* 31.2, 217–230.
Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *The Modern Language Journal* 75.3, 305–313.
Lalande, J. F. (1982). Reducing composition errors: An experiment. *The Modern Language Journal* 66, 140–149.
Lee, I. (1997). ESL learners' performance in error correction in writing: Some implications for teaching. *System* 25.4, 465–477.
Leow, R. P. (1999). The role of attention in second/foreign language classroom research: Methodological issues. In F. M-G. J. Gutiérrez-Rexach (ed.), *Advances in Hispanic linguistics: The 2nd Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla, 60–71.
Mackey, A. & S. Gass (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum.
Mackey, A., S. Gass & K. McDonough (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition* 22.4, 471–497.
McDonough, J. & S. McDonough (1997). *Research methods for English language teachers*. London: Arnold.
Neale, J. M. & R. M. Liebert (1986). *Science and behavior: An introduction to methods of research* (3rd edn.). Englewood Cliffs, NJ: Prentice-Hall.
Polio, C. & S. Gass (1997). Replication and reporting: A commentary. *Studies in Second Language Acquisition* 19, 499–508.
Robb, T., S. Ross & I. Shortreed (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly* 20.1, 83–95.
Rosnow, R. L. & R. Rosenthal (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist* 44, 1276–1284.
Semke, H. D. (1984). Effects of the red pen. *Foreign Language Annals* 17.3, 195–202.
Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal* 23.1, 103–110.
Thompson, B. (1992). Two and one-half decades of leadership in measurement and evaluation. *Journal of Counseling and Development* 70, 434–438.
Thompson, B. (1994). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. *Journal of Personality* 62, 157–176.
Titscher, S., C. Meyer, R. Wodak & E. Vetter (2000). *Methods of text and discourse analysis*. London: Sage Publications.
Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning* 46.2, 327–369.
Valdman, A. (1993). Replication study. *Studies in Second Language Acquisition* 15, 505.
Van Der Veer, R., M. van IJzendoorn & J. Valsiner (1994). General introduction. In R. van der Veer, M. van IJzendoorn & J. Valsiner (eds.), *Reconstructing the mind: Replicability in research on human development*. Norwood, NJ: Ablex, 1–9.
VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition* 12, 287–301.
Wong, W. (2001). Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition* 23.3, 345–368.
Yates, R. & J. Kenkel (2002). Responding to sentence-level errors in writing. *Journal of Second Language Writing* 1, 29–47.

https://doi.org/10.1017/S0261444807004727 Published online by Cambridge University Press