CrossMark

Lang. Teach. (2015), **48.4**, 545–555 © Cambridge University Press 2015
doi:10.1017/S0261444815000038 First published online 20 February 2015

Replication Studies

Replication in interaction and working memory research: Révész (2012) and Goo (2012)

Susan Gass Michigan State University, USA
gass@msu.edu

Lorena Valmori Michigan State University, USA
valmoril@msu.edu

This paper argues for the replication of two studies, both of which consider feedback and working memory. In the first part of this paper, we discuss the role of interaction-based research and working memory research in second language acquisition research. We then describe two studies that have unified these two areas in recent published articles and discuss ways that replication can further our knowledge in how working memory capacity can impact the benefits of feedback. We emphasize the importance of replication in a number of different areas, including grammatical complexity of the target structure and the nature of working memory tests, in scoring, actual tests, and modes of delivery. Additionally, we discuss the possibilities for replication in the area of population differences, in particular, age groups and more lasting effects as would be found in delayed post-tests. Finally, we point out another means of gathering information about thought processes and suggest the value of adding stimulated recalls to the existing data.

1. Introduction

In this paper we make the argument for the replication of two recent studies in the sub-area of second language acquisition (SLA) research known as input and interaction, or more specifically for the purposes of this paper, interaction-based research. Interaction-based research has taken on a central role in the discipline with numerous articles (including meta-analyses) and books on various aspects of this topic.

With this in mind, we have selected two recent articles (both published in 2012) that are important in moving the field forward as we seek explanations for differences among individuals in the way they benefit from interaction. We have selected these two articles for a number of reasons: (1) they represent recent directions in the field, (2) they address an important construct (working memory capacity); (3) they bring to the forefront the construct of working memory capacity, which can be measured in different ways, thereby highlighting

https://doi.org/10.1017/S0261444815000038 Published online by Cambridge University Press

methodological issues, as noted by Révész (2012); and (4) they present detailed descriptions of all aspects of the studies and include tasks and scoring procedures which facilitate replication. In sum, we focus on the role of working memory capacity as it relates to interaction. The shared area of investigation, yet with different designs, participants, age groups and grammatical structures, raises new issues which will undoubtedly move the field forward in understanding the role of interaction (particularly recasts) in second language (L2) learning.

In the following sections, we provide background on interaction-based research and on working memory. We then turn to a description of the studies to be replicated, conceptualize the replication, and make concluding remarks concerning replication in general.

2. Background

2.1 Interaction-based research

SLA research is approached from many disciplinary directions. One area that has been central to an overall understanding of how second and foreign languages are learned is interaction-based research, finding its beginnings in L2 circles in the mid-1970s (see Ferguson 1971; Wagner-Gough & Hatch 1975; Long 1980 and overviews by Gass & Mackey 2006, 2007; Mackey 2012; Mackey, Abbuhl & Gass 2012). Interactionist research is without a doubt at the forefront of L2 research, with numerous empirical studies (see Mackey 2007) investigating the role of interaction in language learning, books supporting (Gass 1997; Mackey 2012) and critiquing (Block 2003) the interactionist approach, meta-analyses on various parts of the approach (Nicholas, Lightbown & Spada 2001; Keck et al. 2006; Russell & Spada 2006; Mackey & Goo 2007; Li 2010; Lyster & Saito 2010; Goo & Mackey 2013; Lyster, Saito & Sato 2013), and entries in prominent handbooks and encyclopedias.

Early research in this area was primarily descriptive and included detailed analyses of conversational structures involving native speakers (NS) and non-native speakers (NNS) with a focus on input to learners and interactions in which they engaged. The emergence of two constructs – output (e.g., Swain 1985, 1995, 2005) and feedback – was central to a fuller understanding of the role of interaction in acquisition. Corrective feedback (see Loewen 2012; Lyster et al. 2013 for an overview) is central to the two articles discussed in this paper. There are many ways that corrective feedback can be delivered, ranging from explicit feedback with or without metalinguistic information, as in example 1 taken from Goo (2012), to implicit feedback such as recasts, as in example 2 taken from Révész (2012).

1. Example of metalinguistic feedback

NNS: Who does John think that is reading a newspaper?
NS: Don't use the conjunction 'that' when you ask about the subject of the subordinate clause. By the way, it's Bill. Okay, let's move to the next item.

2. Example of recast

NNS: And I saw a boy next to the bar. I think he was with his girlfriend. They talking to each other.
NS: They were talking to each other.

https://doi.org/10.1017/S0261444815000038 Published online by Cambridge University Press

The rephrasing by the NS in (2) with the correct form while at the same time maintaining the original meaning is a RECAST and is hypothesized to serve a corrective function, assuming that the NNS recognizes the difference between his/her original utterance (*they talking to each other*) and the subsequent utterance (*they were talking to each other*).

It is claimed that interactions themselves provide opportunities for learners to gain information about the language being learned, with corrective moves being an important part of that process and representing the point at which learners engage cognitive mechanisms that allow them to focus attention on language form and make crucial connections between form and meaning. In other words, corrective feedback draws attention to form and provides opportunities for learners to notice the differences between what they have uttered and what a NS might say.

A question arises as to what might enable some learners to make the crucial comparison between their own utterances and their interlocutors' utterances. Why are some individuals better than others? One such difference is individual working memory capacity.

2.2 Working memory capacity

Working memory, it is argued, has two parts: it refers to the structures AND processes that humans use to store and manipulate information (for an overview see Williams 2012). What is important for the purposes of this paper is that working memory capacity varies from individual to individual and if there is individual variability in this capacity, there might also be individual capacity to benefit from interactional feedback.

There is no single way of measuring working memory capacity. Rather, a number of elicitation measures have been used, the majority favouring a dual-task format combining some form of memory measure with a processing measure. The following gives a brief description of each of the four measures referred to in the two papers discussed in section 3.

1. Phonological Short Term Memory (PSTM)
   a. Digit span (DS). Participants hear a series of random numbers and are asked to repeat these numbers in the order in which they were presented.
   b. Non-word span (NWS). Participants hear non-words (with a range of syllables – in the Révész study) and are asked to repeat each non-word.
2. Reading span. Participants read sentences one at a time and are told to remember the last word of each sentence (or a letter at the end of the sentence). They are asked to determine whether the sentence is plausible or not (to ascertain that they have actually processed the sentence). At the end of a set of sentences (usually 2–6), they are asked to write down (in order) the last word of each sentence (or the letters that appeared after each sentence).
3. Operation span. This is similar to the reading span, but instead of sentences, mathematical operations are used ($[2 \times 3) + 1 = ?]$). After each operation, participants are shown a possible response and they must indicate whether the answer is correct. This is then followed by a letter. At the end of each set, participants are asked to repeat the letters of that set in correct order.

https://doi.org/10.1017/S0261444815000038 Published online by Cambridge University Press

Recently, within the SLA literature, studies have emerged that link working memory capacity and interaction (Mackey et al. 2002; Mackey & Sachs 2012; Gass, Behney & Uzum 2013). It is the combined significance of these two topics that led us to consider the importance of replication of studies in this area.

In the following section we suggest two recent studies that represent current thinking within the interactionist paradigm. They both consider interaction as a form of corrective feedback and examine an individual trait, that of working memory capacity. With new directions highlighting individual differences and particularly working memory capacity, it is time for replication before we move too far down the path that one-off studies take us.

3. The original studies and suggested approaches to replication

3.1 Study 1: Révész (2012)

This study, published in *Language Learning*, investigated how the effectiveness of recasts is influenced by working memory and specifically by two types of working memory measures (phonological short-term memory and complex working memory tasks [span tasks]). A large number of studies have supported the role of recasts in facilitating L2 development (see Long 2007 and Mackey & Goo 2007 for a review; Goo & Mackey 2013), but the nature of the types of L2 learning processes they promote (*inter alia* declarative or procedural) is still controversial (e.g., Ellis, Loewen & Erlam 2006; Révész & Han 2006; Ellis 2007; Loewen & Nabei 2007) and depend upon the different types of outcome measures employed. Moreover, even though a body of research suggests that 'individual differences in working memory [. . .] are likely to be important determinants of how much is learned from recasts' (Révész 2012: 99), results from empirical studies are not conclusive. Révész' study aimed to (1) provide empirical evidence about the effect of the outcome measures on the effectiveness of recasts and (2) to demonstrate the mediating role of working memory in the observed results on various outcome measures.

Participants in this study were 90 English as a foreign language (EFL) Hungarian first language (L1) high-school students in beginner-level language classes. Participants were selected based on their developmental stage on the target structure (English past progressive) assessed by a pretest, and were randomly assigned to one of three groups: recast (n = 36), non-recast (n = 36), or control (n = 18). The design consisted of a pretest, treatment, a posttest and a delayed posttest. The pretest was made up of an untimed grammaticality judgment task (GJT), a written description task, and an oral description task. For the GJT, participants were asked to judge the grammaticality of 36 sentences (target and fillers) and correct those sentences they believed to be ungrammatical. For the written description task, participants were asked to describe what each of eight people was doing in a picture with the time contextualized as being in the past. Finally, there was an oral description task in which participants had to describe photographs.

Treatment consisted of three 15-minute recast (or non-recast) sessions. Participants were told they were taking photographs at a time when a bank robbery took place. The task was

https://doi.org/10.1017/S0261444815000038 Published online by Cambridge University Press

to describe the photographs telling the researcher (whose role was that of a police officer) the activities in the photograph so that the researcher would know what everyone in the photo was doing at the time of the robbery. In the recast group, recasts were provided when there was a past progressive error (and randomly for other errors). In the non-recast group, no feedback was given.

PSTM was assessed by means of a digit span (DS) and a non-word span (NWS) test and complex working memory capacity was measured with a reading span (RS) test. Measurement of PSTM was by means of repetition of Hungarian numbers. To determine an individual's NWS, non-words with Hungarian phonotactics were presented to participants following which they had to repeat the sequences. An RS score was obtained via a test of Hungarian consisting of 60 sentences, presented one by one on index cards. Participants read each sentence aloud in sets and subsets after which they were asked to recall the last word of each sentence. This was followed by a question about one of the sentences, but not all (sentences were randomly selected for questioning) so as to make sure participants were focusing on meaning and not just remembering the last words.

Data were analyzed through three many-facet Rasch measurement analyses: pretest-posttest and pretest-delayed posttest gain scores on the (1) GJT, (2) written task, and (3) oral description task. Correlations were run separately for recast and non-recast groups. The correlations were between each of the WM measures and the pretest-posttest gain scores from the three tasks (GJT, written and oral description). Many-facet Rasch measurement and correlation analyses showed that oral recasts had the greatest gains on oral production tests, lesser gains on written production tests, and the least on untimed grammaticality judgment tests. According to these findings, the researcher suggested that recasts are 'more likely to foster development in procedural knowledge.' (p. 124). On the basis of the strong correlation between high PSTM scores and oral test scores on the one hand and high complex working memory capacity scores and writing test scores on the other, the researcher suggested that 'learners engaged in different types of learning processes as a function of their complex WM and PSTM capacities.' (p.124). Additionally, she argued that the types of outcome measures differently favored the deployment of L2 knowledge depending on learners' complex WM or PSTM capacities.

3.2 Approaches to replication

A first question to be raised through replication is a methodological one. One would want to know, for example, if other tests of working memory or other means of delivery of the input sentences (in this study sentences were written on a card) would yield different or similar results. An important issue when assessing working memory is the need to ensure processing along with storage. When only random sentences were measured for processing rather than all sentences, one is left with an incomplete picture. The issue of scoring is also relevant. In some working memory studies, all sentences with incorrect plausibility judgments following the presentation of the sentence are eliminated. In the current study, that was not possible because not all sentences received plausibility judgments. Thus, a replication of this study

https://doi.org/10.1017/S0261444815000038 Published online by Cambridge University Press

(with computer delivery) would include all sentences tested and would allow one to ensure both processing and storage.

A second area of replication has to do with different results based on differing working memory tests. A replication with extended data analysis is one way to sort out these differences. One of the explanations Révész provides for differences in results depends on the working memory measure used, namely, high RS individuals gained more on written tasks, whereas those with high PSTM improved to a greater extent on oral tasks. She suggests that these differences have to do with learner engagement. To further support this notion, one could include measures of responses to recasts as a way of determining engagement with the recasted form. Given the role of feedback and responses to feedback, this is an important area for further investigation, as Révész herself points out. To do this, a researcher would create a coding method that would provide information on the responses to each recast, for example: (1) no opportunity, (2) opportunity, but did not repeat, (3) repeated the recasted form, (4) negotiated the response, or (5) used the recasted form later in the discourse, with (1) being low engagement and (5) greatest engagement. This would allow an examination of engagement through the type of response to recast and in relation to different working memory tests.

Another method to determine the degree of engagement with the recast could result from a (partial) conceptual replication adding stimulated recall to the experimental design. Stimulated recall is an important technique (Gass & Mackey 2000) used to elicit 'data about thought processes involved in carrying out a task or activity' (p. 1). As Bloom (1954: 25) notes: 'A subject may be enabled to relive an original situation with great vividness and accuracy if he is presented with a large number of the cues or stimuli which occurred during the original situation.' Because some of Révész' arguments rely on the notion of engagement and transfer-appropriate processing which states that individuals can better 'remember what [they] have learned if the cognitive processes that are active during learning are similar to those that are active during rehearsal' (Lightbown 2007: 27), probing individuals' thought processes would be an important way of accessing that information. To do this, the original interaction would need to be taped (video is best) with the tape serving as the stimulus upon which the participant comments (adhering to the recommendations in Gass & Mackey regarding time lapse and other stimulus preparation). In this case, engagement would be operationalized as comments made about the recast. We make this suggestion in full recognition that absence of mention may not signify lack of noticing or engagement. Nonetheless, it could add information that goes beyond speculation and which could be triangulated with other data, such as the response categories mentioned above.

Generalizability of the results in this study depends on the extent to which the results can be extended to different types of individuals (e.g., age groups), different grammatical constructions, different proficiency levels, different L1-L2 relationships, and different contexts (classroom versus non-classroom). It is appropriate and important to consider learners other than high school age participants in order to determine if the results are unique to a particular age group or if they are generalizable to a wider range of ages. How do working memory measures relate to languages that are closely related (e.g., Italian and Spanish)? As Révész discussed in the article, it is important to bring in the construct of grammatical complexity and/or transparency. The past progressive is not complex and may be particularly susceptible

https://doi.org/10.1017/S0261444815000038 Published online by Cambridge University Press

to learning from recasts; complex structures, as is the case with *that*-trace, may not be. Therefore, one could conduct an approximate replication keeping the methodology the same, but varying the grammatical structures. This would allow a researcher to determine the potential generalizability of the original study through an understanding of the role of grammatical complexity.

We turn next to a study by Goo (2012), who dealt with some of the same variables, namely, working memory capacity and feedback.

3.3 Study 2: Goo (2012)

The topic of this paper is similar to the one by Révész discussed above. Goo's article, published in *Studies in Second Language Acquisition*, compares the benefits of two types of corrective feedback, recasts and metalinguistic feedback, on the acquisition of *that*-trace filter (** Who does John think that is reading a newspaper?* versus *Who does John think is reading a newspaper?*). Goo specifically considered the mediating role of working memory capacity on these feedback types. Empirical evidence has been provided on the facilitative role of recasts in L2 development, but findings also suggest that other types of feedback, such as metalinguistic feedback, are more effective than recasts. Recasts and metalinguistic feedback differ, *inter alia*, in the level of implicitness or explicitness of the given feedback and therefore on the attention-control processes involved. Individual differences in working memory capacity can play a crucial role in the effectiveness of corrective feedback in interaction, but research (Mackey et al. 2002, 2010; Sagarra 2007; Trofimovich, Ammar & Gatbonton 2007; Gass et al. 2013) yields mixed findings. Goo's study aimed to further the understanding of the relative effects of recasts and metalinguistic feedback and the potential role of individual differences in working memory capacity on their effectiveness.

Fifty-four EFL Korean L1 university students at intermediate proficiency level were randomly assigned to a recast (n = 14), metalinguistic feedback (n = 32) or control group (n = 8; this group participated in pretest-posttest working memory tests only). In a pretest-posttest design experiment, participants attended two 20-minute sessions during which the target construction was elicited with one-way information-gap activities and corrected respectively with full recasts or metalinguistic explanations. This was done as an in-class activity. A picture was projected onto a screen and each participant was given a list of 15 sentences (ten of which were the target) with some blanks. Students had to ask the teacher a question about the missing item. To avoid issues of modified output, the teacher did not allow participants the opportunity to respond to his/her comments. There was a second treatment session with the researcher three days after the first treatment; this was apparently done as a whole-group activity. Working memory capacity was assessed in the L1 with two span tasks: reading span (RS) task and operation span (OS). The pretest and posttest comprised a GJT and a written production test. In the GJT (computer administered), there were 24 sentence pairs for each of the two tests (eight of which were fillers). Each pair consisted of one declarative sentence and a question asking about an underlined part of that sentence (e.g., *They think that Jeff loved Jennifer. Who do they think that loved Jennifer?*); the task was to judge the question as grammatical

https://doi.org/10.1017/S0261444815000038 Published online by Cambridge University Press

or not. For the written production test, there were 15 critical (declarative) sentences (with one word underlined) for the pretest and for the posttest. The task was to write a question about the underlined word (e.g., *They believe that John is working as a waiter at a Korean restaurant; They think that Mike designed Jane's jacket*).

Findings showed similar beneficial effects of recasts and metalinguistic feedback, in a blocked modified output situation, in facilitating the acquisition of the target construction. With regard to the effect of working memory capacity, the significant correlation between working memory capacity and recasts suggested that working memory capacity mediates and predicts the effectiveness of recasts, but no such significant correlation existed for metalinguistic feedback, suggesting the lack of impact of working memory capacity on metalinguistic feedback. The researcher argued that the two types of feedback 'reflect two different mechanisms for how noticing occurs' (p. 465), as recasts necessitate attention-control processes (essential elements of working memory capacity) which are not particularly required for the noticing of metalinguistic feedback given their highly overt nature.

3.4 Approaches to replication

We turn now to possibilities for replication of the Goo study. The finding that individual differences in working memory capacity influence the effectiveness of recasts, but not of metalinguistic feedback, is particularly relevant with regard to its implications for the cognitive processes involved in the two corrective feedback procedures and how they can be beneficial in different situations and for different students. Yet, we are left with unanswered questions, some of which relate to a consideration of both of these studies together.

Because of the differences in population (age and L1) and context in the two studies reported on here, it is difficult to assess the impact of working memory as a general conclusion. This is made more difficult because the size of the recast group in Goo's study was relatively small ($n = 14$). Other modifications to Goo's study would be important. For example, the target structure used is a complex one and undoubtedly is not one that these learners use in everyday conversation. Further, it is not a structure that one would be likely to receive feedback on given its comprehensibility with or without the use of *that*. One of the criteria for inclusion in Goo's study was the rarity of this construction in natural discourse. This structure '. . . was assumed, to some extent, to minimize impacts of potentially confounding factors (e.g., prior learning or exposure)' (p. 453). A replication study would be particularly useful to see if this confined usefulness of feedback can be generalized. Thus, a replication study could include a wider range of structures, or, at least structures that are more likely to receive feedback in a context that extends beyond the laboratory.

As with the Révész study, one would want to understand better what learners were thinking about as they received feedback. In the recast group, for example, did they notice the recast? Replicating the study with that component would yield a more nuanced understanding of the role of working memory. This could be accomplished through a stimulated recall procedure in which each student would be given a stimulus (the video of the task) and asked questions about what they were thinking about with each instance of feedback. What could be further

https://doi.org/10.1017/S0261444815000038 Published online by Cambridge University Press

explored in this way is the role of direct feedback (to the individual asking the *wh-* question) versus observed feedback (to a classmate).

In Goo's study, there was no delayed posttest, leaving one to wonder what the true impact is. Such an inclusion in a replication study would allow a more comfortable generalization regarding claims about learning rather than to just the immediate impact of the treatment. Finally, as we understand the treatment, it was done as a class and not individually. Because it is important for learners to notice feedback in order for that feedback to have the greatest impact and because a face-to-face interaction is more likely to lead to noticing (see recent overview by Mackey et al. 2012), a replication of the same study in a one-to-one setting might provide information that would help the field determine the extent of impact. A second possibility was mentioned above, namely, a replication in which introspective comments through a stimulated recall could potentially separate feedback to an individual and feedback to another individual.

Finally, since we know that scoring procedures (see Leeser & Sunderman, ms) result in different conclusions, one would want to analyze a set of data with different scoring procedures to determine the stability of the results. The interested reader is referred to Conway et al. 2005 for alternative scoring methods in a possible replication study.

4. Concluding remarks

Attempting to generalize findings is essential before findings in a field become too entrenched in the psyche of a field. It is important to take stock and possibly backtrack at early stages of research areas. These two studies, because they deal with similar topics, and because the field is just beginning to investigate these questions, are appropriate for all types of future replication studies. The major question addressed is: Does working memory mediate effectiveness of feedback or not? In Goo's study it mediates the effects of recasts, but it does not of metalinguistic feedback; in the Révész study, it depends on the type of working memory test given. As suggested in this paper, there are a number of specific ways to replicate the original studies, both of which are important in current SLA research. Replication studies would further this area of research by clarifying some of the conflicting areas, as mentioned above, providing greater clarity on working memory, and extending the results to a wider range of target structures and age groups.

References

Block, D. (2003). *The social turn in second language acquisition*. Edinburgh: Edinburgh University Press.

Bloom, B. (1954). The thought processes of students in discussion. In S. J. French (ed.), *Accent on teaching: Experiments in general education*. New York: Harper, 23–46.

Conway, A. R. A., M. J. Kane, M. F. Bunting, D. Z. Hambrick, O. Wilhelm & R. W. Engle (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review* 12.5, 769–786.

Ellis, R. (2007). The differential effect of corrective feedback on two grammatical structures. In A. Mackey (ed.), *Conversational interaction and second language acquisition: A series of empirical studies*. Oxford, UK: Oxford University Press, 339–360.

https://doi.org/10.1017/S0261444815000038 Published online by Cambridge University Press

Ellis, R., S. Loewen & R. Erlam (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition* 28.2, 339–368.
Ferguson, C. (1971). Absence of copula and the notion of simplicity: A study of normal speech, baby talk, foreigner talk and pidgins. In D. Hymes (ed.), *Pidginization and creolization of languages*. Cambridge, UK: Cambridge University Press, 141–150.
Gass, S. (1997). *Input, interaction and the second language learner*. Mahwah, NJ: Lawrence Erlbaum Associates.
Gass, S. & A. Mackey (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.
Gass, S. & A. Mackey (2006). Input, interaction and output: An overview. In K. Bardovi-Harlig & Z. Dörnyei (eds.), *AILA Review*. Amsterdam: John Benjamins, 3–17.
Gass, S. & A. Mackey (2007). Input, interaction, and output in second language acquisition. In B. VanPatten & J. Williams (eds.), *Theories in second language acquisition: An introduction*. Mahwah, NJ: Lawrence Erlbaum Associates, 175–179.
Gass, S., J. Behney & B. Uzum (2013). Inhibitory control, working memory, and L2 interaction gains. In K. Drożdział-Szelest & M. Pawlak (eds.), *Psycholinguistic and sociolinguistic perspectives on second language learning and teaching*. Berlin: Springer Verlag, 91–114.
Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition* 34.3, 445–474.
Goo, J. & A. Mackey (2013). The case against the case against recasts. *Studies in Second Language Acquisition* 35.1, 127–165.
Keck, C., G. Iberri-Shea, N. Tracy-Ventura & S. Wa-Mbalaka (2006). Investigating the empirical link between interaction and acquisition: A quantitative metanalysis. In J. M. Norris & L. Ortega (eds.), *Synthesizing research on language learning and teaching*. Amsterdam: John Benjamins, 91–131.
Leeser, M. & G. Sunderman (ms). Methodological issues of working memory tasks for L2 processing research.
Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning* 60.2, 309–365.
Lightbown, P. (2007). Transfer appropriate processing as a model for classroom second language acquisition. In Z. H. Han (ed.), *Understanding second language process*. Clevedon, UK: Multilingual Matters, 27–44.
Loewen, S. (2012). The role of feedback. In S. Gass & A. Mackey (eds.), *The Routledge handbook of second language acquisition*. New York: Routledge, 24–40.
Loewen, S. & T. Nabei (2007). Measuring the effects of oral corrective feedback on L2 knowledge. In A. Mackey (ed.), *Conversational interaction in second language acquisition: A series of empirical studies*. Oxford, UK: Oxford University Press, 361–377.
Long, M. H. (1980). *Input, interaction and second language acquisition*. Unpublished doctoral dissertation, University of California, Los Angeles.
Long, M. H. (2007). *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum Associates.
Lyster, R. & K. Saito (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition* 19.1, 37–66.
Lyster, R., K. Saito & M. Sato (2013). Oral corrective feedback in second language classrooms. *Language Teaching* 46.1, 1–40.
Mackey, A. (2012). *Input, interaction, and corrective feedback in L2 learning*. Oxford: Oxford University Press.
Mackey, A. (ed.) (2007). *Conversational interaction in second language acquisition: A collection of empirical studies*. Oxford: Oxford University Press.
Mackey, A. & J. Goo (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (ed.), *Conversational interaction in second language acquisition: A series of empirical studies*. Oxford, UK: Oxford University Press, 407–452.
Mackey, A. & R. Sachs (2012). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning* 62.3, 704–740.
Mackey, A., R. Abbuhl & S. Gass (2012). Interactionist approach. In S. Gass & A. Mackey (eds.), *The Routledge handbook of second language acquisition*. New York: Routledge, 7–23.
Mackey, A., R. Adams, C. Stafford & P. Winke (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning* 60.3, 501–533.
Mackey, A., J. Philp, T. Egi, A. Fujii & T. Tatsumi (2002). Individual differences in working memory, noticing of interactional feedback, and L2 development. In P. Robinson (ed.), *Individual differences and instructed language learning*. Philadelphia: John Benjamins, 181–209.

https://doi.org/10.1017/S0261444815000038 Published online by Cambridge University Press

Nicholas, H., P. Lightbown & N. Spada (2001). Recasts as feedback to language learners. *Language Learning* 51.4, 719–758.
Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning* 62.1, 93–132.
Révész, A. & Z. H. Han (2006). Task content familiarity, task type and efficacy of recasts. *Language Awareness* 3.3, 160–179.
Russell, J. & N. Spada (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J. M. Norris & L. Ortega (eds.), *Synthesizing research on language learning and teaching*. Amsterdam: John Benjamins, 133–164.
Sagarra, N. (2007). Working memory and L2 processing of redundant grammatical forms. In Z. Han (ed.), *Understanding second language processes*. Clevedon: Multilingual Matters, 133–147.
Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (eds.), *Input in second language acquisition*. Rowley, MA: Newbury House, 235–253.
Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (eds.), *Principle and practice in applied linguistics*. Oxford, UK: Oxford University Press, 125–144.
Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (ed.), *Handbook of research in second language teaching and learning*. Mahwah, NJ: Lawrence Erlbaum Associates, 471–483.
Trofimovich, P., A. Ammar & E. Gatbonton (2007). How effective are recasts: The role of attention, memory, and analytic ability. In A. Mackey (ed.), *Conversational interaction in second language acquisition: A collection of empirical studies*. Oxford: Oxford University Press, 171–195.
Wagner–Gough, K. & E. Hatch (1975). The importance of input in second language acquisition studies. *Language Learning* 25.2, 297–308.
Williams, J. (2012). Working memory and SLA. In S. Gass & A. Mackey (eds.), *The Routledge handbook of second language acquisition*. New York: Routledge, 427–441.

SUSAN GASS is University Distinguished Professor at Michigan State University where she directs the Ph.D. in Second Language Studies and is Chair of the Department of Linguistics & Germanic, Slavic, Asian and African Languages. She has published widely in the field of second language acquisition including her most recent book, *Second language acquisition: An introductory course* with Jennifer Behney and Luke Plonsky (Routledge, 2013). She is the recipient of numerous honors and awards and has served as the President of the American Association for Applied Linguistics and the International Association of Applied Linguistics. She can be reached at: English Language Center, Michigan State University, B230 Wells Hall, 619 Red Cedar Road, East Lansing, MI 48824.

LORENA VALMORI is a Ph.D. candidate in the Second Language Studies program at Michigan State University. She previously taught L2 Italian at the Language Center of the Università di Modena e Reggio Emilia (Italy) and Virginia Tech (VA). Her research interests are in how individual differences affect learning, teaching and rating, and in particular in teachers' motivation and rater bias. She can be reached at: English Language Center, Michigan State University, B220 Wells Hall, 619 Red Cedar Road, East Lansing, MI 48824.

https://doi.org/10.1017/S0261444815000038 Published online by Cambridge University Press