CAMBRIDGE
UNIVERSITY PRESS

REPLICATION RESEARCH

The interaction of working memory capacity and engagement with recasts on different L2 outcome measures: A replication of Révész (2012)

Javad Alipour*, Maryam Mohebi and Ali Roohani

Shahrekord University, Shahrekord, Iran
*Corresponding author. Email: alipour.edu@gmail.com; javad.alipour@sku.ac.ir

Abstract

We report on a conceptual replication of Révész (2012) in order to investigate the idea whether learners provided with recasts do engage in different kinds of behavioral engagement as a function of their working memory and if/how this engagement comes to bear on performance on different measures. Engagement with recasts was measured through a coding method categorizing responses to the recasts running the gamut from: (1) no opportunity, (2) opportunity, but did not repeat, (3) repeated the recasted form, (4) negotiated the response, to (5) used the recasted form later in the discourse. Consistent with Révész (2012), though with lower effect sizes, the results showed that recasts were most conducive to gains on an oral task and less so on a written description task, but non-effective on a grammaticality judgment task. Furthermore, it was revealed that learners with a high phonological short-term memory were more prone to recast-induced engagement on an oral production task, whereas those enjoying a higher reading span were considerably less so. We propose that learner engagement be deemed more important in future interaction research.

Keywords: oral feedback, working memory, different outcome measures, recast, engagement with recasts

1. Introduction

A recast involves a teacher's or other interlocutor's reformulation of a learner's erroneous utterance via altering one or more errors in it while retaining its semantic content (Révész, 2012). It typically occurs in a context where interlocutors are engaged in communicative activities as opposed to focusing on forms activities (Long, 2007). The following excerpt (Gass & Varonis, 1989) is often cited in the literature to exemplify the provision of recast in a learner-learner interaction:

> Learner A: A man is uh drinking coffee or tea uh with uh the saucer of the uh uh coffee set is uh in his uh knee
> Learner B: In him knee
> Learner A: Uh on his knee
> Learner B: Yeah
> Learner A: On his knee
> Learner B: So sorry, on his knee

In this conversation, learner B first offers erroneous recast, followed by learner A's recast, which is, then, followed by learner B's confirmation and subsequent modified output.

There are mixed findings in the second language acquisition (SLA) literature regarding the effectiveness of recasts in promoting second language development (Révész, 2012). Several researchers have

CrossMark

underscored their potential benefits (e.g., Gass, 2003; Long, 2007; Nassaji, 2007, 2016, 2017, 2019) including the provision of both positive and negative feedback, immediate juxtaposition of the error along with the reformulated version, facilitation of understanding on the part of the learner, and increased noticing. On the other hand, some have cast doubt on the efficacy of recasts, arguing that they could be potentially ambiguous (e.g., Lyster & Ranta, 1997; Panova & Lyster, 2002) and that learners may conceive them as non-corrective.

Despite this ongoing debate, several studies have investigated the variables that may influence the extent to which recasts can promote second language (L2) acquisition (e.g., Ellis et al., 2001; Goo, 2012; Hawkes & Nassaji, 2016; Lyster & Ranta, 1997; Mackey, 2007; Nassaji, 2017, 2019; Saito, 2013; Sheen, 2011). In this respect, an important question concerns the nature of the relationship between learner factors, including gender (Ross-Feldman, 2007), language anxiety (Sheen, 2008), and language aptitude (Sheen, 2007), with recasts in relation to linguistic development. One learner factor relevant to the present study is working memory (WM), which has been the subject of several investigations (e.g., Goo, 2016; Mackey et al., 2010; Révész, 2012; Trofimovich et al., 2007). It has been defined as the 'the temporary storage and manipulation of information that is assumed to be necessary for a wide range of complex cognitive activities' (Baddeley, 2003, p. 189), with processing and storage competing for a shared pool of limited resources that vary across individuals. This individual variability, referred to as working memory capacity (WMC), could justify learners' differences in benefiting from interactional feedback. Another fertile area of research on recasts concerns the idea whether recasts are more likely to result in declarative or procedural knowledge (e.g., Ellis et al., 2006; Loewen & Nabei, 2007). This idea led Révész (2012) to take this line of research further and examine whether the effects of recasts on various outcome measures are mediated by a difference in WM capacity, a learner-internal factor.

The initial study investigated the role of WM in relation to gains from recasts on different outcome measures. However, it did not indicate whether these differences would carry over to the learners' reactions after receiving recasts (Révész, 2012). The study was concerned with the question why some learners are better than others when it comes to making a comparison between their own erroneous utterance and their interlocutors' more target-like utterance, seeking the answer in the difference in learners' WM. Révész discussed the hypothesis that depending on their phonological short-term memory (PSTM) and complex verbal WMC, two of the four components of working memory (Baddeley, 2000), the participants in her study had possibly engaged in different types of learning processes to various degrees, with high PSTM learners being able to maintain the information in recasts longer in short-term memory. It is possible that these different types of learning processes could be partly accounted for by the different types of behavior that learners engage in after receiving the feedback. This is a significant hypothesis to investigate since it is in line with the idea that learners do not perceive feedback in the same way (Mackey et al., 2010).

Furthermore, since a number of researchers (e.g., McDonough & Mackey, 2006) have argued that the process of modifying output could be as important as the product, an interesting question that arises at this juncture is concerned with how differences in WM capacity translate into possibly different types of behavior, given that some learners may modify their output during interaction whereas others may not and that WM may be responsible for the benefits that learners garner from interaction and even for what they do at the moment of output modification (Mackey et al., 2010). In other words, it is possible that high WM learners may be better able to reprocess and reconstruct their utterances with new grammatical encoding. A second justification for performing a replication on the study concerns the idea whether the three outcomes measures that Révész used did draw on differential types of knowledge, given that the issue was not an established fact but rather an assumption, rendering some of her interpretations speculative (Révész, 2012). A third justification has to do with the idea that the participants in the initial study were Hungarian learners of English, which, as Révész points out, constrains the generalizability of the findings to learner populations with other first languages and in different contexts.

According to the classification proposed by the Language Teaching Review Panel (2008), the current study would fall more toward the conceptual end of the replication continuum in that (a) it was designed to assess the extent to which the findings of the initial study would be generalizable to other learner populations, and (b) it was aimed at assessing an explanation that the researcher of the initial study provided to account for the results she obtained. The explanation underscored the idea that the participants in the initial study had possibly engaged in different types of learning processes to various degrees. As a follow-up on the initial study and its proposed explanation, the current study attempted to investigate into the interaction of learners' gains from recasts and WM capacity as observed on different outcome measures and the extent to which the participants were behaviorally engaged with recasts as a function of their WM capacity. Specifically, what happens after recasts was the focus of the current study. This was accomplished by a coding method that categorized responses to recasts.

2. Recasts and different outcome measures

In an L2 interaction-based context, learners may focus on form while attending to meaning and use at the same time. One of the primary ways that this opportunity can be provided is through recasts (Doughty, 2001; Nassaji, 2016). What sets recasts apart from other types of feedback is that they (a) provide the correct form and (b) try to maintain the focus on meaning (Nassaji, 2017), thereby prompting learners to make cognitive comparisons (Doughty, 2001) between their non-target constructions and target-like utterances. Long (2007) argued that recasts could afford valuable processing resources for form-meaning connections. He also underscored the advantage that recasts occur in context, and thus involve semantic contingency and joint attentional focus between the erroneous utterance and the reformulated one. In other words, as the learner is engaged in and understands at least part of a message, they can notice the gap (Schmidt, 1990), hence the allocation of more cognitive resources to form-meaning connections. It has also been argued that recasts could potentially lead to modified output and automatization of language knowledge (McDonough, 2005; Swain, 1995). These benefits have rendered recasts the most frequently used feedback technique in L2 classrooms (Lyster & Ranta, 1997). It should be noted, however, that despite the above-mentioned theoretical benefits associated with recasts, some have cast doubts on whether they can promote SLA (e.g., Lyster & Ranta, 1997, 2013; Panova & Lyster, 2002).

As to what type of knowledge recasts could foster, a brief overview of a dichotomy in SLA literature is in order, that of declarative and procedural knowledge. According to the skill acquisition theory (DeKeyser, 2009), language learning occurs in several stages and is, in essence, like other complex cognitive skills such as learning how to swim. In other words, L2 learners first obtain information from the input. At this declarative stage, knowledge is characterized as being slow and demanding in terms of processing. The next stage, known as proceduralization, involves applying these rules to production or comprehension (knowledge *How*), resulting in faster performance by drawing on ready-made chunks. Finally, having been exposed to extensive practice opportunities, L2 learners are likely to automatize the procedural knowledge for use in fluent and effortless performance (DeKeyser, 2007a, 2007b). DeKeyser (2009) argues that procedural knowledge itself does not turn into automatized knowledge, but that its presence is conducive to the development of automatized knowledge. He also underscores the idea that practice should involve 'real operating conditions' (p. 292). This emphasis on the provision of real conditions renders oral feedback including recasts a potential candidate for inducing proceduralization and automatization of correct forms.

Révész (2012) concurs with Ellis's (2004) recommendation that different carefully selected assessment tools ought to be included in effects-of-instruction studies to better shed light on the effects of the treatment, if any, on different types of L2 knowledge. Caveat should, however, be exercised that implicit instruction does not necessarily lead to implicit knowledge, nor does explicit instruction to explicit knowledge (Ellis, 2004). Regarding recasts, for example, it has been empirically shown that they can lead to gains on a range of various assessment tools, from metalinguistic tests (e.g., Ellis et al., 2006; Loewen & Nabei, 2007; Nassaji, 2017) to oral production tasks (e.g., Mackey & Philp,

1998; Nassaji, 2017; Révész & Han, 2006). Meta-analyses of feedback studies have also shown that oral communicative tasks are more amenable to change by recasts than metalinguistic tasks, although both types of measure seem to benefit from them substantially (Lyster & Saito, 2010; Mackey & Goo, 2007). Until recently, most recast studies employed a single assessment task (e.g., Mackey & Philp, 1998; McDounough & Mackey, 2006). This has been unwarranted, given the need in SLA research to gather a multiplicity of data sources (Doughty, 2003; Ellis et al., 2006; Norris & Ortega, 2003). Over the past few years, however, several studies have recognized the need for this multiplicity (e.g., Nassaji, 2017, 2019; Révész, 2012).

3. WM

WM refers to the cognitive processes that account for the temporary storage and manipulation of information and is assumed to 'serve the function of facilitating performance of a range of complex tasks' (Baddeley, 2003, p. 190). The most widely used model of WM has been proposed by Baddeley and Hitch (1974). It was originally made up of three separable components assumed to be working in tandem: Two slave systems, namely the phonological loop and the visual-spatial sketch-pad and a central executive with limited attentional capacity. In this model, the phonological loop is held to be responsible for the retention of auditory sequences whereas the visual-spatial sketchpad handles visual and spatial information. The central executive is tasked with monitoring the interaction of these components with other cognitive domains and filtering information. A fourth component was later added to the model: The episodic buffer (Baddeley, 2000). This new slave system has the function of storing information, where information from several different sources can be bound into chunks and episodes and where information from different modalities can be integrated into a single multi-faceted experience.

Phonological loop is further divided into two subsystems, namely a temporary storage system and a subvocal rehearsal system. The temporary storage holds memory traces very briefly such that the information will decay unless the rehearsal component comes into play. The rehearsal process involves both maintaining and registering information within the store, provided that the received visual information can be verbalized (Baddeley, 2003).

Phonological loop capacity is argued to play a key role in certain aspects of SLA (e.g., Ellis & Sinclair, 1996; French & O'Brien, 2008; O'Brien et al., 2006). With specific reference to recasts, some have argued that learners with a higher PSTM are more likely to show grammatical development. Ellis (2005), for example, speculates that learners with high phonological loop capacity have an edge in gaining grammatical benefits from recasts by virtue of their superior ability to hold the reformulated utterance in memory. Research also shows that high PSTM learners outperform low PSTM learners on delayed posttests (Mackey et al., 2002; Trofimovich et al., 2007). It should be noted that there are two widely accepted measures of PSTM, namely the digit span (DS) and non-word span (NWS) tasks (Baddeley, 2003). The former involves the repetition of increasing numbers of digits, while the latter requires subjects to repeat sequences of non-words of varying lengths.

The second component of the model is the central executive, tasked with supervising attentional processes (Baddeley, 2003). As with PSTM, the central executive is limited in capacity and differs across individuals (Baddeley, 2003). Assessment of this component typically involves subjects doing complex WM tasks such as reading or listening to a set of sentences and having to recall the final word of each sentence after the presentation of a set (Conway et al., 2005; Waters & Caplan, 1996), referred to as reading span (RS) and listening span (LS) tasks respectively.

The role of WM in L2 learning and use has been the focus of extensive research over the past few decades. Overall, WM is deemed a core element of language aptitude and overall proficiency, in addition to various L2 processes including reading, writing, sentence processing, speaking, vocabulary development, grammar, and the processing of intake (see Juffs & Harrington, 2011, for an overview). However, there is still a debate in SLA research on whether there is a link between complex WM capacity and gains from recasts, with some studies reporting a positive relationship (e.g., Mackey et al.,

2002; Sagarra, 2007) and others suggesting no effect (e.g., Trofimovich et al., 2007). It has been suggested that the differences in these findings could be explained in terms of differences in methodology (Révész, 2012). For instance, whereas in some studies, recasts are delivered exclusively in response to learners' erroneous utterances (e.g., Mackey et al., 2002; Sagarra, 2007), in others (e.g., Trofimovich et al., 2007), they are provided after each learner utterance, regardless of whether an error occurs, hence more predictability.

4. The initial study

4.1. Method

In order to investigate the effect of different outcome measures on gains from recasts, Révész (2012) collected data as part of a larger study (Révész, 2007), using a pretest-posttest-delayed posttest design. The participants were 90 beginner-level English as a Foreign Language (EFL) learners, who were randomly assigned to recast, non-recast, and control groups. The past progressive construction was chosen as the target structure in the study on the grounds of enjoying perceptual salience and bearing communicative value. Perceptual salience concerns the ease of hearing or perceiving a given structure and is typically measured by three sub-factors, namely the number of phones in the functor, the presence/absence of a vowel in the surface form, and the total relative sonority of the functor (see Goldschneider & DeKeyser, 2001). The past progressive is made up of a free morpheme (*was/were*) and a bound morpheme (*ing*) and is considered to be physically salient. The structure also bears communicative value in that it denotes grammatical tense and aspect. Before and after the treatment, that is, the provision versus absence of recasts, a grammaticality judgment task (GJT), a written production task, and two oral communication tasks were administered. Half of the participants were also invited to take the delayed posttest. In addition, tests of PSTM (DS, NWS) along with a test of complex WM were administered to a few participants in each group (Révész, 2012).

4.2. Materials and procedures

Révész (2012) used three comparable versions of the treatment task, each for one treatment session and each containing ten photos. The materials were designed using the computer program Microsoft PowerPoint and piloted on native speakers of English and Hungarian beginning-level EFL learners. She pointed out that the piloting was intended to determine whether there was an equal number of obligatory contexts for the target structure (i.e., one usage of the past progressive construction), and that the use thereof was natural. Comparability of the three versions in terms of lexical variation was confirmed by Guirad's index and in terms of syntactic complexity by clauses per Analysis of Speech unit (Foster et al., 2000).

The task in the study involved the participants imagining themselves as witnesses of a hypothetical situation, presented in a photo, where a crime was happening. It, then, required them to carefully describe each of their photos within 40 seconds, with the researcher taking on the role of a police officer. The treatment was provided on three separate days, with each session lasting 15 minutes. In the recast group, recasts were consistently provided to learners by the researcher in the event they produced errors in the use of the past progressive construction. A small number of distracting recasts were also presented randomly in response to other erroneous utterances. The study controlled for recast characteristics by providing only simple declarative recasts with falling intonation. The non-recast group did not receive any type of feedback during the treatment sessions while the control group participated only in the pretest and posttests. To assess the participants' growth or lack thereof after the treatment, Révész used three versions of three different outcome measures. A brief overview of these measures follows.

4.3. Assessment measures and scoring

4.3.1. GJT

To tap the declarative knowledge of participants regarding the past progressive construction, Révész (2012) employed a GJT, consisting of 36 sentences and 12 distractors. The error types were determined by recourse to past SLA research (Bardovi-Harlig, 2000). In the piloting phase, she invited 15 Hungarian EFL learners from a similar population to take the test. No significant differences were found among the three versions. In the main study, no time limit was set for the task to allow learners to draw on their declarative knowledge. Each learner response was classified into one of four categories, motivated by the developmental stages that learners pass through in learning this construction (Bardovi-Harlig, 2000), namely stage 1: bare progressive (e.g., *talking*), stage 2: present progressive (e.g., *is talking*), and stage 3: past progressive (e.g., *was talking*). Révész (2012) used the following scoring method: Three points were awarded for grammatical items judged as grammatical and for ungrammatical items supplied with appropriate corrections. Two points were awarded for changing grammatical sentences to present progressive, for changing ungrammatical bare progressive sentences to present progressive, and for judging ungrammatical present progressive sentences as grammatical. One point was given for correcting a grammatical item into a bare progressive, for changing an ungrammatical present progressive form into a bare progressive, and for judging a bare progressive item as grammatical. Zero points were awarded for grammatical items judged as ungrammatical with any non-target-like change to a form different from the ones listed above. Any sentence that was judged ungrammatical but for which no correction was supplied was excluded. Also excluded were items where the corrections indicated that the sentences were judged based on linguistic forms other than the target form.

4.3.2. Written picture description task

Révész (2012) used this task on the grounds that it would allow for the use of both declarative knowledge, by virtue of being unpressured, and procedural knowledge, by requiring the participants to supply the target construction. A 10-minute time limit was set for the task to allow for relatively unpressured written production. It involved describing a picture that showed eight people engaged in different activities. Then, the data were coded, yielding a total score for each participant. The coding involved identifying the obligatory contexts for the target structure, checking to see if any progressive marking had been produced, and analyzing the data based on the four developmental categories suggested in Bardovi-Harlig (2000).

4.3.3. Oral description tasks

Six versions of two time-pressured oral tasks, which involved the participants describing five photos, were developed and piloted. These tasks were employed on the assumption that they would provide appropriate contexts for the use of procedural knowledge. A total of 87 hours of oral data was recorded, transcribed, and subsequently coded (Révész, 2012).

4.3.4. Tests of PSTM and complex WMC

Révész used Hungarian versions of the DS and NWS (Racsmány et al., 2005) tests to measure the participants' PSTM. The DS test involved repeating four lists of numbers increasing progressively from three to nine digits. The NWS test contained 36 non-words that conformed to Hungarian phonotactics and ranged in length from one to nine syllables. It required the participants to repeat the non-words as they were presented one by one. The highest number of syllables of a particular length that a participant could recall at least twice was recorded as the participant's NWS. To measure the participants' RS, sentences of increasing length were presented. In addition to requiring the participants to remember the last word of each sentence after the presentation of each set, the task also involved answering comprehension questions for a random number of sentences to ensure that participants were not merely concentrated on recalling the words (Révész, 2012).

4.4. Data collection and analysis

Révész (2012) administered 40-minute pretests and posttests to all and equivalent delayed posttests to half of the participants. Three separate Many-facet Rasch measurements (MFRM) were used (Linacre, 1989) to assess the impact of the treatment on the participants' development of the target construction, with the three analyses involving the different types of outcomes measures. To examine the relationship between the participants' WMC and gains from recasts, a number of Pearson correlations were conducted.

5. Results

5.1. Research question 1

The first research question was concerned with whether the type of outcome measure employed influences the observed effects of recasts on L2 development. The results suggested that it does. However, the extent of development varied considerably from measure to measure, with recasts producing the greatest impact on gain scores on the oral production test (approximately 5.5 logits), followed by the written production test (about 4 logits) and the GJT (approximately 2.5 logits). These results were interpreted as suggesting that while recasts could promote the acquisition of both declarative and procedural knowledge, they seem to be more geared to the development of procedural knowledge.

5.2. Research question 2

The second research question dealt with the relationship between PSTM and/or complex WMC and the recast-induced gains on different outcome measures. A series of correlations were conducted between the WM measures and the pretest posttest gain scores. In the recast group, the participants' gains on the GJT and written description tests showed moderate to strong correlations with their performance on the RS test scores (GJT: $r = 0.53$; $p < 0.05$; written description: $r = 0.47$; $p < 0.05$). However, no significant correlations were found between the GJT and the DS and NWS test scores. On the other hand, moderate to strong correlations were identified between the gain scores achieved by participants on the oral description test and the results of the PSTM tests (DS: $r = 0.49$; $p < 0.05$; NWS: $r = 0.55$; $p < 0.01$), but the oral test gain scores showed no significant correlations with the RS test results. In contrast, the non-recast group showed no significant correlations between the WM and developmental measures. It was implied that WM did not seem to be related to the development of learners in the non-recast group, but was associated with the gains achieved by the recast group (Table 1).

6. Learner engagement

Based on these results, Révész concluded that recasts are conducive to the development of procedural rather than declarative knowledge. In addition, the strong correlation between PSTM scores and oral test scores and that between complex WM scores and the written test scores were attributed to the extent to which learners were engaged with recasts as a function of their complex WM and PSTM differences (Révész, 2012).

These kinds of differences in terms of the degree of involvement may also be confirmed by any teacher observing his or her students in the classroom (Philp & Duchesne, 2016). While some students may be completely off-task, some others may be compliant, yet inattentive. The term frequently used to talk about learners' interest and participation in an activity is ENGAGEMENT. Definitions of the term are highly variable, possibly because of the varied research contexts and foci, each underscoring particular facets. Generally, it refers to a state of heightened attention and involvement and has been described as a multifaceted construct including, at the least, three components: cognitive, behavioral, and emotional (Fredricks et al., 2004; Philp & Duchesne, 2016). Philp and Duchesne (2016) observe that behavioral engagement concerns the degree and quality of learners' participation, with such indicators as the

Table 1. Correlations of working memory test scores and pretest-posttest gain scores in Révész (2012)

Group	Test	GJT	Written description	Oral description
Recast group (*N* = 23)	Digit span	0.08	0.39	0.49*
	Nonword span	−0.05	0.38	0.55**
	Reading-span	0.53*	0.47*	0.01
Non-recast group (*N* = 22)	Digit span	−0.15	−0.15	−0.04
	Nonword span	−0.30	−0.01	0.07
	Reading-span	−0.10	−0.16	−0.15

amount of effort, persistence, and active involvement. Cognitive involvement refers to sustained attention and mental effort, whereas emotional engagement is found in the learners' subjective or affective responses during tasks. A fourth dimension that some consider to be also involved in tasks concerns social engagement, highlighting the learners' degree of affiliation and willingness to be involved. A perfect example of the positive involvement of all factors is the concept of flow (Csikszentmihalyi, 1990), representing the all-encompassing involvement of the individual to the exclusion of all else.

Given the consensus that engagement could provide an optimal condition for learning, it comes as a surprise that, to date, there has been little principled understanding of the term in applied linguistics research. However, following the emphasis on the need for L2 learners to pay attention to the connections between language form and its meaning in use, the construct has recently drawn the attention of SLA research. This kind of research (e.g., Leow, 2015; Robinson, 1995) recognizes gradations of cognitive involvement, with the term engagement being used as a near synonym, although, as noted above, paying attention is only one dimension of engagement. For the purposes of the current study, only the behavioral dimension, involving such indicators as answering questions or participating in tasks (Philp & Duchesne, 2016), was taken into account.

Considering the important role of responses to feedback in today's interaction research, it would be plausible to explore learners' level of engagement induced by recasts via such methods as coding, as suggested by Gass and Valmori (2015). In the case of recasts, this level of engagement may be manifested in some learners by simply parroting or repeating the feedback they receive, possibly involving immediate recall from a short-term memory store. In contrast, using the correct form learners receive through recasts later in the discourse places higher demands on WM, possibly because learners need to recall their earlier utterance while exploring possibilities for modification (Gass, 2006).

7. The replication study

The first two research questions of this study are the same as the ones in the initial study while the third one extends the research by following up on the learners' reactions to recasts:

1. Does the type of outcome measure employed influence the observed effects of recasts on L2 development?
2. Are PSTM and/or complex WMC related to whether any effects of recasts are observed on different types of outcome measures?
3. What is the relationship between recast-induced behavioral engagement and PSTM and/or complex WMC?

7.1. Method

As the current study is a conceptual replication, we attempted to reduce heterogeneity between the two studies in terms of method, in line with the recommendations of Marsden et al. (2018), with the data

being elicited in the same way, and the participants' age, proficiency level and the learning context being similar. The differences concerned the samples' first languages (L1s), the language of the WM tests, and the addition of the variable LEARNER ENGAGEMENT in the current study, which led to the third research question.

7.1.1. Setting and participants

The current study examined the effect of different outcome measures on the effectiveness of recasts and the relationship of learners' WM capacity with their gains from recasts in terms of their engagement with the recasted form on different task types. The participants were 90 EFL learners of a beginner level at a language school in Iran. Biographical information collected about the participants indicated that their age ranged from 18 to 22 ($M = 18.90$, $SD = 0.6$) and that none had been to an English-speaking country. The approach employed at the language school consisted of an amalgamation of mechanical drills and communicative activities. A pretest, consisting of only the GJT and the written production task, was administered to an original pool of 120 participants. Based on the pretest results, 102 students who showed a lack of knowledge of the target construction were invited to participate in the oral task. Finally, 90 students, 50 males and 40 females, all native speakers of Persian, were randomly selected to take part in the study. Their prior English instruction lasted from two to three years ($M = 2.6$, $SD = 0.2$), and was not found to be significantly different across the three groups.

7.1.2. Materials and procedures

The materials and procedures in the current study were similar to those of Révész (2012). Persian versions of the RS test, DS and NWS tasks were developed and piloted on ten individuals, yielding the internal consistencies of 0.94 and 0.96, respectively. Also, test-retest reliability of the measures was obtained with a five-week interval and reported to be $r = 0.76$, $p < 0.01$ for the DS task and $r = 0.74$, $p < 0.01$ for the NWS task. The tasks were used in line with the guidelines in Conway et al. (2005). The target structure chosen for the current study was, as in Révész (2012), one usage of the past progressive construction and was presented in the same way as in Révész (2012), that is, a description task requiring the participants to describe what was happening in a picture, with one of the researchers taking on the role of the police officer. The reason for not opting for a target structure different from that of Révész (2012) was that, given a different setting, the replication study would have been potentially incomparable to the initial study. Developmental readiness for the target structure was operationalized as the second stage that learners go through to develop knowledge of this construction (Bardovi-Harlig, 2000). In the first stage, the learner has knowledge of bare progressive (e.g., *walking*). In the second stage, they gain knowledge of present progressive (e.g., *is walking*), followed by the third stage in which past progressive is acquired (e.g., *was walking*). Learners' developmental readiness was specified in the same way as in the initial article, namely based on the learners' performance on the written pretest (see Révész, 2012).

A coding method, intended to provide information on responses to recasts, was developed to measure the extent to which the students with different degrees of WM capacity would be engaged with recasts. The reason for the use of a coding scheme was that Révész (2012) attributed differences in performance to the extent to which the participants were engaged in recasts as a function of different WM capacities. In the current study, the coding of learners' reactions towards recasts, based on the categories suggested by Gass and Valmori (2015), ran the gamut from: (1) no opportunity, (2) opportunity, but did not repeat, (3) repeated the recasted form, (4) negotiated the response, to (5) used the recasted form later in the discourse. Examples of the categories are provided in the Appendix. The coding was carried out via collaboration between the first researcher and a research assistant who listened to the audio recordings of the recast-embedded episodes. Collaborative coding (Smagorinsky, 2008) was used to ensure the reliability of the decisions by providing room for discussion in relation to the data, with each decision being the outcome of a thoughtful exchange between the coders regarding what to call each and every data segment.

8. Results

8.1. Results of the GJT

The means and standard deviations for the performance of each group over time are presented in Table 2. The results of an analysis of variance showed that there were no significant differences between the groups on the pretest, $F(2, 127) = 0.15$, $p = 0.85$.

A mixed between-within analysis of variance was used to measure the impact of the treatment on participants' scores on the GJT across the pretest, posttest, and delayed posttest. With Greenhouse-Geisser corrections applied, there was a significant effect for Time × Group interaction, $F(4, 252) = 2.60$, $p = 0.04$, $\eta_P^2 = 0.03$ and a significant main effect for time, $F(2, 126) = 39$, $p = 0.00$, $\eta_P^2 = 0.23$, but no significant main effect for group, $F(2, 127) = 1.51$, $p = 0.22$. It is noteworthy that the data did not meet the assumptions of normal distribution and homoscedasticity, even after applying transformations, hence a possible loss of power to find statistical results, according to Larson-Hall (2010). Post-hoc pairwise comparisons indicated significant within-group gains in the performance of the two experimental groups over time but no significant difference between the groups. Cohen's (1988) effect sizes for comparisons between groups at posttest 1 were: recast versus non-recast, $d = 0.26$; recast versus control, $d = 0.45$; non-recast versus control, $d = 0.19$. Also, comparisons between groups at posttest 2 yielded the following contrast effect sizes: recast versus non-recast, $d = 0.15$; recast versus control, $d = 0.27$; non-recast versus control, $d = 0.11$.

8.2. Results of the oral description task

The descriptive statistics for the oral task description appear in Table 3. An analysis of variance found no significant difference between the three groups on the pretest, $F(2, 127) = 0.09$, $p = 0.91$.

A mixed between-within analysis of variance was run to examine the effect of the treatment on participants' scores on the oral task across the pretest, posttest, and delayed posttest. Significant effects were found for Time × Group interaction, $F(4, 254) = 54$, $\eta_P^2 = 0.46$, $p = 0.00$, and for group, $F(2, 127) = 41$, $p = 0.00$, $\eta_P^2 = 0.39$, and time, $F(2, 126) = 121$, $p = 0.00$, $\eta_P^2 = 0.48$, with Green-Geisser corrections applied. Note that the data did not meet the assumptions of normal distribution and homoscedasticity, even after the use of transformations. Post-hoc pairwise comparisons indicated significant within-group gains in the performance of the two experimental groups over time. Also, Games–Howell post-hoc pairwise comparisons revealed that both on posttests 1 and 2, the recast group performed significantly better than the other two groups, at which time the difference between the non-recast and control group did not reach significance. Contrast effect sizes for comparisons between groups at posttest 1 were as follows: recast versus non-recast, $d = 2.29$; recast versus control, $d = 2.56$; non-recast versus control, $d = 0.49$. Also, comparisons between groups at posttest 2 yielded the following effect sizes: recast versus non-recast, $d = 1.80$; recast versus control, $d = 1.61$; non-recast versus control, $d = 0.02$.

8.3. Results of the written description task

Table 4 presents the group means and standard deviations for each condition over time. The results of an ANOVA showed a similar performance by the three groups on the pretest, $F(2, 127) = 1.00$, $p = 0.36$.

A mixed between-within analysis of variance was conducted on the pretest, posttest, and delayed posttest scores of the written description tasks. There was a significant effect for Time × Group interaction, $F(4, 254) = 15$, $\eta_P^2 = 0.19$, $p = 0.00$, a significant effect for group, $F(2, 127) = 27$, $\eta_P^2 = 0.30$, $p = 0.00$, and time, $F(2, 126) = 21$, $\eta_P^2 = 0.14$, $p = 0.00$, with homoscedasticity and normal distribution assumptions being violated and the Green-Geisser corrections applied. Post-hoc pairwise comparisons indicated significant within-group gains in the performance of only the recast group over time. Also, Games–Howell post-hoc pairwise comparisons showed that the recast group performed significantly better than the other two groups on both posttests 1 and 2, at which time the difference between the

Table 2. Group means and standard deviations for GJT

	Pretest		Posttest 1		Posttest 2	
Groups	M	SD	M	SD	M	SD
Recast ($n = 45$)	17.88	3.24	20.08	4.43	19.51	4.13
Non-recast ($n = 45$)	17.60	4.00	18.97	4.10	18.86	4.20
Control ($n = 40$)	17.45	3.98	18.22	3.67	18.37	4.14

Table 3. Group means and standard deviations for oral description task

	Pretest		Posttest 1		Posttest 2	
Groups	M	SD	M	SD	M	SD
Recast ($n = 45$)	7.64	2.46	14.84	2.55	13.37	3.13
Non-recast ($n = 45$)	7.44	2.51	9.20	2.36	8.64	2.00
Control ($n = 40$)	7.65	2.80	7.90	2.86	8.57	2.80

Table 4. Group means and standard deviations for written description task

	Pretest		Posttest 1		Posttest 2	
Groups	M	SD	M	SD	M	SD
Recast ($n = 45$)	8.06	0.83	10.15	1.78	9.86	1.39
Non-recast ($n = 45$)	8.24	0.64	8.48	1.60	8.22	1.18
Control ($n = 40$)	8.00	0.98	8.40	1.25	7.85	1.07

non-recast and control group was not significant. Contrast effect sizes for comparisons between groups at posttest 1 were as follows: recast versus non-recast, $d = 0.98$; recast versus control, $d = 1.13$; non-recast versus control, $d = 0.05$. Similar comparisons between groups at posttest 2 yielded the following effect sizes: recast versus non-recast, $d = 1.27$; recast versus control, $d = 1.62$; non-recast versus control, $d = 0.32$.

As for the second research question, a series of correlation tests conducted on each of the WM test scores and gains on the three measures yielded two significant correlations only in the recast group: one between the oral gains and NWS scores ($r = 0.88$, $p = 0.00$) and one between the gains on the written description task and the RS scores ($r = 0.73$, $p = 0.00$).

8.4. Recast-induced engagement and WM

To address the third research question concerned with exploring the relationship between recast-induced engagement and WM capacity, a series of Pearson product moment correlations was conducted. The data included the participants' engagement scores calculated based on the average of coding scores obtained by each participant in the recast group and the scores on the DS, NWS, and RS tasks. A significant correlation was found between the DS and engagement scores ($r = 0.63$, $p = 0.00$), and NWS and engagement scores ($r = 0.77$, $p = 0.00$), but the correlation between RS and engagement scores did not reach significance ($r = -0.05$, $p = -0.7$).

9. Discussion and conclusion

According to the statistical analyses, the answer to the first research question for the most part echoed the finding reported in the initial study. On the one hand, it was indicated that the recast group had an

advantage over the non-recast and control groups on two of the measures, that is, the written and oral description tasks. On the other, unlike what was the case in Révész (2012), the recast group did not significantly outperform the other two groups on the GJT on the posttest and the delayed posttest. However, consistent with Révész's finding, recasts seemed to engender a greater impact on the oral production task than the written production task. We arrived at this finding on the grounds that, based on the new field-specific benchmarks for interpreting effect sizes (Plonsky & Oswald, 2014), greater comparison effect sizes were found for the oral production task. This could suggest that recasts were more conducive to the later use of the target structures on such tasks. Furthermore, the pairwise comparisons revealed within-group gains on the written task scores only in the recast group but for both experimental groups on the oral production task. Therefore, it could be argued that the difference between the two tasks is not as pronounced as in the initial study.

The idea that the GJT, unlike the oral and written production tasks, was least impacted by recasts may lie in the different nature of knowledge that it seems to draw on. The GJT, by nature, involves the use of metalinguistic knowledge, which is declarative, whereas the oral production task, which benefited most from the provision of recasts, draws on procedural knowledge. Written production tasks could be argued to fall somewhere between the declarative-procedural knowledge types continuum, allowing for unpressured language production on the one hand and a focus on meaning on the other. Therefore, as reported in a number of previous studies (e.g., Ellis et al., 2006; Révész & Han, 2006), and in the initial study, recasts are more geared toward tasks that require procedural knowledge.

A possible explanation for the non-effectiveness of recasts for the GJT in this study may be that the participants in our study were from a slightly different instructional context. In other words, the participants in the two studies had different L1 backgrounds. It remains to be seen if a different L1 background, possibly interacting with learners' RS, could predispose learners to perform differently on certain WM measures. That is, the language of the WM tasks could have acted as an intervening variable, despite the claim that WM tasks are language independent (Osaka & Osaka, 1992).

The reason the effects found for recasts were weaker, compared with the ones in the initial study, could be sought in what has come to be referred to as the Proteus effect, where large effects are reported in the early phases of an area of research, with the effects gradually regressing toward the mean effect and even non-significance. However, as Thompson (2001) aptly warned, interpreting effect sizes with fixed benchmarks would merely be being stupid in another metric. With that in mind, in evaluating the potential benefits of recasts in the present study, it could be argued that even the poorer effects reported in this replication study could justify the treatment. That is, the minimal manipulation and effort associated with delivering recasts, coupled with the positive outcomes they can induce in learners' L2, encourage the use of this implicit type of feedback since the investment of time and effort seems to be lower, compared with explicit kinds of feedback. Alternatively, it could be that the violation of normal distribution and homoscedasticity assumptions in the current study has led to a loss of power to find statistical differences between the groups.

The better performance of the recast group on the oral production task could be attributable to the notion of transfer-appropriate processing (TAP) (Lightbown, 2008; Révész, 2012). TAP basically posits that learning considerably hinges on the extent to which the learning and retrieval conditions are similar. The implication of this for the results of the current study would be that the retrieval conditions for the GJT were quite dissimilar to the conditions afforded by recasts, which occurred in a communicative context with a primary focus on meaning and the existence of time pressure, whereas the oral task conditions were congruent with the nature of the recasts, hence better gains. From this perspective, the superiority of the recasts on the written description task, compared with the GJT, could possibly be attributed to the idea that in the former, the participants were required to describe pictures, similarly to what was the case during the oral tasks in which recasts were delivered.

The second research question dealt with the relationship between PSTM and/or complex WMC and the recast-induced gains on different outcome measures. As in the initial study, a possible explanation for the higher engagement levels reached by high PSTM learners is that a high phonological capacity possibly predisposes learners to go beyond merely repeating their interlocutor's utterance and

venture more demanding responses to recasts. It is possible that participants with high complex WM, on the other hand, were not intrinsically primed for such noticing. This result was confirmed by the follow-up coding method employed to address the third research question concerned with the relationship between recast-induced behavioral engagement and PSTM and/or complex WMC. It should be noted that the correlation found in the initial study between GJT gain scores and the RS scores and the one between the oral task gain scores and the DS scores were absent in the replication study, possibly showing that WM tests could be influenced by the language being used, unlike what has been claimed to be the case (Osaka & Osaka, 1992).

The novelty of this replication study was the addition of the variable engagement and the significant correlation found between DS and engagement scores, and that between NWS and engagement scores and the lack thereof between RS and engagement scores. Future research can use more ecologically sound approaches, such as motivation questionnaires and introspective interviews including stimulated recall, to look more closely at the learners' level of engagement and provide insight into learner perceptions (Philp & Duchesne, 2016) in relation to recasts from a multidimensional perspective, reflecting the cognitive, social, affective, and behavioral aspects of student engagement.

References

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. doi:10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839. https://psycnet.apa.org/doi/10.1038/nrn1201

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–90). Academic Press.

Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Blackwell.

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. doi:10.3758/BF03196772

Csikszentmihalyi, M. (1990). *Flow*. Harper & Row.

DeKeyser, R. (2007a). Situating the concept of practice. In R. DeKeyser (Ed.), *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 1–18). Cambridge University Press.

DeKeyser, R. (2007b). Skill acquisition theory. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition* (pp. 97–113). Erlbaum.

DeKeyser, R. (2009). Cognitive-psychological processes in second language learning. In M. H. Long, & C. J. Doughty (Eds.), *Handbook of second language teaching* (pp. 119–138). Blackwell.

Doughty, C. J. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206–257). Cambridge University Press.

Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 256–310). Blackwell.

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*(2), 305–352. doi:10.1017/S027226310505014X

Ellis, N. C., & Sinclair, S. (1996). Working memory in the acquisition of vocabulary and syntax. *Quarterly Journal of Experimental Psychology*, *49*(1), 234–250. doi:10.1080/713755604

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*(2), 227–275. doi:10.1111/j.1467-9922.2004.00255.x

Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning*, *51*(2), 281–318. doi:10.1111/1467-9922.00156

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*(2), 339–368. doi:10.1017/S0272263106060141

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354–375. doi:10.1093/applin/21.3.354

Fredricks, J., Blumenfeld, P., & Paris, A. (2004). School engagement: Potential of the concept, state of evidence. *Review of Educational Research*, *74*(1), 59–105. doi:10.3102%2F00346543074001059

French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, 29(3), 463–487. doi:10.1017/S0142716408080211

Gass, S. (2003). Input and interaction. In C. Doughty, & M. Long (Eds.), *The handbook of second language acquisition* (pp. 224–255). Blackwell.

Gass, S. M. (2006). *Input, interaction and output: An overview. AILA review* (pp. 3–17). Benjamins.

Gass, S., & Valmori, L. (2015). Replication in interaction and working memory research: Révész (2012) and Goo (2012). *Language Teaching*, 48(4), 545–555. doi:10.1017/S0261444815000038

Gass, S. M., & Varonis, E. M. (1989). Incorporated repairs in nonnative discourse. In M. R. Eisenstein (Ed.), *The dynamic interlanguage: Empirical studies in second language variation* (pp. 71–86). Plenum Press.

Goldschneider, J., & DeKeyser, R. (2001). Explaining the "natural order of L2 morpheme acquisition" in English. A meta-analysis of multiple determinants. *Language Learning*, 51(1), 1–50. doi:10.1111/1467-9922.00147

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, 34(3), 445–474. doi:10.1017/S0272263112000149

Goo, J. (2016). Corrective feedback and working memory capacity: A replication. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 279–302). John Benjamins.

Hawkes, L., & Nassaji, H. (2016). The role of extensive recasts in error detection and correction by adult ESL students. *Studies in Second Language Learning and Teaching*, 6(1), 19–41. doi:10.14746/ssllt.2016.6.1.2

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

Language Teaching Review Panel. (2008). Replication studies in language learning and teaching: Questions and answers. *Language Teaching*, 41(1), 1–14. doi:10.1017/S0261444807004727

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge.

Leow, R. (2015). *Explicit learning in the L2 classroom. A student-centered approach*. Routledge.

Lightbown, P. M. (2008). Transfer appropriate processing as a model for classroom second language acquisition. In Z. Han (Ed.), *Understanding second language process* (pp. 27–44). Multilingual Matters.

Linacre, J. M. (1989). *Many-facet rasch measurement*. MESA Press.

Loewen, S., & Nabei, T. (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* (pp. 361–377). Oxford University Press.

Long, M. H. (2007). *Problems in SLA*. Lawrence Erlbaum.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, 19(1), 37–66. doi:10.1017/S0272263197001034

Lyster, R., & Ranta, L. (2013). Counterpoint piece: The case for variety in corrective feedback research. *Studies in Second Language Acquisition*, 35(1), 167–184. doi:10.1017/s027226311200071x

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32(2), 265–302. doi:10.1017/S0272263109990520

Mackey, A. (2007). *Conversational interaction in second language acquisition: A collection of empirical studies*. Oxford University Press.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 407–452). Oxford University Press.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal*, 82(3), 338–356. doi:10.1111/j.1540-4781.1998.tb01211.x

Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning*, 60(3), 501–533. doi:10.1111/j.1467-9922.2010.00565.x

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181–209). John Benjamins.

McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition*, 27(1), 79–103. doi:10.1017/S0272263105050047

McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning*, 56(4), 693–720. doi:10.1111/j.1467-9922.2006.00393.x

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning*, 68(2), 321–391. doi:10.1111/lang.12286

Nassaji, H. (2007). Elicitation and reformulation and their relationship with learner repair in dyadic interaction. *Language Learning*, 57(4), 511–548. doi:10.1111/j.1467-9922.2007.00427.x

Nassaji, H. (2016). Anniversary article: Interactional feedback in second language teaching and learning: A synthesis and analysis of current research. *Language Teaching Research*, 20(4), 535–562. doi:10.1177%2F1362168816644940

Nassaji, H. (2017). The effectiveness of extensive versus intensive recasts for learning L2 grammar. *Modern Language Journal*, 101(2), 353–368. doi:10.1111/modl.12387

Nassaji, H. (2019). The effects of recasts versus prompts on immediate uptake and learning of a complex target structure. In DeKeyser, R. M. & Botana, G. P. (Eds.), *Doing SLA Research with Implications for the Classroom: Reconciling Methodological Demands and Pedagogical Applicability* (pp. 107–126). John Benjamins Publishing Company.

Norris, J. M., & Ortega, L. (2003). Defining and measuring SLA. In C. J. Doughty, & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 716–761). Blackwell.

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, 27(3), 377–402. doi:10.1017/S0142716406060322

Osaka, M., & Osaka, N. (1992). Language independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society*, 30(4), 287–289. doi:10.3758/BF03330466

Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, 36(4), 573–595. doi:10.2307/3588241

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, 36, 50–72. doi:10.1017/S0267190515000094

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. doi:10.1111/lang.12079

Racsmány, M., Lukács, A´, Németh, D., & Pléh, C. (2005). A verba´lis munkamem´oria magyar nyelv^u vizsg´al´oelj´ar´asai [Verbal working memory testing procedures in Hungarian]. *Pszichol´Ogiai Szemle*, 60, 479–506.

Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, 62(1), 93–132. doi:10.1111/j.1467-9922.2011.00690.x

Révész, A., & Han, Z.-H. (2006). Task content familiarity, task type, and efficacy of recasts. *Language Awareness*, 15(3), 160–179.

Révész, A. (2007). *Focus on form in task-based language teaching: Recasts, task complexity, and L2 learning*. Unpublished doctoral dissertation. Teachers College, Columbia University, New York.

Robinson, P. (1995). Attention, memory and the noticing hypothesis. *Language Learning*, 45(2), 283–331. doi:10.1111/j.1467-1770.1995.tb00441.x

Ross-Feldman, L. (2007). Interaction in the L2 classroom: Does gender influence learning opportunities? In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 53–77). Oxford University Press.

Sagarra, N. (2007). From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 229–248). Oxford University Press.

Saito, K. (2013). The acquisitional value of recasts in instructed second language speech learning: Teaching the perception and production of English to adult Japanese learners. *Language Learning*, 63(3), 499–529. doi:10.1111/lang.12015

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. doi:10.1093/applin/11.2.129

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41(2), 255–283. doi:10.1002/j.1545-7249.2007.tb00059.x

Sheen, Y. (2008). Recasts, language anxiety, modified output, and L2 learning. *Language Learning*, 58(4), 835–874. doi:10.1111/j.1467-9922.2008.00480.x

Sheen, Y. (2011). *Corrective feedback, individual differences and second language learning*. Springer.

Smagorinsky, P. (2008). The method section as conceptual epicenter in constructing social science research reports. *Written Communication*, 25(3), 389–411. doi:10.1177%2F0741088308317815

Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 125–144). Oxford University Press.

Thompson, B. (2001). Significance, effect sizes, stepwise methods, and other issues: Strong arguments move the field. *Journal of Experimental Education*, 70(1), 80–93. doi:10.1080/00220970109599499

Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytical ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 144–171). Oxford University Press.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology*, 49(1), 51–79. doi:10.1080%2F713755607

Appendix:

Example 1. No opportunity

Learner: A man *comes and then he *walk and there is a girl…

Example 2. Ignoring the opportunity to repeat the recasted form.

Learner: A *girl is talking with a boy.
Researcher: Yeah, a girl *was talking* with a boy.
Learner: Ok, then, there was a man …

Example 3. Repeating the recasted form.

Learner: A man *walk in the park.
Researcher: Yeah, a man *was walking* in the park.
Learner: Yes, a man was walking in the park.

Example 4. Negotiating the response.

Learner: There was a boy. I think he *runs very fast.
Researcher: Yeah, he *was running* very fast.
Learner: You mean I should say he was running?
Researcher: Yes, go ahead.

Example 5. Using the recasted form later in the discourse

Learner: Two men *are walking in the street.
Researcher: Yes, two men *were walking* there.
Learner: And two girls were sitting, I think.