White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/208106/

Version: Published Version

Article:
Mifka-Profozic, Nadia, Behney, Jennifer, Gass, Susan M et al. (2023) Effects of form-focused practice and feedback: a multisite replication study of Yang and Lyster (2010). Language Learning. 1164–1210. ISSN: 0023-8333

https://doi.org/10.1111/lang.12623

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

Check for updates

REGISTERED REPORT

Effects of Form-Focused Practice and Feedback: A Multisite Replication Study of Yang and Lyster (2010)

Nadia Mifka-Profozic,[a] Jennifer Behney,[b] Susan M. Gass,[c] Marijana Macis,[d] Gaia Chiuchiù,[e] and Giulia Bovolenta[a]

[a]University of York [b]Youngstown State University [c]Michigan State University [d]Manchester Metropolitan University [e]Accademia Lingua Italiana Assisi

Abstract: We conducted a multisite replication of Yang and Lyster's (2010) study investigating the effects of recasts and prompts on learning English regular and irregular past

---

CRediT author statement – **Nadia Mifka-Profozic**: conceptualization; methodology (equal); investigation (equal); writing–original draft preparation (equal); writing–review and editing (supporting); data curation (equal); formal analysis (equal); project administration; **Jennifer Behney**: conceptualization; methodology (equal); investigation (equal); data curation (equal); formal analysis (supporting); writing–original draft preparation (supporting); writing–review and editing (supporting); **Susan M. Gass**: conceptualization; methodology (equal); investigation (equal;) writing–original draft preparation (equal); writing–review and editing (lead); data curation (equal); formal analysis (equal); **Marijana Macis**: resources, investigation (equal), data curation (equal), funding acquisition; **Gaia Chiuchiù**: investigation (equal), data curation (equal); **Giulia Bovolenta**: formal analysis (linear mixed effects), visualization.

A one-page Accessible Summary of this article in nontechnical language is freely available in the Supporting Information online and at https://oasis-database.org

We are grateful for the cooperation of teachers and administrators in Bosnia and Italy. They were cooperative and eager to help with our project. We are also thankful to Sible Andringa and Aline Godfroid for bringing us together to share common interests. Their helpful comments and those of anonymous reviewers as we wrote our registered report helped us refine our questions and sharpen our arguments. Finally, we are grateful to Luke Plonsky who was always prompt (no pun intended) in responding to statistical questions. His advice was fast and thoughtful—thank you, Luke, for your help. Of course, we acknowledge that all errors that remain (if there are any!) are our own.

Correspondence concerning this article should be addressed to Nadia Mifka-Profozic, University of York, York YO10 5DD, UK. Email: nadia.mifka-profozic@york.ac.uk

The handling editor for this article was Sible Andringa.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Language Learning* published by Wiley Periodicals LLC on behalf of *Language Learning* Research Club, University of Michigan.
DOI: 10.1111/lang.12623

tense. Our study was conducted with intact high school and vocational school classes in Italy and Bosnia. Our participants were young adolescents (14–15 and 16–17 years old), a population that has been largely ignored in second language acquisition (SLA) research. We followed the design of the original study, but we also included a few modifications regarding the elicitation materials. The findings from our study did not fully align with Yang and Lyster's results. We found no effect of group and no evidence of the superiority of either prompts or recasts in either written or oral data in either Bosnia or Italy. However, we found a steady increase in scores over time from pretest to posttests in oral data in all groups at both sites.

Keywords corrective feedback; prompts; recast

Introduction

By taking stock of the extent to which findings from previous second language acquisition (SLA) research can be generalized, the project *SLA for All* is intended to advance researchers' understanding of how nonprimary languages are learned. Convenience sampling predominates in much of the behavioral research literature including SLA research (see Plonsky's, 2015, estimate of 67% coming from postsecondary students). As Andringa and Godfroid (2020) noted, "[i]f the selection of participants is somehow biased, the reliability of researchers' statements about the behavior under investigation is compromised" (p. 134). They proceeded to point out that in psychology it had been shown (Arnett, 2008) that samples were primarily drawn from WEIRD groups, that is, from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies (Henrich et al., 2010). The current replication study[1] focuses on oral corrective feedback (CF) and attempts to fill the age gap of most current research within this domain. The majority of CF studies rely on participants who are university students or, to a lesser extent, on preadolescents. The participants in this study represent an understudied population, namely early high school students; the study thus provides a more complete understanding of the role of feedback from childhood to the university-aged population. The specific population of "inbetween" learners, namely those of high school age (i.e., 14–18), does not figure prominently in the SLA literature. Yet, because an understanding of the effects of feedback is dependent on concepts that rely on increased cognitive maturity, such as noticing, working memory, attention, and awareness (Gass & Mackey, 2020; Schmidt, 2001), it is important to have data that take us through the entire age continuum. The participants of this study are less cognitively mature than university students and more mature than those classified as younger children (Nicholas & Lightbown, 2008), thus offering a complementary perspective on feedback effectiveness.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Cognitive maturity can be conceptualized as a continuum[2] that in some ways corresponds to an age continuum, especially for people in the educational system. Newport (1990) argued that perception and memory capacity are central to language processing and representation, and pointed out that these cognitive abilities are age-related. Importantly, Newport (1991) included the older child (not specifically defined) in a category with adults and considered the possibility that:

> the cognitive limitations of the young child during the time of language learning may likewise provide a computational advantage for the acquisition of language, and that the less limited cognitive abilities of the older child and the adult may provide a computational disadvantage for the acquisition of language. (p. 125)

Whether adolescents pattern like adults in their ability to notice and respond to feedback is an empirical question. Berk (2006) pointed out that school divisions (e.g., elementary, middle and high school) correspond to changes in ways of thinking, with abstract thinking and memory capacity changing with greater cognitive maturity. As children grow, there is an increasing ability for abstract thought and the ability to focus on linguistic analyses (Berman, 2007; Muñoz, 2007).

CF is particularly important because of the cognitive mechanisms involved in responding to feedback (recasts/prompts); the ultimate goal for a learner is incorporating that feedback into their second language linguistic system. This process requires multiple stages. Most important is the concept of attention. In all types of feedback, attention is drawn to an error; this can be done through explicit metalinguistic correction (e.g., *That is wrong. The past tense must be used*.) or implicit correction, as in recasts where an incorrect utterance is repeated with a corrected form in the next turn.

For recasts to be successful (i.e., resulting in an internalized correct form), learners must engage in a complex process that requires cognitive sophistication. That is, they must:

1. notice a difference between the input (in our case, oral input) and their own linguistic system (which in and of itself involves a sensitivity to their own system);
2. understand the source of the difference (at times made even more difficult when the feedback involves a complex correction or is in itself complex because multiple differences are targeted;

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3. be able to understand how to correct;
4. incorporate the corrected version into their linguistic system.

Prompts, on the other hand, do not provide input. Rather, they elicit forms from the learner. Therefore, for prompts to be successful, learners must have some knowledge of the form being corrected. Unlike recasts where the learner may or may not know the correct form, with prompts, a reformulated response requires learners to draw on their own linguistic resources.

Thus, in both feedback types, learners must pay attention to language, and their memory is invoked to make comparisons. As an individual increases in age and in education level, these tasks become more commonplace. An important part of this continuum is missed if only younger learners and older learners are the focus of research.

In general, it is clear that findings cannot be generalized to all age groups or populations that are less educated/less literate than university-level students (see also Mackey & Sachs, 2012), who represent the majority of SLA studies. Within the context of interaction, Oliver (1998) made a similar claim when she said that "findings from adult studies cannot be generalized to child studies without adequate and appropriate research involving child learners" (p 372). It should not be assumed that learning mechanisms necessary for L2 learning remain the same through the age continuum. As an example, consider a study on massed and spaced instruction by Kim and Webb (2019). They reported data from primary, secondary, and university students, finding differences across the groups. Memory capacities and cognitive reasoning differed across the ages; it can therefore be expected that corrective feedback, which involves both memory capacity and cognitive reasoning, is also likely to differ (see Gass et al., 2013). Thus, without a full investigation of the entire age continuum, it is impossible to understand if previous results are skewed given the preponderance of studies of adults and a more limited number of studies of preadolescent children.

Background Literature

Corrective Feedback

The current study investigates CF in the form of recasts and prompts. Both feedback types have been seen to be effective in language learning (Long, 2007). In particular, studies have shown that feedback is useful in many instances (although not all) as learners develop second language (L2) knowledge (see in particular Nassaji & Kartchava, 2017, 2021, as well as overviews and meta-analyses [Braidi, 2002; Brown, 2016; Chen, 2016; Ellis & Sheen, 2006; Gass & Mackey, 2020; Kang & Han, 2015; Li, 2010; Loewen, 2012;

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Loewen & Sato, 2018; Long, 2007; Lyster et al., 2013; Mackey, 2012; Nassaji, 2016; Nicholas et al., 2001; Plonsky & Brown, 2015; Russell & Spada, 2006; Yousefi & Nassaji, 2019]). These studies include those that demonstrate the differential effects of recasts and prompts as well as other types of feedback. Of these, recasts (see Example 1) are perhaps the most contentious type of oral CF resulting in controversy related to their effectiveness (Goo & Mackey, 2013; Lyster & Ranta, 2013).

(1) Nonnative speaker (NNS): …the boy is holding the girl hand… ← error–trigger
Native speaker (NS): …the boy is holding the **girl's** hand. ← **recast**
Oliver (1995, p. 473)

In 2007, Long noted that there had already been more than 60 studies that looked into the effects of recasts. The interest in this type of feedback has grown, with studies involving a comparison between recasts and explicit correction, recasts and prompts, or recasts and one specific type of prompt (e.g., clarification requests, elicitation). Within this research tradition, the participants in most of the studies have been pooled from a skewed population, most notably from highly literate, college or university educated students with English either as their first language (L1) or L2. Children, adolescents, and older learners have been underrepresented in all areas of applied linguistics research, as we noted earlier. There are notable exceptions, and in this review of CF, we emphasize studies conducted with participants who do not represent the majority adult university/college population (cf. Plonsky, 2015).

Most research has considered the role of recasts and various other forms of correction, generally grouped under the umbrella term of negotiation for meaning and/or prompts. Working within the interactionist framework, Long (2007) defines recasts as:

> …a reformulation of all or part of a learner's immediately preceding utterance in which one or more nontargetlike (lexical, grammatical, etc.) items is/are replaced by the corresponding target language form(s), and where, throughout the exchange, the focus of the interlocutors is on meaning, not language as object. (p. 77)

Whatever view one maintains, the basic notion is the same: Information is provided to a learner that an utterance is erroneous in some way (e.g., phonological, syntactic, pragmatic).

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Within the interactionist tradition the juxtaposition of incorrect/correct forms can facilitate noticing of the discrepancy between what has been said by the learner and what the NS (or the more proficient speaker) has said. Within the child language literature, Saxton (1997, 2000) proposed the direct contrast hypothesis focusing on the contingency of erroneous and correct forms. It is the contingency that fosters an understanding of a contrast between two forms. Goo and Mackey (2013, pp. 129–130) make a similar argument pointing to the benefits of immediate juxtaposition (semantic transparency, salience, and ease of comparison of erroneous and target language forms). Cognitive resources are required for a successful outcome.

Recasts are more subtle than other forms of feedback and may not be noticed by the learner. Further, it is often difficult to determine their effectiveness because the participatory demands on the part of the learner are minimal. Additionally, a lack of response to a recast makes it difficult to interpret its effectiveness, as has been pointed out by, inter alia, Oliver (1995), Oliver et al. (2008), Bryfonski and Sanz (2018), and Gass and Mackey (2020). There are several reasons for a lack of response: One possibility is that learners may be so focused on meaning that they do not recognize a difference between their output and the following response; alternatively, there may not be an opportunity to respond when a recast is provided and then is followed immediately by continued speaking, as is illustrated in (2):

(2) Learner: How many sister you have?
NS: How many sisters do I have? I have one sister
(McDonough & Mackey, 2006, p. 702)

In other words, recasts do not force a learner to come up with a revised form, making it difficult for a researcher to know whether there has been recognition of a correction and the extent to which processing has occurred.

Prompts, on the other hand, rather than providing a model of what is correct, alert the learner (with different degrees of explicitness) to a problem with a previous erroneous utterance. As noted earlier, for prompts to be successful, learners must have some prior knowledge of the form in question. Examples of prompts are seen in (3)–(6).

(3) Metalinguistic clue
Student: I went to the train station and pick up my aunt. ← error–trigger
Teacher: **Use past tense consistently**. ← **metalinguistic clue**
Student: I went to the train station and picked up my aunt.
(Yang & Lyster, 2010, p. 243)

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(4) Repetition

Student: Mrs. Jones travel a lot last year. ← error–trigger
Teacher: Mrs. Jones **travel** a lot last year? ← **repetition of error**
Student: Mrs. Jones travelled a lot last year.

(Yang & Lyster, 2010, p. 243)

(5) Clarification Request

Student: Why does he fly to Korea last year? ← error–trigger
Teacher: **Pardon? ← clarification request**
Student: Why did he fly to Korea last year?

(Yang & Lyster, 2010, p. 243–244)

(6) Elicitation

Student: Once upon a time, there lives a poor girl named Cinderella. ← error –trigger
Teacher: **Once upon a time, there… ← elicitation**
Student: there lived a girl.

(Yang & Lyster, 2010, p. 244)

Nonuniversity Populations

In this section we discuss the literature that has focused on nonuniversity populations, namely younger learners, older learners, and less literate learners. We begin with a discussion of younger learners. In general, the task of incorporating corrective feedback is not easy. Older children are better at using corrective feedback than younger children (Oliver, 1995, 1998, 2000, 2002, 2009; Oliver & Mackey, 2003).

In a study that involved two groups of children from Australian schools (native speakers and immigrants, ages 8–13) working in dyads on communicative tasks, Oliver (1995) found that implicit negative feedback, consisting of negotiation and recasts, followed 61% of the errors produced by learners. The children incorporated 9.93% of all recasts in their subsequent utterances. Oliver further noted that one third of all recasts would have been incorporated in subsequent speech if the learners had been given an appropriate opportunity to do so. In another study with immigrant children in Australia aged 8–12 (Mackey & Oliver, 2002), adult native speakers used interactional feedback in English question formation. The treatment included both recasts and prompts, but they were not teased apart. In the group that had received interactional feedback, eight out of 11 learners showed sustained development in question formation already by the first posttest, whereas in the control group only three out of 11 learners achieved the same improvement.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Whittle and Lyster (2016) also used both recasts and prompts[3] to provide interactional feedback on the use of Italian present indicative to a group of young Chinese children (second grade, approximately seven years old) learning Italian in an Italian school. The focus of their study was on the effects of form-focused instruction; the treatment effect in the experimental groups was strong ($d = 1.42$). Thus, even at a young age, children appeared to respond to the salience of the input created by corrective feedback, although it is not clear if this outcome was in part due to additional awareness raising activities. In a similar attempt to integrate prompts and recasts in learning the English passive construction, Li et al. (2016) conducted a study with 150 Chinese middle school learners of English as a foreign language (EFL). Prompts were followed by reformulation of the incorrect utterance if self-correction had failed. On both a grammaticality judgment test and an elicited imitation test the group that received both explicit instruction and CF outperformed the other groups that received either explicit instruction or CF only. On the elicited imitation test, the results became clear only when the learners were divided into those with some or no prior knowledge, which indicated that only those who had some prior knowledge benefited from CF combined with prior explicit instruction. In an earlier study, Doughty and Varela (1998) conducted a pretest–treatment–posttest–delayed posttest study in a content-based classroom in the United States of America, providing CF on errors in the formation of the English past simple tense. They successfully used a repetition of a student error followed by recasting to make recasts more salient to the learners.

More relevant to this review is the paucity of studies conducted within the interactionist framework and dealing with the adolescent population.[4] Yet a few exceptions exist. First, Havranek and Cesnik (2001) and Havranek (2002) investigated feedback that included a wide age range (from age 10 to university students). However, they organized their results by situational and linguistic differences and not by age differences. A second exception is a study by Mifka-Profozic (2014, 2015) who looked at the acquisition of aspect by L1 English/ L2 French 16-year-old students and compared two types of implicit feedback: recasts and clarification requests. Her studies involved the development of aspectual morphology and used a pretest-treatment-posttest-delayed posttest design. The studies found that recasts were more effective than clarification requests.

Within the same age group, Alcón and García Mayo (2008) investigated focus on form including recasts and prompts in a classroom of 14–15-year-old EFL learners in Spain. As they stated, this is "an age range that falls into

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

its own category, being neither 'classic' child SLA, nor classic adult SLA" (p. 177). They see the advantage for adolescents in a greater capacity for abstract thinking than younger learners. Adolescents are also able to reflect on language issues, "which could be an advantage for the use of form-focused instructional approaches" (p. 177). Their findings showed that focus-on-form activities within this population were successful particularly when the students themselves were aware of their own linguistic shortcomings.

In a study that directly compared age groups, García Mayo and Labandibar (2017) focused on two groups of teenagers (one with a mean age of 13.3 and the other with a mean age of 16.2). There was substantial evidence that learners in both age groups noticed feedback provided on a writing task; the younger learners focused more on lexical feedback and the older learners on grammatical/content feedback. From our perspective, these findings suggest a greater ability to notice feedback on the part of older, more cognitively mature students, although the authors discussed their results in terms of proficiency and not age differences.

As can be seen from the cited studies, older children are capable of noticing corrective feedback. Whether they will incorporate it in their subsequent utterances largely depends on factors such as their proficiency level or whether or not they have the opportunity to use the feedback. We expect that the learners in our study may be able to respond to feedback in a similar way to adults.

At the other end of the age spectrum, Mackey and Sachs (2012) conducted an exploratory study within the interactionist framework including feedback that focused on nine older learners, aged 65–89, all Spanish L1 speakers with English as their L2, of whom only four had education beyond high school. Of the nine participants, only four showed improvement on posttests and only two of the four maintained development at one of the two posttests. The other two showed development on one of the delayed posttests. It may be hard to disambiguate age and literacy in the Mackey and Sachs study: Three of the four participants who showed development (either immediate or sustained) had either graduated from university/college or had attended university/college. Only one of the participants who did not have a high school level of education showed development.

Finally, we look at literacy effects. In a replication of Philp (2003), Bigelow et al. (2006) investigated the role of recasts with low-literacy learners with a L1 Somali and a L2 English. Their participants were similar in age to those of the current study. Of their eight participants, six were high school age, and two were in their twenties. The authors found that differences in literacy levels were related to differences in recall accuracy. Referring to this study, Tarone

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and Bigelow (2007) pointed to research in cognitive psychology to show that literacy affects syntactic manipulation but not lexical segmentation (see also Tarone, 2010). Recasts are more easily noticed as lexical feedback than as morphosyntactic feedback (see also García Mayo & Labandibar, 2017; Mackey et al., 2000; Tsang, 2004, where this was the case for younger/less proficient learners).

Prompts Versus Recasts

In contrast to the studies that tested the effectiveness of either recasts or prompts compared to no feedback or a combination of recasts and prompts to no feedback, there are studies that compare recasts with prompts. This comparison has been done with both children and adults.

In a study of four French immersion classes in Canada with 179 young learners (10 and 11-year olds), Lyster (2004) compared the two feedback types. Prompts consisted of clarification requests, elicitations, and metalinguistic clues. The study had French grammatical gender as its target. The results on the posttests showed significant gains after the treatment for the three groups that received form-focused instruction, but prompts proved more effective than both recasts and no feedback.

Ammar and Spada (2006) also investigated the effectiveness of recasts and prompts in a study of Francophone learners learning the English possessive *his/her*. Sixty-four sixth grade students participated in a four-week study. This study showed that both experimental groups benefited from being exposed to CF in the classroom; but recasts benefited only higher-proficiency learners whereas prompts benefited both lower and higher proficiency learners. Thus, Ammar and Spada's study confirmed the results of Mackey and Philp's (1998) laboratory study: Learners who are developmentally ready to acquire a grammatical structure can benefit from recasts.

It seems that contrasting prompts and recasts, the method employed in the study we are replicating (Yang & Lyster, 2010), was also a motivation for some later studies that used the same or similar design. Doski and Cele (2018) investigated the effects of recasts and prompts on article errors in English learned as a third language by primary school children (sixth grade, 11–12 years old) in Kurdistan. In their study of 39 lower-intermediate Kurdish–Arabic bilinguals, the authors found that both recasts and prompts were more effective than no feedback, but on a delayed posttest, the prompt group outperformed the recast group, which indicated that prompts had longer lasting effects. Similarly, Van de Guchte et al.'s (2015) study of 64 14-year old Dutch learners explored the impact of recasts and prompts on learning two grammatical structures in

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

German as a FL: dative case after a preposition and comparatives. Both groups outperformed the control group on posttests, but prompts were superior to recasts and proved to be more effective for comparatives than for dative case.

With somewhat different goals, Nassaji (2009) compared recasts with one type of prompts (elicitation) in an adult population and found an advantage of recasts over elicitation. Rahimi and Zhang (2016) demonstrated different results. Their participants were 60 advanced adult EFL learners with Persian L1. Both prompts and recasts improved grammatical accuracy, but the prompt group outperformed the recast group on both immediate and delayed posttests. However, Lyster and Izquierdo (2009) investigated the effects of recasts and prompts delivered individually to French L2 undergraduate students and found that both prompts and recasts were equally effective. In research involving 74 Thai university students, McDonough (2007) examined the effects of recasts and clarification requests as a type of prompts on the emergence of the lexical activity aspect in English. She found no significant difference between the effects of recasts and clarification requests. Sato and Loewen (2018) examined four different conditions in which recasts and clarification requests were provided—: either each of them alone or preceded by metalinguistic instruction. The study was conducted with 83 university students in Chile who were corrected on their errors on 3rd person singular -s and on possessive determiner *his/her*. Results suggested that both types of feedback were equally effective when preceded by metalinguistic instruction, but clarification requests proved to be superior when there was no prior metalinguistic instruction.

In sum, both types of feedback are beneficial, but prompts in general have been shown to have a greater effect than recasts on younger learners. Proficiency is also undoubtedly an important variable in trying to understand feedback effects.

The Present Study

In the study we replicated, Yang and Lyster (2010), henceforth YL, compared the effectiveness of recasts and prompts in a Chinese university EFL context. They found that prompts facilitated the acquisition of rule-based regular past tense forms, whereas both recasts and prompts were shown to be facilitative of learning the exemplar-based irregular past. They explained their findings in terms of the learning benefits of modified output (Swain, 1985, 1998, 2005). They also related their findings to the dual mode system (Skehan, 1998) and speculated that "during online communication, prompts more than recasts trigger access to the rule-based system whereas recasts and prompts alike trigger access to the exemplar-based system" (Yang & Lyster, 2010, p. 259).

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

As we have shown, the study of CF has been approached from various perspectives and has included a range of participants, most of whom were university students, followed by younger children (preadolescents). What is clearly missing is research on other groups of learners, for example, older learners, learners with different literacy backgrounds, and younger adolescents, the largest group of foreign language learners in the world (Kormos & Sáfár, 2008) and a group investigated by the current study. To investigate the role of CF, our study examined two groups of first- and second-year high school students (ages 14–16) at two different sites, in Bosnia and in Italy.

There are a number of reasons that led us to replicate the YL (2010) study. First, we felt it important to replicate a study where there was substantial time for feedback to be effective. In the YL study, feedback was provided over a two-week period. Given the constraints on the time that we were allotted for our study, we felt that this time frame was appropriate. Second, we wanted to replicate a study where the materials were appropriate for our population; if, however, changes were needed, they could be introduced while being faithful to the original (see Appendix S1 in the online Supporting Information for the changes made). Third, we needed a target structure that was likely to be learnable by our population. In other words, we were interested in a target structure that our learners were familiar with but over which they lacked control. Finally, we wanted to replicate a study that has been important in the recent literature. With currently more than 320 citations, YL met that goal.

In order to replicate YL (2010), we used the same target structure as was used in their study. Their original arguments for selecting past tense included the following consideration:

1. the forms are introduced early in most textbooks;
2. despite the early introduction, learners are known not to have full control of these forms even well into advanced levels of proficiency[5] (Ellis et al., 2006);
3. regular and irregular forms allow an analysis of the effects of feedback on rule-based forms (regular) and item-based forms (irregular) with the same semantic and functional values.

The last of these differences (i.e., rule-based vs. item-based forms) has important theoretical value. A usage-based perspective would see the differences in regular and irregular forms as being dependent on frequency in the input (Ellis & Wulff, 2020) and not on processing differences. On the other hand, one could posit an account whereby the two forms are processed differently (Pinker & Ullman, 2002). In Pinker and Ulman's view, regular past tense involves

computation because learners have to add a morpheme to the base form of a verb and then make phonological adjustments. This process would be carried out by the procedural system. Irregular verbs, on the other hand, do not involve such computations; rather, one learns the form as an isolated unit. These units are stored in declarative memory. By looking at the effects of recasts and prompts in adolescents, one can determine the extent to which they may be guided by the same principles as adults, a group with greater cognitive maturity. Further, prompts are more likely to involve greater processing than recasts, and inherent in prompts is a greater degree of salience than there is in recasts. As we discussed earlier, prompts (if successful) require a learner to rely on prior knowledge. Recasts, on the other hand, do not call on prior knowledge in the learner's response. But it is important to note that uptake of recasts does imply some degree of noticing. As Tarone and Bigelow (2007) pointed out, "accurate uptake logically requires that the speaker notice the gap in form between their original utterance…and the recast" (p. 101). However, on the surface, uptake is manifested as repetition and does not obligatorily rely on prior knowledge.[6]

Our target structure is English past tense (both regular and irregular verbs). Much research has demonstrated that morphosyntactic feedback is less noticeable than lexical or phonological feedback (e.g., Carpenter et al., 2006; Mackey et al., 2000; Saito, 2015). Mackey et al. specifically investigated the perception of feedback and noted that morphosyntactic feedback was less successful than phonological or lexical feedback in terms of learners' awareness of what the feedback was targeting. Carpenter et al. and Saito investigated phonological feedback and highlighted the relevance of salience. More salient forms were more readily noticed and hence more likely to result in successful structure acquisition than nonsalient forms. With regard to our study, it is likely that CF on irregular past tense forms (because of their syllable change) is more likely to be perceived as more salient than the feedback on regular past tense forms, as the *-ed* ending may not be noticed, especially in the case of recasts.

In the YL study, prompts were more effective than recasts in learning regular past tense forms, but the two feedback types were equally effective in the learning of irregular past tense forms. Thus, each feedback type contributed to learning, albeit in different ways. When computation is involved, prompts have an effect. When learning requires exposure to exemplars (irregular forms), both forms of feedback are equally effective. The question then is: Have young adolescents developed systems that are both exemplar-based and rule-based, and are they able to utilize feedback in the same way as children and adults? The

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

current study intends to add to the state of the art by examining an underrepresented age group.

Our research questions are guided by those of YL and are as follows:

1. Do the groups that had CF outperform the control group where no CF was provided?
2a. Do recasts have a differential effect on the acquisition of regular versus irregular past tense verb forms?
2b. Do prompts have a differential effect on the acquisition of regular versus irregular past tense verb forms?
3a. Does the prompt group outperform the recast group for regular verbs?
3b. Does the prompt group outperform the recast group for irregular verbs?

We hypothesized that the young adolescents in our study would pattern in a similar way to the educated adults in the YL study, but given that their cognitive maturity is less developed, we anticipated slight differences. With regard to the two sites used in this study (Bosnia and Italy), we predicted that the findings would be similar due to the closeness in our participants' age (and hence cognitive maturity).

Method

Participants

Students

The present study was conducted with 10 intact classes, four in central Italy and six in Bosnia (of the six in Bosnia, two classes were placed in each of the three treatment groups). The classes in each location were assigned to one of three conditions, two treatment (prompt and recast) and one control. Each of the conditions was planned to have approximately 18–25 students. The participants in Bosnia were in their second year of vocational school, 16–17 years old, and the participants in Italy were in their first year of high school, 14–15 years old.

Italian students attending public schools begin their study of English in primary school. In first grade, students attend class for one hour a week; in second grade, English classes take up two hours; and from third to eighth grade, English instruction takes up three hours a week. Therefore, prior to the onset of this study, the participants in Italy had studied English formally for eight years and had had approximately 630 hours of instruction. In the school in Italy where our project took place, the students follow a textbook called "Talent 2" by Cambridge University Press (notional-functional syllabus).

The students in Bosnia also begin studying English in primary school (second grade). English classes last 45 minutes. The number of lessons per week

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

varies across grades. In second grade, students have one English class per week; from third to seventh grade, they have two English classes per week; in eighth grade, there are three English classes per week; in ninth grade, frequency decreases to two English classes per week. By the time students start their second year in vocational school, they will have studied English for approximately nine years and will have had 524 hours of formal English instruction. In the school in Bosnia where our project took place, the students follow a textbook called "Solutions" (preintermediate) by Oxford University Press (notional-functional syllabus).

Teachers

At each site, all teachers were native speakers of Italian (Italy) or Serbian (Bosnia) and were experienced English teachers.[7] In Bosnia, one local teacher taught the control group, and the researchers taught the recast group and the prompt group. In Italy, we had three local teachers teaching the control group, the recast group, and two experimental prompt groups[8], respectively. See Appendix S2 in the online Supporting Information for a visual representation of teacher roles.

Feedback Conditions

Training on feedback types took place prior to the onset of data collection. A common training booklet was used. The booklet included (a) a general introduction to the project and the purpose of the study; (b) a timeline of the project with tasks; and (c) information about feedback types. Training consisted of videos exemplifying feedback (videos prepared at the university of one of the researchers and commonly available videos) and written examples. Teachers were given instructions (e.g., prompts or recasts that they were expected to use; each teacher was told not to provide the feedback type given by the other group). Teachers in the prompt and recast groups were told to provide the appropriate feedback type as soon as possible after an error was made. In most cases, this would occur at the end of the sentence. For example, in the question and answer activity, if a student says *Yes, the concert beginned at 9*, the teacher would intervene after *at 9*. In the dictogloss activity, if the student said *Cinderella likes parties*, the teacher would say *Yes, Cinderella liked parties*. In all of our activities, student responses generally consisted of one sentence. Thus, CF easily came very close to the error produced. We used the study instruments as stimuli for the practice session so that the teachers would get used to giving feedback in the natural context of the classroom. The training

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

session continued until the teacher could consistently produce the appropriate feedback type.

Procedure

We followed the same procedure as YL (2010). The entire project took place over a 5-week period and included pretests, treatment sessions, and two posttest sessions (immediate and delayed). A background questionnaire (see Appendix S3 in the online Supporting Information) and student assent forms were administered a day before or on the day of the pretests. Parental consent forms were sent to parents approximately two weeks before data collection.

There were four treatment sessions, each lasting approximately 15–20 minutes. The first and the third sessions utilized a dictogloss format; the second a picture-cued narrative, and the fourth involved a questions and answers activity. The sessions took place over a 2-week period.

The pretests were administered 1–3 days before the treatment; the immediate posttest took place within three days after the treatment, and the delayed posttests were conducted two weeks later. Figure 1 is a graphic display of the design. For all tests, for purposes of counterbalancing, we used a Latin square design, with three subgroups in each class. Thus, in each class, three written and three oral tasks were used, with rotation occurring for each test (pretest, immediate posttest, delayed posttest).

Control

During the treatment tasks, the control group completed the same activities as the experimental groups, but with no feedback given following error production of past tense forms. Feedback, if any, was on content or vocabulary.

Materials

The materials used were, to the extent possible, taken from YL (2010), although some modifications had to be made for various reasons, as described below.

Treatment

The treatment session materials, two dictogloss tasks, one picture-cued narrative, and a questions and answers activity, are given in Appendix S4 in the online Supporting Information. One dictogloss was based on the fairy tale Cinderella, while the other used an adapted short story A Quiet Walk. In Appendix S4, we provide information about verbs (regular and irregular) used in the treatment tasks.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

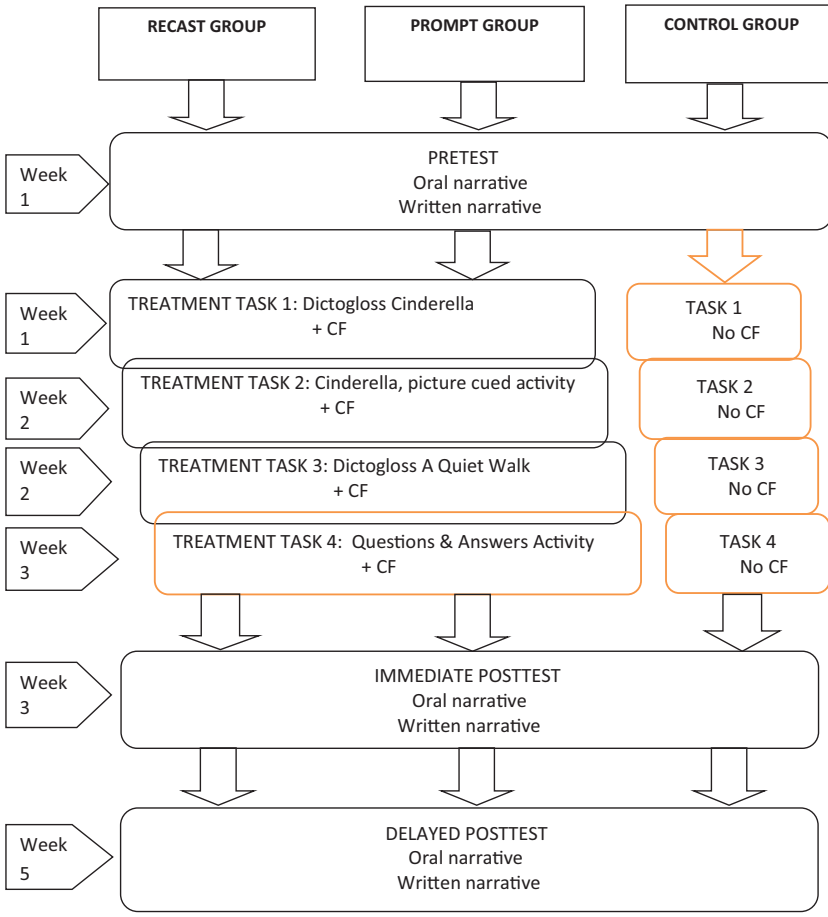

Figure 1 Design of the study.

For the dictogloss tasks, the students were divided in pairs, and the teacher read the story twice at normal speed. In the first reading, the students did not take notes; in the second reading, they wrote as much as possible of what they heard from the teacher. In each pair, students then compared their notes and attempted to reconstruct the text, after which each individual in each pair was asked to narrate part of the story. As the students narrated the story, the teachers in the treatment groups provided either recasts or prompts to the learners' erroneous utterances.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The picture-cued activity, based on the Cinderella story, contained 10 pictures that had to be described in order to complete the whole story. Each group of 2–3 students received one picture. The students were given 2–3 minutes to prepare sentences to describe their picture. After the preparation, each student was told to say at least one sentence related to their picture, so that eventually the whole class completed the story.

The questions-and-answers activity was a conversation practice task in which two sets of cards were prepared in advance, one set for students and the other for teachers. On each of the teacher's cards was a question about an action in the past (e.g., *Did the concert begin on time?*). Below each question on the teacher's card was a specific noun or phrase (e.g., *6:00pm*). The teacher had 25 cards, each with a number on it. There was a second set of cards (student cards) distributed at the beginning of the activity. Each card had a response (e.g., *6:00pm*) and a number written on it. The teacher read the question and then called a number. The student with that number provided the response. The student completed the sentence quickly using the information on the card by responding *Yes, the concert began at 6:00pm*. If the student made an error in the use of the past tense, feedback was provided (in experimental groups).

Test Materials

Oral Tests

All students were tested individually. Oral tests required students to retell a story based on a series of word cues. Each student was handed a story and was given three minutes to read the story silently. The story was then returned to the examiner. To help the learners remember what the story was about, the examiner provided them with a series of word cues including content words, verbs in their base form, and adverbials that indicated that something happened in the past. The learners were then asked to retell the story using the appropriate forms of the verbs provided as word cues. There was no planning time for learners after the word cues were shown to them. They were told to retell the story with the help of some word prompts. The three stories used were A Beach Party, A Day With My Family, and The Stolen Bicycle. In Appendix S5 in the online Supporting Information, we provide the tests and the lists of regular and irregular verbs. The oral data were collected with each student individually and responses were audio-recorded, transcribed, and then coded for accuracy.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Written Tests

The written production took the form of a written narrative and aimed to test the students' productive knowledge of past tense forms. Participants were provided with a topic and specific verbs to be used (see Appendix S5 in the online Supporting Information) and were asked to compose a story within 20 minutes. The three topics were A Happy Day, The Best Memory From My Holiday, and My Favorite Birthday.

Modifications

In order to be true to the original study, we made as few changes as possible. When we did make changes, there were generally three reasons for doing so: (a) to make materials age-appropriate; (b) to avoid repetition of content; or (c) to make them appropriate for our students' background knowledge (they had been studying English for fewer years than the university students in the YL study).

Given the profile of the participants and the number of years of English study, we felt confident that the participants would be familiar with the 2,000 most frequent words in English. To ensure that the materials would not pose comprehension problems due to unfamiliar vocabulary, we ran the materials through the lexical profiler at http://www.lextutor.ca/vp/ (Cobb, 2019) and found that 97–98% of the words in each story, with the exception of Cinderella,[9] were within the 2,000 most frequent words. We then showed the stories to the teachers, and they confirmed that the students should not have comprehension problems. Thus, we felt confident that the materials were well within these students' reading capacity. Information regarding all the stories is provided in Table 1. All materials were uploaded to the IRIS database.

Coding and Scoring Procedure

There were two types of coding: student responses and teacher feedback. Student data had as its focus grammatical accuracy of regular and irregular past-tense forms. Accuracy in the present study was operationalized as "the correct use of past-tense forms in appropriate past tense context" (Yang, 2008, p. 113).

The oral data were transcribed, and we used the transcriptions and the written stories for coding and scoring. The criteria for coding and scoring both written and oral data followed the original study: (a) the suppliance of the past tense forms in obligatory context; and (b) the accuracy of the past tense forms used. One point was awarded in case of the combination of (a) and (b) (i.e., the use of the correct form of the simple past tense in appropriate context). For example, if the student says *I flew to Greece for the first time in 2018*,

Table 1 Stories used in present study

Title	Words *N*	Verbs		Lexical Coverage at 2,000 words	Source
		Regular *N*	Irregular *n*		
Cinderella	213	16	16	89.3%	http://www.abcteach.com
A Quiet Walk	211	16	15	97.2%	https://cdn.shopify.com/s/files/1/0252/4723/files/Simple-Past-Random-Pages-Sample2.pdf
A Beach Party	210	16	17	98.1%	http://www.eslgo.com/quizzes/irregpast2.html
A Day With My Family	212	16	17	97.6%	https://livvbruce.wordpress.com/2013/11/23/a-sunny-day-at-the-beach-children-short-story/
The Stolen Bicycle	209	15	17	97.6%	https://www.learningprintable.com/wp-content/uploads/2018/10/Third-Grade-Worksheets-Reading.jpg

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

next to the verb *fly* on the scoring sheet, the rater indicated one point because the learner used the correct past-tense form of the verb *fly* (i.e., *flew*) in the appropriate past-tense context. There were many possibilities for the awarding of zero. Examples are listed in Appendix S6 in the online Supporting Information. Only the target verbs were used in coding and scoring. All coding of student data was done by one researcher at each site, with an interrater reliability check carried out by having an additional rater code approximately 25% of the data. Discrepancies were resolved by a third researcher through a discussion with the other two raters.

Following difficulties discussed in Goo and Mackey (2013), we planned to investigate differences across prompt types. Goo and Mackey (2013) made the important "apples and oranges" argument that a comparison of one type of feedback (recasts) with many types (prompts) was misleading. As they have noted, "learners receiving multiple types of feedback have more opportunities to benefit from contextually appropriate feedback than those exposed to only one type of feedback during the entire task" (p. 150). We considered the recordings of teacher feedback and analyzed feedback types in all contexts (recasts and prompts). One of the researchers coded all feedback episodes and identified each as either recast or one of the four prompt types: metalinguistic clue, repetition, clarification request, or elicitation. Feedback episodes were coded as follows: *recast* (1), *metalinguistic* (2), *repetition* (3), *clarification request* (4), *elicitation* (5). Interrater reliability is reported. In Appendix S6 in online Supporting Information, we include a sample coding sheet for both the teacher and the student data.

Analysis

Following YL (2010) and Yang (2008), a mixed design repeated measures ANOVA (analysis of variance) was used in the statistical analysis. The G*Power 3.1.9.4 software (available at www.g-power.hhu.de) was used to calculate the sample size required to expect a power of .95 and a medium effect size Cohen's $f \geq 0.25$, which equals Cohen's $d \geq 0.5$. An a priori power analysis for a three-group repeated measures ANOVA with between-within interactions and a standard probability $\alpha = .05$ returned an estimate of the required total sample size of 54 for the Bosnian site. For the Italian site, where there were four groups, the required total sample size was 60. Repeated measures ANOVA was employed to determine:

1. the differences in various treatment groups' use of simple past-tense forms;

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2. learners' performance on the use of past-tense forms across testing times;
3. the interaction effect between treatment conditions and testing time (i.e., the differences in treatment effects across time on learners' accuracy scores).

Effect sizes were calculated for each effect, using Cohen *d*, for which the sample size, the mean score and standard deviation are needed. In addition, we determined the absolute number of each feedback type and, within the prompt group, the number of each prompt type. We report the number of errors as well as the range of verbs used. These are reported by site.

In Appendix S7 in online Supporting Information, we list modifications due to COVID.

The preregistration of our study is available via OSF at https://osf.io/chzad/.

The final report with appendices and all data are available via OSF at https://osf.io/chzad/.

Materials are available at IRIS database (https://www.iris-database.org/).

Results

The section is divided into two subsections; first, we provide the details of CF treatment using prompts and recasts on the basis of classroom transcripts. Second, we present the results of statistical analyses, organized by site and by mode (oral, written).

Analysis of Classroom Transcripts

We first determined the number of CF episodes provided in each activity. At the Bosnian site, one researcher listened to all recordings, checked the transcripts, and counted the number of CF episodes. Reliability was tested by the second researcher on about 25 % of the data. Interrater reliability agreement was substantial, $\kappa = .745$, Bosnia; $\kappa = .611$, Italy. Discrepancies were resolved in discussion.

Tables 2 and 3 show the distribution of errors, CF episodes, and repair across activities and groups.

At the Italian site, unfortunately, recordings of treatment sessions are available only for the prompt groups due to a recording malfunction of the recast and control groups. In addition, for the prompt groups, only three of the four tasks had useable recordings (the question and answer task did not).

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2 Distribution of errors, feedback episodes, and repair across activities

	Errors		Feedback		Repair	
Activity	Bosnia	Italy	Bosnia	Italy	Bosnia	Italy
Dictogloss 1 Cinderella	28	13	25	25[a]	9	14
Dictogloss 2 A Quiet Walk	45	10	34	13	10	10
Picture narrative	46	6	37	10	8	6
Questions & answers	85	11	110	n.a.[b]	32	10
TOTAL	204	40	206	48	59	40

Notes. [a]In many cases, there were more instances of feedback than the actual number of errors because the teacher would often give feedback to a single error multiple times if the student repeated the error. Students received individual feedback in the classroom, but all students were able to hear the same feedback.
[b]During the questions and answers task in Italy, the recorder stopped working. The teacher took notes during the treatment of the errors that the students made but did not take notes on specific feedback provided, nor on what repair, if any, occurred.

Analysis of Oral and Written Tests

At the Bosnian site, 75 students (with parental consent) agreed to participate in the study. One student missed the oral pretest, and two students missed the written pretest, so their data were removed from the analysis. At the Italian site, 66 students agreed (with parental consent) to participate in the study; two students were eliminated from all analyses because they had one English native speaker parent, and English was used at home.

A professional company transcribed the oral data. One of the researchers listened to 20% of the oral data and verified the accuracy of the transcripts. For the Italian data, accuracy was 97.78%. A third rater listened to the original recordings in cases of discrepancy. Coding of the oral data was done by one of the researchers, and coding of the written data was done by another (for the data collected in Bosnia), and both the oral and the written data by a research assistant (for the data collected in Italy). Twenty percent of the data at each research site were checked by a second researcher. Interrater reliability was almost perfect agreement, oral: $\kappa = .958$, Bosnia; $\kappa = .922$, Italy; written: $\kappa = .945$, Bosnia; $\kappa = .938$, Italy. When discrepancies occurred, a third rater coded the data until 100% agreement was reached.

We ran repeated measures ANOVA on the percentages of correct verb forms in the three groups. To interpret statistical significance, alpha was set at $p = .05$. We calculated effect sizes using Cohen's d; following Plonsky and

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3 Distribution of errors, feedback episodes and repair across groups

		Feedback episodes						
Group	Errors	Clarification requests	Elicitations	Repetitions	Metaling. feedback	Prompts Total	Recasts	Repair
Bosnia Prompt	99[a]	17	61[c]	0	24	102	6	34
Bosnia Recast	105[b]	2	5	0	0	7	91	25
Italy Prompt	40	7	9	17	17	48	n.a.	40

Notes. [a] Out of this number, 24 errors were repetitions of the same error after CF was provided
[b] Out of this number, four errors were repeated after CF was provided
[c] Out of this number, five elicitations were repeated after the repetition of t he error

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 4 Oral tests, descriptive statistics (%) for regular verbs (Bosnia)

		Pretest		Immediate Posttest		Delayed Posttest	
Group	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Prompt	24	23.46	25.52	32.32	25.69	37.67	32.52
Recast	24	23.80	20.69	30.01	20.79	36.56	26.29
Control	22	26.13	24.06	29.66	30.57	37.14	31.84

Table 5 Oral tests, descriptive statistics (%) for irregular verbs (Bosnia)

		Pretest		Immediate Posttest		Delayed Posttest	
Group	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Prompt	24	26.06	27.25	32.71	34.29	36.24	32.37
Recast	24	20.64	20.41	32.81	24.22	42.35	28.09
Control	22	28.52	28.16	36.55	31.98	38.66	36.78

Oswald (2014), we considered an effect of 0.25 to be small, 0.40 medium, and above 0.60 large. For all analyses involving regular and irregular verbs together, descriptive statistics and results can be found in Tables S8.1–S8.4 of Appendix S8 in online Supporting Information. In addition to our originally planned statistical analyses, we also analyzed our data using linear mixed-effects modelling (LMM). LMM results are presented in Appendix S9 in online Supporting Information along with graphs.[10]

Oral Tests

Bosnia

A one-way ANOVA on the pretest found no significant group differences, F (2, 71) = 0.66, p = .519 between the groups. Possible differential effects of CF on regular and irregular verbs were addressed by considering separately the performance of the prompt and the recast group on regular and irregular verbs. See Table 4 for regular and Table 5 for irregular verbs.

A repeated measures ANOVA on regular verbs, with the assumption of sphericity met, found a significant effect of time, F (2, 134) = 18.06, p < .001, no significant effect of group, F (2, 67) = 0.012, p = .988, and no significant Group × Time interaction, F (4, 134) = 0.267, p = .900. Pairwise comparisons[11] for time difference with Bonferroni adjustments showed that the prompt group performed significantly better on the immediate posttest

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

than on the pretest, *M* difference = 8.87, $p = .029$, 95% CI [0.683, 17.052], $d = 0.35$, and on the delayed posttest than the pretest, *M* difference = 14.22, $p = .001$, 95% CI [4.774, 23.658], $d = 0.48$. The recast group had improved significantly from the pretest to the delayed posttest, *M* difference = 12.75, $p = .004$, 95% CI [3.312, 22.196], $d = 0.54$. The control group also had improved significantly from the pretest to the delayed posttest, *M* difference = 11.01, $p = .024$, 95% CI [1.146, 20.869], $d = 0.39$. See Table S9.3 in Appendix S9 in online Supporting Information for LMM results, where significant differences are confirmed only for delayed posttests in both the prompt and the recast groups but not in the control group.

A repeated measures ANOVA on the irregular verbs, with the sphericity assumption met, found a significant effect of time, $F(2, 134) = 18.25$, $p < .001$; no effect of group, $F(2, 67) = 0.78$, $p = .925$, and no Group × Time interaction, $F(4, 134) = 1.40$, $p = .239$.

Pairwise comparisons with Bonferroni adjustments showed that the prompt group had improved significantly from pretest to the delayed posttest, *M* difference = 10.18, $p = .039$, 95% CI [0.399, 19.960], $d = 0.34$; the recast group had improved significantly from the pretest to the immediate posttest, *M* difference = 12.17, $p = .025$, 95% CI [1.171, 23.175], $d = 0.54$, and the delayed posttest, *M* difference = 21.71, $p < .001$, 95% CI [11.933, 31.494], $d = 0.88$, and from the immediate posttest to the delayed posttest, *M* difference = 9.54, $p = .025$, 95% CI [0.942, 18.139], $d = 0.36$. Control group approached statistical significance from the pretest to the delayed posttest, *M* difference = 10.14, $p = .052$, 95% CI [–0.073, 20.358], $d = 0.36$. See Table S9.3 in Appendix S9 in online Supporting Information for LMM results, where the significant differences are confirmed for both the immediate and the delayed posttest in the recast groups but not in the prompt and the control group.

Italy

A one-way ANOVA on the pretest showed no significant differences across groups, $F(2, 68) = 173.71$, $p = .706$. We next considered separately the performance on regular and irregular verbs. The regular verb data are given in Table 6 and irregular verb data are presented in Table 7.

A repeated-measures ANOVA on the regular verb percentages found a significant effect of time, $F(2, 102) = 11.61$, $p = .001$, no significant effect of group, $F(2, 51) = 1.21$, $p = .306$, and a significant Time × Group interaction, $F(4, 102) = 3.52$, $p = .010$. Pairwise comparisons revealed that the recast group had improved significantly from the immediate to the delayed posttest, *M* difference = 14.76, $p = .042$, 95% CI [0.42, 29.11], $d = 0.62$. The control

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 6 Oral tests, descriptive statistics (%) for regular verbs (Italy)

		Pretest		Immediate Posttest		Delayed Posttest	
Group	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Prompt	29	50.24	24.88	55.75	22.36	60.32	28.99
Recast	13	39.32	26.49	40.75	23.55	55.51	24.16
Control	12	35.06	24.77	67.04	18.20	57.50	14.98

Table 7 Oral data, descriptive statistics (%) for irregular verbs (Italy)

		Pretest		Immediate Posttest		Delayed Posttest	
Group	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Prompt	29	43.92	24.97	56.12	26.71	60.00	25.63
Recast	13	34.03	22.47	52.01	30.25	48.36	27.06
Control	12	48.24	17.34	52.45	20.23	61.77	19.64

group had improved significantly from the pretest to the immediate posttest, *M* difference = 31.98, 95% CI [12.65, 51.31], $p = .001$, $d = 1.47$, and the delayed posttest, *M* difference = 22.44, $p = .007$, 95% CI [5.03, 39.85], $d = 1.10$. Pairwise comparisons also revealed that the control group performed significantly better than the recast group on the immediate posttest, *M* difference = 26.29, $p = .012$, 95% CI [47.92, 4.66], $d = 1.25$. See Table S9.3 in Appendix S9 in online Supporting Information for LMM results, where significant differences are confirmed on both the immediate and the delayed posttests for the control group but not for the recast group. Also, there was no significant difference between the recast and the control group.

A repeated-measures ANOVA on the irregular verb percentages found a significant effect of time, F (2, 102) = 15.06, $p = .001$, no significant effect of group, F (2, 51) = 0.78, $p = .463$, and no Time × Group interaction, F (4, 102) = 1.07, $p = .378$. Pairwise comparisons revealed that the prompt group had improved significantly from the pretest to the immediate posttest, *M* difference = 12.20, $p = .008$, 95% CI [1.98, 26.38], $d = 0.47$, and to the delayed posttest, *M* difference = 16.08, $p = .001$, 95% CI [6.82, 25.34], $d = 0.64$. The recast group had improved significantly from the pretest to the immediate posttest, *M* difference = 17.98, $p = .009$, 95% CI [3.75, 32.21], $d = 0.68$, and to the delayed posttest, *M* difference = 14.33, $p = .040$, 95% CI [0.50, 28.16], $d = 0.58$. See Table S9.3 in Appendix S9 in online Supporting

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 8 Written tests, descriptive statistics (%) for regular verbs (Bosnia)

		Pretest		Immediate Posttest		Delayed Posttest	
Group	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Prompt	24	31.67	35.14	40.61	33.99	35.40	36.28
Recast	23	41.14	34.29	39.90	29.20	44.91	31.67
Control	21	31.59	28.30	39.66	36.37	28.66	32.06

Table 9 Written tests, descriptive statistics (%) for irregular verbs (Bosnia)

		Pretest		Immediate Posttest		Delayed Posttest	
Group	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Prompt	24	30.99	33.68	27.71	30.59	29.35	26.75
Recast	23	36.18	27.38	43.74	27.33	41.11	28.66
Control	21	36.97	33.80	40.50	34.46	35.89	30.34

Information for LMM results, where the significant results are confirmed for the prompt group on both the immediate and the delayed posttests, but not for the recast group.

Written Tests

Bosnia

A one-way ANOVA on the pretest showed no significant differences across groups, $F\,(2, 72) = 0.22$, $p = .801$. Table 8 presents the accuracy percentages of regular verbs, and Table 9 presents the accuracy percentages of irregular verbs.

A repeated-measures ANOVA on regular verbs found no significant effect of time, $F\,(2, 130) = 1.57$, $p = .211$, or group, $F\,(2, 65) = 0.50$, $p = .611$, and no significant Group × Time interaction, $F\,(4, 130) = 1.33$, $p = .263$. Table 9 shows the accuracy percentages of irregular verbs.

A repeated-measures ANOVA on irregular verbs found no significant effects of time, $F\,(2, 130) = 0.47$, $p = .626$, or group, $F\,(2, 65) = 1.04$, $p = .361$, and no significant Group × Time interaction, $F\,(4, 130) = 0.80$, $p = .528$. Due to the nonsignificant results of the repeated-measures ANOVA, pairwise comparisons were not carried out.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 10 Written tests, descriptive statistics (%) for regular verbs (Italy)

		Pretest		Immediate Posttest		Delayed Posttest	
Group	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Prompt	29	49.51	25.56	61.71	27.62	56.85	31.17
Recast	10	52.19	32.30	53.19	33.62	48.29	30.48
Control	11	52.21	28.19	78.53	19.25	64.94	29.58

Table 11 Written tests, descriptive statistics (%) for irregular verb (Italy)

		Pretest		Immediate Posttest		Delayed Posttest	
Group	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Prompt	29	49.18	27.42	61.27	26.14	53.92	29.53
Recast	10	48.81	34.38	47.66	39.63	50.38	28.36
Control	11	61.27	24.84	77.38	21.22	69.77	25.50

Italy

Three participants' data were removed from the analyses because they had 100% accurate verb usage on the written pretest (two participants in the prompt group and one in the recast group). A one-way ANOVA on the pretest showed no significant differences across groups, $F\,(2, 65) = 0.713$, $p = .494$.

We analyzed the groups' performance in terms of regular and irregular verbs. The descriptive statistics of regular verbs can be seen in Table 10, and the descriptive statistics for irregular verbs can be found in Table 11.

A repeated-measures ANOVA on regular verbs found a significant effect of time, $F\,(2, 94) = 4.12$, $p = .019$, no significant effect of group, $F\,(2, 47) = 1.04$, $p = .363$, and no significant Time × Group interaction, $F\,(4, 94) = 1.10$, $p = .363$. Pairwise comparisons revealed that the control group had improved significantly from the pretest to the immediate posttest (*M* difference = 26.32, $p = .020$, 95% CI [3.30, 49.34], $d = 1.09$).

A repeated-measures ANOVA on irregular verbs found no significant effects of time, $F\,(2, 94) = 2.33$, $p = .137$, or of group, $F\,(2, 47) = 2.32$, $p = .110$, and no significant Time × Group interaction, $F\,(4, 94) = 0.59$, $p = .669$. See Table S9.5 in Appendix S9 in online Supporting Information for LMM results, which do not differ from the results obtained by repeated measures ANOVA using SPSS software.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Discussion

Research Question 1: Do the Groups That Had CF Outperform the Control Group Where No CF Was Provided?

In our study, we did not find an effect of group in either the written or the oral data at either of our testing sites. However, when regular and irregular verbs were considered separately, a somewhat different picture emerged.

We found no group effect in Bosnia; in Italy, the control group outperformed the recast group on the immediate posttest but only for regular verbs (only oral data). We note that in the analyses using mixed-effects modeling, this difference was not significant. The improvement patterns showed a steady increase for the control group in Italy from pre to post to delayed posttest. Both CF groups in Italy showed significant improvement by the time of the delayed posttest, and the same pattern was shown in the prompt and the control groups in Bosnia, whereas the recast group in Bosnia significantly improved from pretest to both posttests, as well as from the immediate posttest to the delayed posttest.

The strong performance of the control group,[12] especially in Italy, is surprising but not that different from the results of the YL study, a topic we turn to below when we compare our results with theirs. One reason for this strong performance may have to do with proficiency, which was higher in Italy than in Bosnia, as can be seen from a comparison of pretest scores for all three groups at both sites. Proficiency levels were not a direct concern of this study. However, when the pretest results from the different sites are considered, it is clear that the students in Italy scored higher than those in Bosnia. A posthoc *t*-test on pretest accuracy scores showed a significant difference for both the written and oral data, written: $t = 6.50, p = .003$; oral: $t = 5.20, p = .007$. This was supported by mixed-effects modeling where the difference between students in Italy and Bosnia was significant ($p < .001$) in all the tests and for both verb types. Further, students in Italy were in a linguistic high school (Liceo Linguistico) and had language learning as a curricular focus. The pretest data are limited to past tense verbs, but it is likely that the overall proficiency of the Italian students was higher than the proficiency level of students in Bosnia. Thus, it may be that because of the higher proficiency of the Italian students, they needed less correction, particularly in instances where rule-based learning was involved. This suggests that at higher levels of proficiency, input alone is sufficient, which helps to explain the high performance of the control group in Italy.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Research Question 2: Do Recasts Have a Differential Effect on the Acquisition of Regular Versus Irregular Past Tense Verb Forms? Do Prompts Have a Differential Effect on the Acquisition of Regular Versus Irregular Past Tense Verb Forms?

We had anticipated that our data would shed light on whether feedback differentially affects rule-based (regular verbs) versus item-based (irregular verbs) forms. Oral data provides some evidence that this is the case, but proficiency may enter the picture. In Italy, where students had greater knowledge of past tense verbs, recasts yielded greater improvement on irregular as opposed to regular verbs. In Bosnia, recasts resulted in significant improvement on both verb types at the time of the delayed posttest.

Conclusions based on the oral data from the prompt groups are less straightforward than they are for the recast groups, although some interesting and consistent findings did emerge. For the Italian students, there was significant improvement only on irregular verbs. For Bosnia, feedback resulted in improvement on both verb types. Thus, in general, feedback was beneficial for the higher proficiency students only for irregular verbs, whereas feedback was beneficial for regular verbs for the lower proficiency students.

A usage-based approach claims that frequency of input can help to explain differences in gains on the basis of feedback to errors in irregular versus regular verb usage. To address this issue, we considered the errors that received feedback in each of our experimental groups; that is, were the errors on regular or irregular verbs? In both feedback types (recasts and prompts), there was more feedback during treatment sessions on irregular verbs than on regular verbs. The most complete information comes from Bosnia, where the ratio of corrective feedback to irregular and regular verbs in the recast group was 47 to 35 and in the prompt group 43 to 31. Less complete but similar information comes from the Italian prompt group data, where the ratio was 45 to 18. Based on Italian data alone, the answer is clear: Where there is greater frequency in the input, there is also greater learning. We argue that our results support approaches that rely on frequency of input to account for improvement as opposed to processing differences. However, the Bosnian data paint a different picture because, regardless of feedback type (recasts or prompts), significant improvement could be seen in both verb types (although improvement was often seen from delayed posttest data rather than immediate posttest data). It appears that frequency of input alone may not be sufficient at early stages of learning. Rather, feedback type (recasts for irregular verbs and prompts for regular verbs) also appears to play a role. We turn to this topic when considering our third research question.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Research Question 3: Does the Prompt Group Outperform the Recast Group for Regular Verbs? Does the Prompt Group Outperform the Recast Group for Irregular Verbs?

We did not find evidence of the superiority of either feedback type. For regular verbs, neither CF type yielded significant gains. However, in Bosnia, an interesting outcome was noted: for regular verbs, prompts yielded improvement on delayed posttests, and, similarly, recasts were more effective on delayed posttests. For irregular verbs, consistent gains came in the recast group, with the prompt group only seeing significant improvement at the time of the delayed posttest. We suggest that for learning irregular verbs any feedback is useful for learners who obtained higher scores on pretest (Italy). In Bosnia, prompts appear to be more effective or equal to recasts in the case of regular verbs, whereas recasts seem to be superior in the case of irregular verbs. Thus, the extra processing involved in the case of prompts may be useful when dealing with rule-based learning. But with irregular verbs recasts are sufficient for learning where little processing is needed to learn a single item. The extra salience required for regular verbs is not necessary for irregular verbs. This study then suggests that adolescent data do not pattern in the same way as adult data, but it also suggests that proficiency plays a role in this understanding.

Interaction of Feedback Types and Verb Types

We next turn to a discussion of feedback types and verb types. Our participants were foreign language learners with limited L2 exposure, who mostly rely on explicit knowledge. With both verb types, they needed support to produce the forms for which they had explicit knowledge (a pedagogical rule [regular verbs] or a specific verb form [irregular verbs]). In Italy, fewer errors were made and less feedback was received on regular verbs than on irregular verbs, as in the case of irregular verbs, a variety of forms had to be remembered and used when narrating a story. Both types of CF were more effective with irregular verbs, though recasts did result in significant improvement on regular verb delayed posttests. We suggest that these students with greater knowledge of the past tense (see pretest scores) managed to deploy the pedagogical rule for past tense with less need of explicit feedback than for irregular forms.

In Bosnia, where the students' knowledge of past tense prior to the treatment was lower than in Italy, both prompts and recasts seem to have achieved their purpose in a complementary mode: Prompts resulted in significant improvements in the use of regular verbs, whereas recasts were superior in the case of irregular verbs. In both situations, as evidenced, improvements could be observed over time, with strongest performance recorded on delayed posttests.

As noted above, these results are consistent with the usage-based approach because more errors occurred and more feedback was provided on irregular than regular verbs, which resulted in greater effects of feedback on irregular verbs.

Comparison With Yang & Lyster's (2010) study

This study set out to replicate YL (2010) with a population of adolescents. Table 12 makes a direct comparison of the statistical results for the three sites (Bosnia, Italy, China).

As can be seen from the table, the findings from our study did not fully align with the YL results. Table 13 presents a comparison of results in our multisite study with the results of YL.

One surprising finding in our data was the performance of the control group that in many cases performed at the level of the experimental groups or even better (in Italy). This also appeared to be the case in YL's study, as in some instances (e.g., irregular verbs), the control group showed improvement with effect sizes that were comparable to those of the experimental groups. One explanation for the improved performance of control groups at both sites may be related to the tasks used, which were more appropriate for the experimental groups and less so for the control group. The tasks used were designed to elicit numerous instances of past tense verb forms, both regular and irregular, so that there would be a sufficient number of opportunities for the provision of CF. This means that the control group also had significant exposure to the verb forms, which could explain why the control group showed as much improvement as they did.

Another difference between our results and YL's results involves the effectiveness of prompts for regular versus irregular verbs. YL found that prompts affected both regular and irregular verbs. In our study, there was a differential effect, but the evidence points in different directions at the different sites: In Bosnia, the greater effect was in the case of regular verbs, and in Italy, the effect was higher in the case of irregular verbs. We, therefore, note the progression of the effectiveness of prompts and increased proficiency from regular verbs to irregular verbs to an equal effect. Perhaps at lower proficiency levels, as was presumably the case with the Bosnian students, prompts might be more effective for regular verbs. Both regular and irregular verb knowledge started low (23.46% in the case of regular verbs and 26.06% in the case of irregular verbs) with significant room for improvement. With time (i.e., at the time of the delayed posttests), the Bosnian students in the prompt group were seen to have improved on irregular verbs as well. In Italy, where the students had higher

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 12 Significant within-group time contrasts, probability levels (Bosnia and Italy only), and magnitude of effect sizes

	Prompt Group			Recast Group			Control Group		
Test	Bosnia	Italy	YL	Bosnia	Italy	YL	Bosnia	Italy	YL
Oral Regular									
Immediate	$p = .029$	ns		ns	ns	ns	ns	$p = .001$	ns
	$d = 0.35$		$d = 0.57$					$d = 1.47$	
Delayed	$p = .001$	ns		$p = .004$	$p = .042$	ns	$p = .024$	$p = .007$	ns
	$d = 0.48$		$d = 0.69$	$d = 0.54$	$d = 0.62$		$d = 0.39$	$d = 1.10$	
Oral Irregular									
Immediate	ns	$p = .008$		$p = .025$	$p = .009$		ns	ns	
		$d = 0.47$	$d = 0.94$	$d = 0.54$	$d = 0.68$	$d = 0.81$			$d = 0.78$
Delayed	$p = .039$	$p = .001$		$p = .001$	$p = .040$	ns	$p = .052$	ns	
	$d = 0.34$	$d = 0.64$	$d = 0.66$	$d = 0.88$	$d = 0.58$		$d = 0.36$		$d = 0.46$
Written Regular									
Immediate	ns	ns	$d = 1.92$	ns	ns	$d = 1.11$	ns	$p = .020$	$d = 0.82$
Delayed	ns	ns	$d = 0.98$	ns	ns	ns	ns	$d = 1.09$	ns
Written Irregular									
Immediate	ns	ns	$d = 1.51$	ns	ns	$d = 1.92$	ns	ns	ns
Delayed	ns	ns	$d = 1.65$	ns	ns	$d = 0.98$	ns	ns	ns

Note: YL = Yang and Lyster (2010); ns = nonsignificant results.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 13 Comparison of Bosnia/Italy data with data from Yang and Lyster (2010)

RQ	Bosnia/Italy	Yang & Lyster
1	Do the groups that had CF outperform the control group where no CF was provided?	Do the groups that performed form-focused production activities while receiving CF show an overall superiority in learning regular and irregular past tense over the control group, which performed the same communicative classroom activities but without receiving CF on past tense errors?
	No. No group effect for either written or oral production data. For oral data, all three groups improved over a longer time; the recast group significantly improved from pretest to both posttests at both sites, but only for irregular verbs. Prompt and control groups at both sites showed more improvement on regular verbs.	**Yes,** although the results of YL did not address overall superiority. YL showed that the prompt group outperformed the control group in each verb type (using effect sizes as a means of comparison). The results were less clear for recasts.
	For written data, only the control group in Italy showed significant improvement from pretest to posttest.	
2a	Do **recasts** have a differential effect on the acquisition of regular versus irregular past tense verb forms?	Do **recasts** have differential effects on the acquisition of regular versus irregular English past tense forms?
	Partially. Recasts had a differential effect, greater on irregular verbs, in both Bosnia and Italy, but only in oral production.	**Partially**. For oral data, based on effect sizes, there was only a large effect for irregular verbs on the immediate posttest.
	For written data, there were no differences.	For written data, there were large effect sizes for both irregular and regular verbs on immediate posttests, but only for irregular verbs on the delayed posttest.

(*Continued*)

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 13 (Continued)

RQ	Bosnia/Italy	Yang & Lyster
2b	Do **prompts** have a differential effect on the acquisition of regular versus irregular past tense verb forms? **Partially yes, but in different directions**. There was improved performance in Italy on irregular verbs but not on regular verbs. In Bosnia, there was generally better performance on regular verbs than on irregular verbs. No differences were observed on written data.	Do **prompts** have differential effects on the acquisition of regular versus irregular English past tense forms? **Mostly no**. For oral data, there were either large or medium effect sizes for both irregular and regular verbs on both immediate and delayed posttests. For written data, there were mostly large effects on all tests and for both verb types.
3a	Does the prompt group outperform the recast group for regular verbs? **No**. In Bosnia, greater short-term improvement in prompt than in recast group (but larger effect size for recast group). No improvement in written data at either site.	No comparable question asked, although in the discussion, YL note the superiority of the prompt group over the recast group but only for regular verbs; for irregular verbs, both feedback types were equal in effectiveness. Difficult to assess. YL presented their results separately for irregular and regular verbs. Another complication in interpreting YL results was that the control group had lower scores than the other groups (oral data). It appears that the answer to this research question is yes for prompts but not necessarily for recasts.
3b	Does the prompt group outperform the recast group for irregular verbs? **No**. Generally greater improvement for the recast group.	

Note: RQ = research question; CF = corrective feedback; YL = Yang and Lyster (2010).

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

accuracy rates for past tense verb forms on pretests and where the pretest results were higher than those in Bosnia (50.24% for regular verbs and 43.92% for irregular verbs), irregular verbs posed a greater challenge, and the students needed an additional feedback type. For adults (the YL study), prompts were equally effective for both verb types. So, the question is: Why are prompts equally effective for regular and irregular verb types for adults but not for adolescents? One could argue that this is a matter of cognitive maturity. If that were the case, we would expect that the results would be the same for Bosnian and Italian students since they were of the same age and presumably at relatively similar levels of cognitive maturity. However, the results for our two groups are not uniform, with prompts being more effective for regular verbs in the Bosnian group and more effective for irregular verbs in the Italian group. We suggest that this is a matter of proficiency (or, minimally, greater knowledge of past tense verb forms) rather than cognitive maturity: The Bosnian and Italian students were at an equal level of cognitive maturity, but most likely not of proficiency. However, whether this is an issue of proficiency or cognitive maturity remains to be seen in additional research.

Limitations and Future Directions

There are numerous limitations of this study, some of which concern the actual data, some of which relate to the tasks, and some of which relate to differences in execution between our study and that of YL. First, the sample size in Italy was lower than we had anticipated, particularly for the control and recast groups. Second, even though the teachers in Italy went through a rigorous training session, they were still not trained researchers; on the other hand, in Bosnia, given an unfortunate turn of events, the authors of this study also provided instruction to both experimental groups.

Following YL, we coded only the verbs that we had asked students to use. However, some students used other verbs correctly, particularly on the written tests, and therefore our data may not represent the true knowledge of our participants.

We noted earlier the pretest accuracy differences between the students at the two sites, but the students were different in other ways as well. Italian students were in a linguistic high school (Liceo Linguistico) and were therefore focused on and interested in language learning and possibly had high degrees of language aptitude. On the other hand, the Bosnian students were from a vocational school, trained in hospitality and tourism industry.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Finally, we noted earlier that the control groups as well as the experimental groups were exposed to a large number of past tense verb forms due to the fact that the control groups performed the same tasks as the experimental groups.

There are two differences between our methodology and that of YL that might have affected our ability to make meaningful comparisons. First, YL had two versions of tests, so students had the same pretest and delayed posttest. In our study, each student (for both oral and written tests) had three different test versions. Second, in YL's dictogloss task, the teachers in all groups "provided the original texts with the past tense verbs highlighted in bold for students to compare with their own texts" (p. 245). What this essentially means is that YL's control group did receive feedback as part of the treatment sessions.

Conclusion

Motivated by the need for greater inclusivity in data sources for SLA, our study was an attempt to expand current practice by examining behaviours of high school students, who, in a sense, straddle the line between child and adult learners. Working with this population can be both challenging and rewarding. Conducting research at a high school level brings its specific challenges connected with classroom teaching and learning. The need to adhere to the school/class schedule and not to interrupt the established process of teaching proved to be a constraint that affected our data collection to some extent (for example, the class sizes or the order of scheduled classes). However, the teachers' and students' enthusiasm and motivation to participate in the study were the driving forces helping us to move forward with data collection until the last tests were completed. The welcoming atmosphere at both sites made up for all the challenges that research with this population may pose. At both sites, the data collection was long delayed due to the school closures during the COVID pandemic. Even at the time of our data collection in spring of 2022, there was a constant threat of a new closure. Therefore, when the last (delayed) posttests were completed, we finally felt relieved as we knew the most precarious part of our study was over. The data that our adolescent participants at both sites contributed to this replication study have immense value: Even though more research is needed to come at final conclusions, the data that we collected show that findings obtained in studies of educated adults are not always generalizable and applicable to other, less educated or less cognitively mature populations.

This study shows the benefits and the challenges of conducting multisite research. On the one hand, replication is necessary and significant for the field; on the other hand, the population differences may have been such that a

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

common conclusion for both sites could not always be drawn. Yet, despite this, we were heartened at the similarity between the sites, which yielded greater confidence in our findings.

Final revised version accepted 29 August 2023

Open Research Badges

This article has earned Open Data and Open Materials badges for making all data and materials used in the study publicly available. The article has also earned a Preregistered Research Designs badge for having a preregistered research design. All data, materials and design are available at https://osf.io/chzad/ and https://www.iris-database.org/.

Notes

1 The current study is a partial replication, given that some of the materials were modified to be more age-appropriate for the adolescents who participated in this study.
2 We recognize that there are different ways of characterizing cognitive growth (e.g., is it continuous or can it be better thought of as discontinuous stages?) This discussion goes beyond the scope of this paper.
3 The effects of the two feedback types were not investigated, as the primary focus was on form-focused instruction.
4 We do not wish to imply that the high school population that we focus on in this study has been ignored in all SLA research. For example, in the meta-analysis by Li (2016) on aptitude, there are almost as many high school samples as university-based samples. What we do find, however, is that biased sampling is prevalent in interaction-based research.
5 It is important to point out that in Chinese there is no morphological marking of past tense; this is different for the first languages of the participants in the present study.
6 Repetition or uptake is only possible after explicit correction or after a recast but not after prompts.
7 At the Bosnian site, both teachers had a degree in English language and literature and over 10 years of teaching experience. The principal of the school in Italy was not comfortable providing detailed information about the teachers.
8 Because in Italy we had four classes, we had two prompt groups, so that we could conduct posthoc analyses and possibly tease apart the role of specific types of prompts by comparing the two prompt groups to determine if outcomes may be related to feedback type.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

9 The coverage in Cinderella was 89%, but the words that were outside the 2,000 most frequent vocabulary items (e.g., *stepmother*, *pumpkin*, *fairy*) were words that all teachers felt that the students were familiar with.

10 Because this was a replication study, our analysis was faithful to YL's original analysis. However, based on reviewers' and editors' suggestions, we included the linear mixed-effects modeling results in Appendix S9 in online Supporting Information.

11 We present only significant pairwise comparisons. For all significant and nonsignificant pairwise comparisons, see Appendices S10 and S11 via the OSF link https://osf.io/chzad/.

12 Based on linear mixed-effect modelling, there is one outlier, namely a very large improvement by the Italian control on regular verbs from pre to posttest. The effect does not quite reach statistical significance due to adjustment for multiple comparisons ($p = .07$), but the numerical difference is large. We do note that, in actuality, we had a comparison group rather than a true control group. Had we truly had a control group without the same exposure to the same verbs, it is likely that the corrective feedback groups would have outperformed the control group at all times and with both verb types.

References

Alcón, E., & García Mayo, M. (2008). Incidental focus on form and learning outcomes with young foreign language classroom learners. In J. Philp, R. Oliver, & A. Mackey (Eds.), *Second language acquisition and the younger learner: Child's play?* (pp. 173–192). Multilingual Matters. https://doi.org/10.1075/lllt.23.12sol

Ammar, A., & Spada, N. (2006). One size fits all?: Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, *28*(4), 543–574. https://doi.org/10.1017/S0272263106060268

Andringa, S., & Godfroid, A. (2020). On the foundations of knowledge in applied linguistic research: Sampling bias and the problem of generalizability. *Annual Review of Applied Linguistics*, *40*, 134–142. https://doi.org/10.1017/s0267190520000033

Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, *63*(7), 602–614. https://doi.org/10.1037/0003-066x.63.7.602

Berk, L. (2006). *Child development* (7th ed.). Pearson Education.

Berman, R. (2007). Language knowledge and language use across adolescence. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 347–367). Blackwell. https://doi.org/10.1002/9780470757833.ch17

Bigelow, M., Delmas, R., Hansen, K., & Tarone, E. (2006). Literacy and the processing of oral recasts in SLA. *TESOL Quarterly*, *40*(4), 665–668. https://doi.org/10.2307/40264303

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Braidi, S. M. (2002). Reexamining the role of recasts in native-speaker/nonnative-speaker interactions. *Language Learning*, *52*(1), 1–42. https://doi.org/10.1111/1467-9922.00176

Brown, D. (2016). The type and linguistic foci of oral corrective feedback in the L2 classroom: A meta-analysis. *Language Teaching Research*, *20*, 436–458. https://doi.org/10.1177/1362168814563200

Bryfonski, L., & Sanz, C. (2018). Opportunities for corrective feedback during study abroad: A mixed methods approach. *Annual Review of Applied Linguistics*, *38*, 1–32. https://doi.org/10.1017/s0267190518000016

Carpenter, H., Jeon, S., MacGregor, D., & Mackey, A. (2006). Learners' interpretations of recasts. *Studies in Second Language Acquisition*, *28*, 209–236. https://doi.org/10.1017/s0272263106060104

Chen, T. (2016). Technology-supported peer feedback in ESL/EFL writing classes: A research synthesis. *Computer Assisted Language Learning*, *29*, 365–397. https://doi.org/10.1080/09588221.2014.960942

Cobb, T. (2019, September 15). *Vocabprofile*. Lextutor. https://www.lextutor.ca/

Doski, P. M., & Cele, F. (2018). The effect of oral corrective feedback on article errors in L3 English: Prompts vs. recasts. *English Language Teaching*, *11*(8), 143–158. https://doi.org/10.5539/elt.v11n8p143

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition*. Cambridge University Press.

Ellis, N., & Wulff, S. (2020). Usage-based approaches to L2 acquisition. In B. VanPatten, G. Keating, & S. Wulff (Eds.), *Theories in second language acquisition* (3rd ed., pp. 63–82). Routledge. https://doi.org/10.4324/9780429503986-4

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. https://doi.org/10.21832/9781847691767-015

Ellis, R., & Sheen, Y. (2006). Reexamining the role of recasts in L2 acquisition. *Studies in Second Language Acquisition*, *28*(4), 575–600. https://doi.org/10.1017/s027226310606027x

García Mayo, M., & Labandibar, U. (2017). The use of models as written corrective feedback in English as a foreign language (EFL) writing. *Annual Review of Applied Linguistics*, *37*, 110–127. https://doi.org/10.1017/S0267190517000071

Gass, S., Behney, J., & Uzum, B. (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: Studies in honor of Waldemar Marton* (pp. 91–114). Springer. https://doi.org/10.1007/978-3-642-23547-4_6

Gass, S., & Mackey, A. (2020). Input, interaction, and output in L2 acquisition. In B. VanPatten, G. Keating, & S. Wulff (Eds.), *Theories in second language acquisition:*

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

An introduction (3rd ed., pp. 192–222). Routledge. https://doi.org/10.4324/9780429503986-9

Goo, J., & Mackey, A. (2013). The case against the case against recasts. *Studies in Second Language Acquisition*, *35*(1), 127–165. https://doi.org/10.1017/s0272263112000708

Havranek, G. (2002). When is corrective feedback most likely to succeed? *International Journal of Educational Research*, *37*, 255–270. https://doi.org/10.1016/s0883-0355(03)00004-1

Havranek, G., & Cesnik, H. (2001). Factors affecting the success of corrective feedback. In S. Foster-Cohen, & A. Nizegorodcew (Eds.), *EUROSLA yearbook*, (Vol. 1, pp. 99–122). John Benjamins. https://doi.org/10.1075/eurosla.1.10hav

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–135. https://doi.org/10.1017/S0140525X0999152X

Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *Modern Language Journal*, *99*, 1–18. https://doi.org/10.1111/modl.12189

Kim, S., & Webb, S. (2019, September 20). *The effect of spaced practice on second language learning: A meta-analytic review*. [Paper presentation]. Second Language Research Forum, East Lansing, MI, United States.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, *11*(02), 261–271. https://doi.org/10.1017/S1366728908003416

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*(2), 309–365. https://doi.org/10.1111/j.1467-9922.2010.00561.x

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Applied Linguistics*, *36*(3), 345–366. https://doi.org/10.1093/applin/amu040

Li, S., Ellis, R., & Zhu, Y. (2016). Task-based versus task-supported language instruction: An experimental study. *Annual Review of Applied Linguistics*, *36*, 205–229. https://doi.org/10.1017/S0267190515000069

Loewen, S. (2012). The role of feedback. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 24–40). Routledge. https://doi.org/10.4324/9780203808184.ch2

Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching*, *51*, 285–329. https://doi.org/10.1017/s0261444818000125

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

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, *26*(3), 399–432. https://doi.org/10.1017/S0272263104263021

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Lyster, R., & Izquierdo, J. (2009). Prompts versus recasts in dyadic interaction. *Language Learning*, *59*(2), 453–498. https://doi.org/10.1111/j.1467-9922.2009.00512.x

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

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, *46*(1), 1–40. https://doi.org/10.1017/s0261444812000365

Mackey, A. (Ed.). (2012). *Input, interaction and corrective feedback in L2 learning*. Oxford University Press.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, *22*(4), 471–497. https://doi.org/10.1017/s0272263100004010

Mackey, A., & Oliver, R. (2002). Interactional feedback and children's L2 development. *System*, *30*(4), 459–477. https://doi.org/10.1016/S0346-251X(02)00049-0

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

Mackey, A., & Sachs, R. (2012). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning*, *62*(3), 704–740. https://doi.org/10.1111/j.1467-9922.2011.00649.x

McDonough, K. (2007). Interactional feedback and the emergence of simple past activity verbs in L2 English. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 323–338). Oxford University Press.

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

Mifka-Profozic, N. (2014). Effectiveness of implicit negative feedback in foreign language classroom: The role of input, frequency, and saliency. In L. Roberts, I. Vedder & J. Hulstijn (Eds.), *EUROSLA yearbook* (Vol. 14, pp. 111–142). John Benjamins. https://doi.org/10.1075/eurosla.14.05mif

Mifka-Profozic, N. (2015). Effects of corrective feedback on L2 acquisition of tense-aspect verbal morphology. *Language, Interaction and Acquisition*, *6*(1), 149–180. https://doi.org/10.1075/lia.6.1.05mif

Muñoz, C. (2007). Age-related differences and second language learning practice. In R. DeKeyser (Ed.), *Practice in a second language* (pp. 229–255). Cambridge University Press. https://doi.org/10.1017/cbo9780511667275.014

Nassaji, H. (2009). Effects of recasts and elicitations in dyadic interaction and the role of feedback explicitness. *Language Learning*, *59*(2), 411–452. https://doi.org/10.1111/j.1467-9922.2009.00511.x

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Nassaji, H. (2016). Anniversary article: Interactional feedback in second language teaching and learning: A synthesis and analysis of current research. *Language Teaching Research*, *20*, 535–562. https://doi.org/10.1177/1362168816644940

Nassaji, H., & Kartchava, E. (2017). *Corrective feedback in second language teaching and learning*. Routledge.

Nassaji, H., & Kartchava, E. (2021). *The Cambridge handbook of corrective feedback in language learning and teaching*. Cambridge University Press.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*(1), 11–28. https://doi.org/10.1207/s15516709cog1401_2

Newport, E. (1991). Contrasting conceptions of the critical period for language. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind* (pp. 111–130). Lawrence Erlbaum Associates.

Nicholas, H., & Lightbown, P. (2008). Defining child second language acquisition, defining roles for L2 instruction. In J. Philp, R. Oliver, & A. Mackey (Eds.), *Second language acquisition and the younger learner: Child's play?* (pp. 27–52). Multilingual Matters. https://doi.org/10.1075/lllt.23.04nic

Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, *51*(4), 719–758. https://doi.org/10.1111/0023-8333.00172

Oliver, R. (1995). Negative feedback in child NS-NNS conversation. *Studies in Second Language Acquisition*, *17*(4), 459–481. https://doi.org/10.1017/s0272263100014418

Oliver, R. (1998). Negotiation of meaning in child interactions. *Modern Language Journal*, *82*, 372–386. https://doi.org/10.1111/j.1540-4781.1998.tb01215.x

Oliver, R. (2000). Age differences in negotiation and feedback in classroom and pairwork. *Language Learning*, *50*(1), 119–151. https://doi.org/10.1111/0023-8333.00113

Oliver, R. (2002). The patterns of negotiation of meaning in child interactions. *Modern Language Journal*, *86*, 97–111. https://doi.org/10.1111/1540-4781.00138

Oliver, R. (2009). How young is too young? Investigating negotiation of meaning and corrective feedback in children aged five to seven years. In A. Mackey & C. Polio (Eds.), *Multiple perspectives on interaction: Second language research in honor of Susan M. Gass* (pp. 135–156). Routledge. https://doi.org/10.4324/9780203880852

Oliver, R., & Mackey, A. (2003). Interactional context and feedback in child ESL classrooms. *Modern Language Journal*, *87*, 519–533. https://doi.org/10.1111/1540-4781.00205

Oliver, R., Philp, J., & Mackey, A. (2008). The impact of teacher input, guidance and feedback on ESL children's task-based interactions. In J. Philp, R. Oliver, & A. Mackey (Eds.), *Second language acquisition and the younger learner: Child's play?* (pp. 131–147). Multilingual Matters. https://doi.org/10.1075/lllt.23.09oli

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Philp, J. (2003). Constraints on "Noticing the gap": Nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition*, *25*(1), 99–126. https://doi.org/10.1017/s0272263103000044

Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Science*, *6*, 456–463. https://doi.org/10.1016/s1364-6613(02)01990-3

Plonsky, L. (2015, October 29). *Demographics in SLA: A systematic review of sampling practices in L2 research*. [Paper presentation]. Second Language Research Forum (SLRF), Atlanta, GA, United States.

Plonsky, L., & Brown, D. (2015). Domain definition and search techniques in meta-analyses of L2 research (Or why 18 meta-analyses of feedback have different results). *Second Language Research*, *31*, 267–278. https://doi.org/10.1177/0267658314536436

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

Rahimi, M., & Zhang, L. J. (2016). The role of incidental unfocused prompts and recasts in improving English as a foreign language learners' accuracy. *The Language Learning Journal*, *44*(2), 257–268. https://doi.org/10.1080/09571736.2013.858368

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for second language acquisition: A meta-analysis of the research. In J. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 131–164). John Benjamins. https://doi.org/10.1075/lllt.13.09val

Saito, K. (2015). Variables affecting the effects of recasts on L2 pronunciation development. *Language Teaching Research*, *19*, 276–300. https://doi.org/10.1177/1362168814541753

Sato, M., & Loewen, S. (2018). Metacognitive instruction enhances the effectiveness of corrective feedback: Variable effects of feedback types and linguistic targets. *Language Learning*, *68*(2), 507–545. https://doi.org/10.1111/lang.12283

Saxton, M. (1997). The contrast theory of negative input. *Journal of Child Language*, *24*, 139–161. https://doi.org/10.1017/s030500099600298x

Saxton, M. (2000). Negative evidence and negative feedback: Immediate effects on the grammaticality of child speech. *First Language*, *20*, 221–252. https://doi.org/10.1177/014272370002006001

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press. https://doi.org/10.1017/cbo9781139524780.003

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press. https://doi.org/10.5070/L4111005027

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House Publishers.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom: Second language acquisition* (pp. 64–81). Cambridge University Press.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410612700-38

Tarone, E. (2010). Second language acquisition by low-literate learners: An under-studied population. *Language Teaching*, *43*(1), 75–83. https://doi.org/10.1017/S0261444809005734

Tarone, E., & Bigelow, M. (2007). Alphabetic print literacy and oral language processing in SLA. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 101–122). Oxford University Press.

Tsang, W. (2004). Feedback and uptake in teacher-student interaction: An analysis of 18 English lessons in Hong Kong secondary classrooms. *Regional Language Centre Journal*, *35*, 187–209. https://doi.org/10.1177/003368820403500207

Van De Guchte, M., Braaksma, M., Rijlaarsdam, G., & Bimmel, P. (2015). Learning new grammatical structures in task-based language learning: The effects of recasts and prompts. *The Modern Language Journal*, *99*(2), 246–262. https://doi.org/10.1111/modl.12211

Whittle, A., & Lyster, R. (2016). Focus on Italian verbal morphology in multilingual classes. *Language Learning*, *66*(1), 31–59. https://doi.org/10.1111/lang.12131

Yang, Y. (2008). Corrective feedback and Chinese learners' acquisition of English past tense. [Unpublished doctoral dissertation] McGill University.

Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition*, *32*(2), 235–263. https://doi.org/10.1017/s0272263109990519

Yousefi, M., & Nassaji, H. (2019). A meta-analysis of the effects of instruction and corrective feedback on L2 pragmatics and the role of moderator variables. *International Journal of Applied Linguistics*. https://doi.org/10.1075/itl.19012.you

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Accessible Summary

Appendix S1. Modifications Made to Original Treatment and Testing Materials.

Appendix S2. Teacher Roles.

Appendix S3. Background Information.

Appendix S4. Treatment Tasks.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendix S5. Tests.
Appendix S6. Coding.
Appendix S7. Modifications Due to COVID.
Appendix S8. Descriptive statistics: Goups (Total Verbs).
Appendix S9. Linear Mixed-Effects Modelling.

14679922, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/lang.12623 by Test, Wiley Online Library on [22/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License