



OPEN ACCESS

EDITED BY

Zhiru Sun,
University of Southern Denmark, Denmark

REVIEWED BY

Lucian Lupu-Dima,
Universitatea Din Petrosani, Romania
Achmad Yudi Wahyudin,
Universitas Teknokrat Indonesia, Indonesia

*CORRESPONDENCE

Xiwen Lu
✉ xiwenlu@brandeis.edu

RECEIVED 27 June 2025

ACCEPTED 03 October 2025

PUBLISHED 22 October 2025

CITATION

Lu X, Yang Q, Wei J and Heffernan NT (2025)
From belief to evidence: simpler immediate
feedback improves language learning and
confidence in semi-open-ended questions.
Front. Educ. 10:1654809.
doi: 10.3389/feduc.2025.1654809

COPYRIGHT

© 2025 Lu, Yang, Wei and Heffernan. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

From belief to evidence: simpler immediate feedback improves language learning and confidence in semi-open-ended questions

Xiwen Lu^{1*}, Qingyu Yang², Jian Wei¹ and Neil T. Heffernan³

¹Department of German, Russian and Asian Languages and Literature, Brandeis University, Waltham, MA, United States, ²Modern and Classical Languages and Literatures, University of Rhode Island, Kingston, RI, United States, ³Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA, United States

Semi-open-ended questions (SOE) are commonly used in online foreign language assignments; however, the effects of online feedback on these question types have received limited research attention. To address this gap, the present study employed an instructor survey and an in-class experiment to investigate how different forms of immediate feedback to SOE questions influence learning outcomes and judgment of learning (JoL). Survey findings revealed that although most instructors had access to platforms enabling immediate feedback, few utilized this feature for open-ended questions. Moreover, a majority of instructors believed that the most effective approach was to provide both correct responses (CR) and elaborated feedback (EF). Contrary to these beliefs, our experimental findings demonstrated that CR alone—feedback that simply provides the correct answer—was more effective than more complex feedback conditions, including a combination of CR and EF. Students who received CR required the least time, achieved the highest learning gains and reported stronger JoL, while CR and EF together did not yield additional benefits despite requiring more time to process. These findings challenge the common assumption that “more feedback is always better” and highlight the efficiency and pedagogical value of CR feedback for SOE questions in online foreign language learning. The study offers practical implications for instructors designing feedback strategies in technology-mediated environments.

KEYWORDS

semi-open-ended questions, online learning, foreign language learning, immediate feedback, correct response, elaborated feedback, judgment of learning

1 Introduction

Given the ongoing digital transformation, online automated feedback has become readily available across diverse educational contexts, and its positive effects on learning outcomes have been widely recognized (Swart et al., 2019). A substantial body of research (e.g., Ma et al., 2014) has demonstrated the effectiveness of immediate online feedback in promoting learning. Among the current available research on immediate online feedback, most focuses on close-ended questions in the STEM field, there is a paucity of studies on immediate feedback for semi-open-ended (SOE) questions in the realm of online learning as well as in the field of foreign language (FL) learning.

The limited research in this area can be attributed to the variability of correct answers to open-ended questions and SOE questions, particularly in terms of sentence form and word usage. Instructors are often confronted with a dilemma: either acquiring expertise in natural language processing and machine learning techniques to design assignments capable of

autonomously providing immediate feedback to SOE questions (Zhang et al., 2022), or they must rely on pre-designed questions available through platforms that are closely tied to specific textbooks and limited in the scope of questions they can support.

At the same time, SOE questions are commonly employed in FL assignments and are used more frequently than multiple-choice questions or questions with concise answers. However, grading these assignments subjectively is a time-consuming task. As a result of the grading challenges associated with SOE questions, instructors may be hesitant to utilize them. Nevertheless, the provision of immediate feedback during online learning should not be restricted solely to question types with single correct answers. The present project aims to explore the effects of immediate feedback beyond close-ended questions, with a specific focus on SOE questions in the context of FL learning. This research sheds light on the need to alleviate the workload of instructors while ensuring timely and consistent feedback for students.

In this study, we distinguish among three types of assessment items commonly used in online foreign language learning: closed-ended questions, SOE questions, and open-ended questions. Closed-ended questions, such as multiple-choice or true/false items, allow for a fixed set of responses and are well-suited for automated scoring and feedback. At the other extreme, fully open-ended questions require extended responses, such as essays or personal reflections, which demand higher levels of interpretation, critical analysis, and subjective evaluation (Bonotto, 2013). Positioned between these two, SOE questions involve short but flexible responses, such as applying a target grammar structure in a sentence or producing a lexical item in context (Zhang et al., 2022). Because SOE questions combine the efficiency of shorter responses with the variability of open-ended production, they represent a particularly challenging area for automated feedback in online learning environments. Accordingly, the present study focuses specifically on SOE questions, with an emphasis on their role in assessing grammatical accuracy and lexical knowledge.

2 Literature review

A comprehensive literature review on the feedback types in student learning and the effects of feedback on the judgment of learning is presented, along with the selection of four feedback types for the present study.

2.1 Immediate online feedback types

The provision of immediate feedback via online platforms for student learning holds immense importance in today's digital era. With the increasing prevalence of online education and remote learning, the availability of immediate feedback has become even more crucial. Immediate feedback online allows students to receive timely guidance about their performance, comprehension, and progress, enabling them to correct errors, reinforce their understanding, and adjust their learning strategies accordingly (Epstein et al., 2002).

Shute (2008) summarized six types of feedback: no feedback (NF), verification (also called knowledge of results), correct response (CR), try again, error flagging, and elaborated feedback (EF). Swart et al. (2019) provided a summary of previous studies on

feedback types and emphasized the categorization of feedback into two sub-categories, namely verification and elaboration, based on an information processing perspective. In our study, verification refers to feedback that involves confirming the accuracy or correctness of a task. Feedback types that do not necessitate verification of students' responses have been employed to provide immediate feedback to SOE questions in the experiment of this study, assuming that instructors do not have access to online auto-grading engines. When considering the absence of verification, the feedback types commonly utilized for SOE questions, as outlined by Shute (2008), included NF, CR, and EF addressing the target concept. These feedback types were arranged in ascending order of information complexity, ranging from the simplest to the most complex. In addition, they differ in whether they include a verifiable model (CR) or consist solely of explanation without a verifiable model (EF).

Different types of feedback have varying complexities and effects on learning, as supported by previous research (Enders et al., 2021). Existing studies have generally supported the notion that immediate feedback of different complexities enhances error correction compared to the absence of feedback (Kuklick et al., 2023). A large number of studies have supported the idea that the effectiveness of feedback increases with its complexity (Van der Kleij et al., 2015). Wisniewski et al. (2020) ran a meta-analysis on educational feedback research and found that the greater the amount of information contained in feedback, the more effective it tended to be. Swart et al. (2019) found that elaboration feedback outperformed verification feedback in learning from text. Enders et al. (2021) conducted a study and found that elaborated feedback yielded superior results compared to verification plus correct response in an assessment involving close-ended questions, with Kuklick et al. (2023) further supporting this finding.

However, a body of research has yielded findings contrary to the commonly held belief that more elaborated feedback is beneficial for learning (Golke et al., 2015). Kulhavy et al. (1985) demonstrated that feedback versions with higher complexity have a minor impact on students' self-correction of errors. Conversely, feedback with lower complexity offers greater advantages to learners in terms of efficiency and outcomes compared to complex feedback. R  th et al. (2021) compared the effects of verification plus correct response feedback and elaborated feedback on learning scores with mobile quiz apps but did not find a difference. Kuklick et al. (2023) found that although all feedback enhanced learning compared to no feedback, EF did not outperform CR. Jaehnig and Miller (2007) observed that students might necessitate additional time to process more complex messages, potentially diminishing the efficiency of elaborated feedback. While a longer duration of time spent can serve as a positive indicator of motivational attentiveness, it does not necessarily increase learning.

It is important to note that in the majority of the reviewed studies, both elaborated feedback and correct response feedback were accompanied by verification (Van der Kleij et al., 2015). One major reason is that most studies have focused on close-ended questions. Providing verification before presenting more complex information for close-ended questions is a common practice to establish a solid foundation for learning. Feedback that includes verifications has been observed to outperform feedback without verifications (Kuklick and Lindner, 2021). Unfortunately, immediate feedback to open-ended questions typically lacks verification.

The complexity of feedback and the presence of a verifiable model in relation to open-ended and SOE questions has received relatively little research attention. The majority of research on feedback for open-ended questions has primarily centered on natural language processing methods, which facilitate the automatic evaluation of students' answers. In the broader field of educational research, [Attali and Powers \(2010\)](#) conducted two experiments to examine the effects of providing immediate versus no feedback on open-ended verbal and math questions in an online environment. Their findings indicated that participants who received immediate feedback scored significantly higher than those who did not receive feedback. Unfortunately, we did not come across any studies in the FL learning field specifically focusing on the provision of immediate feedback for open-ended questions in the absence of auto-grading mechanisms.

To obtain a more comprehensive understanding of the influence of feedback types provided following SOE questions, the current study employed four distinct feedback types that varied in complexity, differing in whether they included a verifiable model or a detailed explanation. These types included (1) No Feedback (NF), which did not provide feedback except for telling students "answer recorded"; (2) Correct Response (CR), which provided an implicit form of verification through a model answer to the question; (3) Elaborated Feedback (EF), which provided a detailed explanation of the required learning target by listing all the attributes; and (4) All Feedback (AF), which encompassed both CR and EF.

2.2 Feedback and judgment of learning

Beyond its effects on learning outcomes, feedback has also been suggested as a valuable instrument for enhancing students' metacognitive skills related to performance ([Butler et al., 2008](#); [Labuhn et al., 2010](#); [Stone, 2000](#); [Urban and Urban, 2021](#)). Judgment of learning (JoL) is a metacognitive process through which individuals predict or assess their own learning and memory performance on a particular task or material. It involves estimating the likelihood or certainty of being able to recall or recognize information in the future ([Rhodes, 2016](#)). By examining JoL, we can gain a better understanding of how students assess their own learning from different types of feedback and how different types of feedback affect students' JoL.

There are several learning theories that help explain the concept of JoL and its relationship to feedback. Among these is cognitive load theory, which suggests that learners have limited cognitive resources and that the allocation of these resources can impact their ability to make accurate judgments about their learning ([Sweller, 1994](#)). Based on this theory, feedback that is concise, clear, and well-structured can reduce cognitive load and facilitate more accurate JoL.

Another relevant theory that contributes to the understanding of JoL and feedback is metacognitive theory. According to this theory, individuals possess metacognitive awareness and monitoring abilities that allow them to assess their own learning and make judgments about their future performance. Within this framework, feedback plays an important role by providing learners with valuable information about their current level of understanding or performance. Armed with this feedback, learners can then evaluate their progress and make informed decisions about how to adjust their learning strategies accordingly ([Flavell, 1979](#); [Pintrich, 2002](#)).

Furthermore, self-regulated learning theory emphasizes the role of feedback in self-regulation processes such as goal setting, monitoring, and adjusting strategies ([Zimmerman, 2000](#)). Feedback provides learners with information about the effectiveness of their strategies and helps them make informed decisions about how to improve their learning.

Given the existing body of research indicating a positive correlation between higher confidence levels and enhanced information retrieval from memory, we incorporated students' JoL into our investigation and examined whether different types of feedback had an impact on students' JoL. In the context of this study, JoL refers to students' self-evaluated scores indicating their predictions of their own learning outcomes following the receipt of feedback. To minimize the potential impact of JoL on learning outcomes, the cues of our JoL questions were intentionally designed differently from the test questions ([Myers et al., 2020](#)). The example of JoL question can be found in the section of 5.1.4.

2.3 Semi-open-ended questions in foreign language learning

Online teaching and learning in the FL field have been extensively examined from various perspectives, including design and teaching guidelines, particularly in response to the COVID-19 pandemic. Moreover, an increasing application of technology has been witnessed in the context of Chinese FL learning ([Liu et al., 2022](#)). Research has shown that open-ended questions develop students' problem-solving skills ([Bonotto, 2013](#)). The utilization of open-ended items allows FL students to engage in real-life problem-solving scenarios where certain information may be missing and where there is not a single predetermined solution. This approach encourages FL students to employ reasoning skills and actively contribute to the discussion by making assumptions and providing comments about the missing information.

Compared to multiple-choice questions that provide an answer list with options for students to select from, short-answer questions require respondents to construct a response. Responses to short-answer questions are usually between one and a few sentences in length. [Ye and Manoharan \(2018\)](#) classified the answers to short answer questions with two categories, concise and descriptive, which are also often referred to as close-ended questions and open-ended questions. Concise answers are fixed, and students are required to provide a number, a word, or a phrase that is unique. For example, for the question "When was the Qin dynasty founded?" the sole answer should be "221BC." A descriptive answer consists of one or a few sentences. "What majors are you interested in? Please explain your reasons" is one example in which the answers are open-ended and students can answer based on their own experiences. [Zhang et al. \(2022\)](#) identified a third category of short answer questions, known as SOE questions, which falls between the two previously described categories. SOE questions require students to express their subjective opinions within a given context, often using specific grammatical structures or key words. For example, to help students practice the Chinese grammar pattern involving the structure 把 (bǎ), an SOE question might ask: "Give a command to a robot using the 把 (bǎ) structure to tell it to do a chore for you." A sample answer could be a sentence using the grammar structure: "把桌子擦干净 (Bǎ zhuōzi cā gānjìng)," which translates to "Clean the table."

In the present study, our focus was specifically on SOE questions that elicit simple answers with a restricted range of acceptable responses. We did not include open-ended questions that require more complex answers involving critical thinking, information analysis, and detailed explanations or arguments. For the sake of brevity and clarity, we refer to our research targets as SOE questions that require the learner to provide a short, concise answer to a question while still allowing for some variability in the response. SOE questions often begin with an open-ended prompt but then provide specific instructions or parameters for the response. These questions are often used in FL education settings to assess a student's understanding of a specific topic or usage of language. Commonly used SOE question types in FL learning include translating, making sentences with a provided structure, completing dialogues, and answering questions based on the text.

Some studies have used constructed-response questions to indicate open-ended questions. We do not use constructed-response questions because this is a broader concept that ranges from fill-in-the-blank questions to essay writing questions (Kuechler and Simkin, 2010), and it is more often used for essay writing assessment (Livingston, 2009). Due to the fact that certain scoring engines have the technical capability to grade SOE questions (Livingston, 2009), we do not use non-computer-gradable questions to define the question type we are investigating either, although these technologies are not yet widely accessible to frontline instructors.

3 Study aims

3.1 Research questions

Building on previous research, the primary objectives of this study are to examine Chinese foreign language (FL) instructors' beliefs and practices regarding the provision of immediate online feedback, to explore how different types of immediate online feedback influence students' performance in responding to SOE questions, and to investigate the impact of immediate feedback on students' judgments of learning (JoL). To address these goals, the study employed an instructor survey and an experimental design to investigate the following research questions:

RQ1. What are instructors' beliefs and practices regarding the provision of immediate feedback, particularly for SOE questions?

RQ2. How do different types of immediate online feedback (CR, EF, AF, NF) affect students' learning outcomes on SOE questions?

RQ3. How does the complexity of feedback influence students' metacognitive JoL?

RQ4. Do students' performances in response to different types of feedback align with instructors' perspectives on the effectiveness of feedback types?

3.2 Research hypotheses

Given the previous literature results, we hypothesized that instructors believe that high complexity feedback (AF) to SOE

questions would lead to greater learning gains (H1), and that their instructional practices would align with these beliefs (H2). CR, EF, and AF would outperform NF (H3), and AF would potentially be the most effective in enhancing learning (H4). Students receiving more complex feedback with EF will achieve greater learning gains than those receiving only CR (H5). For the relationship between time spent on feedback and learning, we anticipated that the duration of time spent on feedback messages would positively predict learning gains (H6). We also hypothesized that feedback with high complexity (AF) would yield greater JoL scores compared to CR and EF (H7). Finally, students' performances in response to different types of feedback align with instructors' perspectives on the effectiveness of feedback types (H8).

To better align the research questions with the hypotheses, Table 1 presents a map that links each research question to its corresponding hypothesis and clarifies how the data collection instruments (Appendices A and B) were designed to address them.

4 Instructor survey

4.1 Survey objectives

Despite the popularity of SOE questions in the field of FL learning and the recognized role of immediate feedback in online learning, the provision of immediate feedback on SOE questions during online learning is limited, probably due to the challenges associated with subjective grading. It is important to note that although many instructors do not have access to ready-made question banks that offer auto-graded immediate feedback with verifications for SOE questions, this does not imply that feedback for SOE questions cannot be provided or that instructors are unable to offer timely responses. Two common approaches for providing immediate feedback to SOE questions without the support of auto-grading systems are CR, which offers correct sample answers, and EF, which provides detailed explanations such as grammar pattern rules.

Prior to delving deeper into the experiment, we first gathered the perspectives of instructors on this topic. Instructors are the ones directly responsible for delivering the assignments and feedback. Their understanding of the students, teaching styles, and learning needs is invaluable. By conducting an instructor survey, we can gain insights into the challenges they face, their preferences, and their expectations for the experiment. An instructor survey (see Appendix A) was designed to explore the attitudes, opinions, and practices of K–16 Chinese FL instructors regarding feedback on online assignments and to investigate their views on providing immediate feedback, especially for SOE questions.

4.2 Instrument

The survey used for data collection was developed by the author and reviewed by two Chinese language instructors. The survey was written in English and included four sections with 15 questions in total.

Section 1 included two general questions asking about the type(s) of institutions where the instructors taught and at what language level they taught.

TABLE 1 Alignment of research questions, hypotheses, and data collection instruments.

Research question (RQ)	Corresponding hypothesis (H)	Data source/instrument (appendices)	Analysis
RQ1. What are instructors' beliefs and practices regarding the provision of immediate online feedback, particularly for semi-open-ended (SOE) questions?	H1. Instructors generally believe that more complex feedback (e.g., AF) leads to higher learning gains. H2. Instructors' practices align with their beliefs.	Instructor Survey (Appendix 2: Survey Questionnaire)	Descriptive and inferential statistics of survey responses (frequency, percentages).
RQ2. How do different types of immediate online feedback (CR, EF, AF, NF) affect students' learning outcomes on SOE questions?	H3. Feedback conditions (CR, EF, AF) will outperform no feedback (NF). H4. AF will be the most effective condition for enhancing learning outcomes. H5. Students receiving more complex feedback with EF will achieve greater learning gains than those receiving only CR. H6. Longer time spent reading feedback messages predicts higher learning gains.	In-class Experiment (Appendix 2: Experimental Materials)	Robust ANOVA and boxplots of posttest scores across conditions: identification of which feedback type(s) led to the greatest learning gains. Robust linear regression models for each posttest result: assess the effects of two factors (with or without CR and with or without EF). Robust linear regression model of time spent on feedback and learning gains.
RQ3. How does the complexity of feedback influence students' metacognitive judgments of learning (JoL)?	H7. More complex feedback (AF) will yield higher JoL scores than simpler feedback (CR).	In-class Experiment (Appendix 2: Experimental Materials; JoL ratings)	Robust linear models of JoL confidence and learning gain scores: compare JoL scores across CR, EF, AF conditions.
RQ4. Do students' performances in response to different types of feedback align with instructors' perspectives on the effectiveness of feedback types?	H8. Students' performances in response to different types of feedback align with instructors' perspectives on the effectiveness of feedback types.	Instructor Survey (Appendix 1) and In-class Experiment (Appendix 2)	Compare survey response findings to the experiment findings.

Section 2 included seven questions in total. The first six questions were related to homework assignments and perspectives on providing immediate feedback to close-ended questions, open-ended questions, and SOE questions. To ensure that instructors understood the definition of SOE questions, an example was provided. The seventh question (Question 9) asked instructors which method they used more often when assigning homework: an online platform with the option to provide immediate feedback or another format that could not provide immediate feedback (e.g., written homework or email/upload answer sheet). Based on the feedback, instructors were asked to answer questions either in Section 3 or Section 4.

Section 3 was designed for instructors who used an online platform with the option to provide immediate feedback for homework assignments. This section consisted of three questions: How often do you use online tools to provide immediate auto-feedback to close-ended questions, open-ended questions, and SOE questions when assigning homework through an online platform? How do you provide immediate feedback to questions that are not auto-gradable? When assigning homework without immediate auto-feedback, how long does it usually take for students to receive feedback?

Section 4 was designed for instructors who did not use an online platform with the option to provide immediate feedback, and three questions were asked: How would you rate the helpfulness of providing immediate feedback to questions that are not auto-gradable for

students? Which type of immediate feedback for non-auto-gradable questions do you believe would be most effective? When assigning homework, how long does it usually take for students to receive feedback?

4.3 Participants

Participants were Chinese language instructors in the United States who teach K–16.

4.4 Data collection and analysis

The survey was conducted using a web-based platform called Qualtrics that allowed participants to respond to a series of questions via their personal electronic devices. After receiving official IRB approval in late March 2023, the recruitment of participants for the survey was carried out using a combination of email, social network platform WeChat, and personal or professional contacts. The survey was designed to take approximately 10 min to complete and was available to interested Chinese instructors for a period of 2 months. The data were analyzed with Qualtrics built-in tools and Excel.

4.5 Results

A total of 60 Chinese language instructors took the survey, of which 17 were K–12 instructors and 43 were college instructors. A total of 35 (58.33%) of the participants taught beginning-level language courses, 20 (33.33%) taught intermediate courses, and 5 (8.33%) taught advanced courses.

In total, an overwhelming majority of the respondents (83.94%) indicated that they assign homework to their students at least once per week. 30.36% of the participants reported assigning homework every day, 26.79% a few times a week, and 26.79% every week. When asked about the components of student assignments, instructors reported that, on average, 38.11% of the questions were close-ended and up to 71.49% were open-ended, including 33.53% that were semi-open-ended (SOE) questions. The top three question types used most often in assignments were essay writing, multiple choice, and translation; two out of three were open-ended questions.

The survey also indicated that most instructors have access to online platforms with the option to provide immediate feedback. Of the 56 instructors who answered this question, 46 (82.14%) reported having access to and using such platforms. Among them, 26 instructors (46.43%) reported using online platforms with immediate feedback more frequently, while 20 (35.71%) reported using both online platforms and other formats that do not support immediate feedback. Only 10 instructors (17.86%) stated that they primarily used non-digital formats without immediate feedback, such as on-paper practice.

When asked to rate on a 0–100% scale how helpful it would be for students to receive immediate feedback on questions that are not

auto-gradable, such as SOE questions, 55 out of 60 instructors answered the question. On average, they expressed 74.89% confidence that it would be helpful. However, even with the immediate feedback option available and instructors' belief that immediate feedback for open-ended and SOE questions is helpful, only a low percentage of instructors utilized the immediate feedback function, even for close-ended questions. Among the instructors who could use an online platform with the option to provide immediate feedback, only 17.78% always provided immediate feedback to close-ended questions all of the time, and 15.56% never provided immediate feedback to close-ended questions or other question types. The percentage of respondents providing immediate feedback was even lower for open-ended questions and SOE questions. Only 9.76% of instructors always provided immediate feedback to SOE questions. Figure 1 presents the percentage of instructors who provided immediate feedback to different types of questions with varying frequencies.

In terms of providing immediate feedback when assigning homework online, the respondents were mostly positive (76.61%) about its helpfulness/necessity. Fifty-five instructors who answered the question expressed 74.89% confidence that immediate feedback is helpful for questions that are not auto-gradable, such as SOE questions. Among the 56 instructors who answered the question, the majority of them (85.71%) believed that feedback containing both CR and EF is the most effective for SOE questions. Only two (3.57%) of them believed that CR is the most effective, while four (7.14%) believed that EF is the most effective.

In summary, our survey found that most instructors already had access to online platforms capable of providing immediate feedback and generally held positive attitudes toward offering such feedback

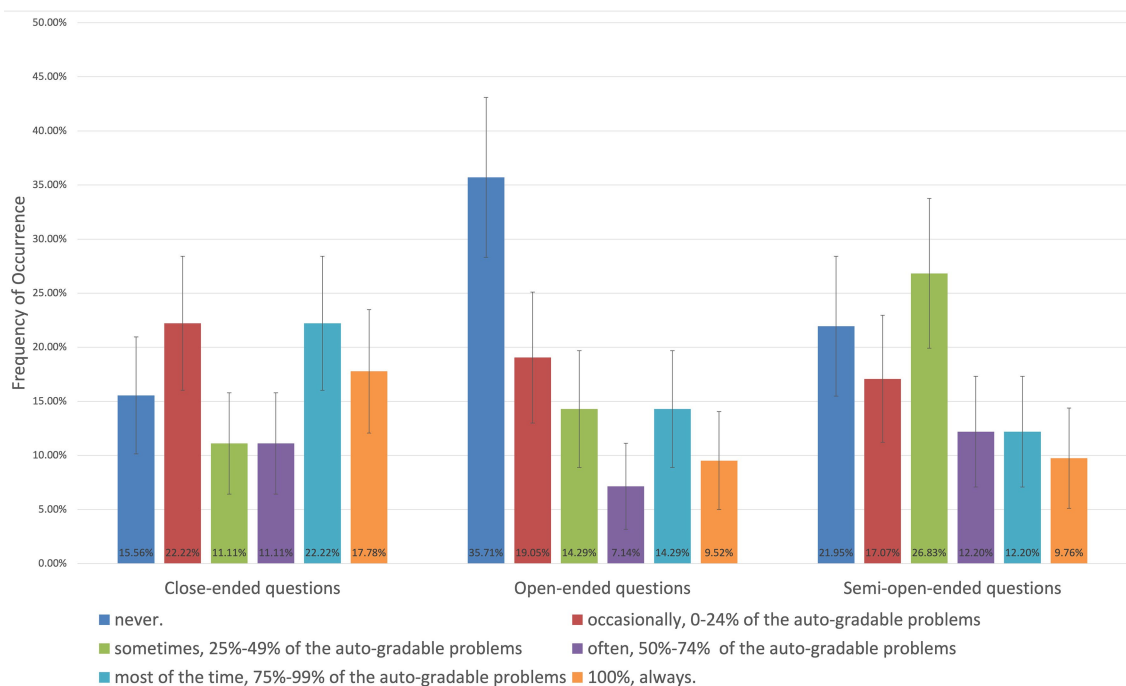


FIGURE 1

Frequency distribution of instructors' reported provision of immediate feedback across three types of questions: close-ended, open-ended, and semi-open-ended (SOE). The total answer occurrences for the close-ended questions, open-ended questions, and semi-open-ended questions were 45, 42, and 41, respectively. The errors bars stand for standard error.

across all question types. However, only a small percentage of instructors regularly provided immediate feedback, even when the option was available, and even fewer offered feedback on SOE questions. This imbalance highlights the need for continued adjustment in instructional strategies. Importantly, there are immediate, relatively low-effort actions instructors can take to enhance their teaching effectiveness without requiring significant additional infrastructure. Further discussion of these implications can be found in the General Discussion section.

When considering different types of immediate feedback for SOE questions, a large majority (85.71%) of instructors believed that providing both corrective response and detailed explanation is the most effective approach. These survey results support Hypothesis 1 (H1), suggesting that instructors generally believe that providing high-complexity feedback—such as AF—for SOE questions leads to greater student learning gains. However, despite having access to immediate feedback options and expressing the belief that immediate feedback for open-ended and SOE questions is beneficial, only a small percentage of instructors actually utilized this function, even for close-ended questions. Thus, instructors' reported practices did not align with their stated beliefs, contradicting Hypothesis 2 (H2). Building on these findings, we are now positioned to compare instructors' expectations with students' actual learning outcomes, in order to better inform both the amount and the type of immediate online feedback to be provided.

5 Experiment

5.1 Methods

5.1.1 Design

The experiment incorporated two factors and four conditions as well as four test points. The two factors examined were the provision of CR and the provision of EF. These factors were manipulated between participants, resulting in four distinct between-subject conditions: the NF condition (no EF, no CR), the CR condition (with CR, no EF), the EF condition (with EF, no CR), and the AF condition (with CR, with EF). All participants underwent the pretest as well as the three subsequent posttests. This experiment was conducted under the oversight of the ASSISTments IRB, and all participants provided informed consent prior to participating in the study.

5.1.2 The online learning platform

The experiment was conducted in-person using the ASSISTments platform, which is an online learning platform designed to provide support to students through hints, feedback, and scaffolding (Heffernan and Heffernan, 2014). All assignments, pretests, and posttests were administered during class time. Figure 2 displays the interfaces of ASSISTments as seen by students, instructors, and researchers. The pretest, quiz assignments, immediate feedback, and posttest were all delivered through the ASSISTments platform. Prior to analysis, the research data was anonymized by the platform to ensure confidentiality. For each course, daily in-class quizzes were administered to all students using their personal laptop computers. As a result, all students were already accustomed to taking online quizzes and were provided with laptops at the start of the semester to ensure inclusivity.

5.1.3 Participants

A total of 91 undergraduate students participated in the experiment as part of their in-class learning. Participants were recruited from four intermediate-level Chinese courses in the spring 2023 semester at two universities. Chinese was not their native language, and students were placed into the for-credit classes based on placement tests at the beginning of the semester. Seven students who achieved an average score of 100 on the pretest were excluded from the analysis as this indicated that they had already mastered the material.

5.1.4 Materials

The Chinese grammatical structure *ba* (把) within the scope of intermediate-level Chinese FL courses was selected to provide the content of questions, and the three primary usages of the *ba* pattern were chosen as the specific learning targets for the study. Seven SOE questions of the same difficulty but in different communicative situations were created for each usage of the *ba* pattern. Out of these seven questions, four were intentionally designed to be highly similar, with only variations in the usage of situations. These four questions were randomly selected to be pretest and posttests prior to the commencement of the experiment, in order to effectively minimize any testing differences. The rest three questions were used for learning practice. An example of the four test questions, the three practice questions can be found in Appendix B.

Three types of feedback were developed for each practice question. The first type was CR, which provides an example of the correct answer for the question. The second type was EF, which presents the necessary grammar instruction required to answer the question. The third type is AF, which combines both CR and EF. An example question with its CR, EF, and AF feedback can be found in Appendix B.

In calculating the content validity of the grammar tests and feedback, we collected expert opinions. Three CFL instructors at the college level reviewed the content to ensure the questions, situations provided, examples, and grammar explanations were appropriate and at the same difficulty level.

After receiving a feedback (except for no feedback) in the learning practice question, students were asked a JoL question to rate how likely they were to be able to use the grammar pattern correctly on a later test based on the feedback they received within a range from 0 to 100%. See the example below:

With the feedback given above, how likely will you be able to use this structure correctly on a later test? Indicate from 0% to 100%.

0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%.

5.2 Procedure

We oversaw the procedure and trained instructors before the experiment to help them familiarize themselves with the procedure of delivering the experiment. Participants were randomly assigned to one of the four conditions before the experiment started. On experiment day, before providing consent, participants were informed that they would study grammar pattern *ba* and would be asked to take tests later on the same grammar patterns. Participants were not told what type of feedback they would receive prior to studying the grammar pattern, although they were told they might not be given feedback. They were

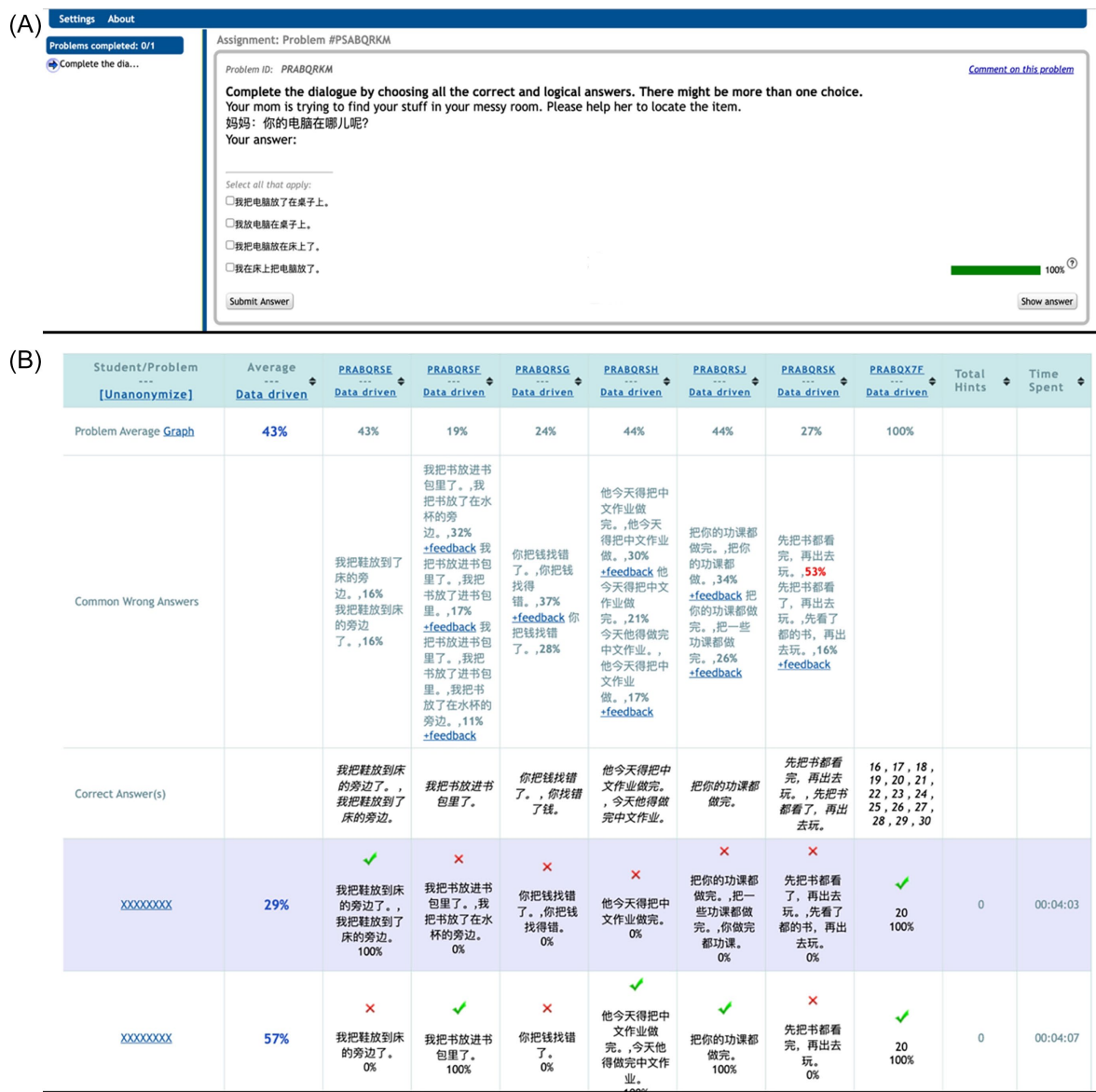


FIGURE 2
ASSISTments interfaces for students (A) and instructors (B).

told that the testing results were not connected to their course grades. All students logged into ASSISTments with laptops/computers/iPads to complete the tasks.

On the first day of the experiment, after signing the consent form, each participant responded to three pretest questions in a randomized order; completed nine practice questions covering the three usages in a randomized presentation order; and answered three questions for Posttest 1, also in a randomized order. The order of the three questions in each usage was also randomized. If assigned to a CR, EF, or AF condition, each feedback was immediately followed by a JoL question in practice. On the second day, participants completed three questions for Posttest 2 in a randomized order on ASSISTments. One week later (eighth day), participants took three questions for Posttest 3, again in a randomized order on ASSISTments. No feedback was provided for either test.

To ensure that the study met ethical requirements, all students in the four conditions received all types of feedback by the end of the experiment to guarantee equal learning opportunities. Instructors also

went over the target grammar usage in the classroom after the experiment to eliminate any remaining confusion. All the experiments took place in the classrooms to guarantee that students did not receive extra support beyond the feedback. The experimental procedure is displayed in Figure 3.

Throughout the experimental process, all participants' attempts on the learning platform were logged automatically, which served as a statistical analysis of the data. Participants from all universities that participated in this study used the same materials and followed an identical experimental procedure.

5.3 Scoring and analysis

All questions, including the pretest, practice questions, and the three posttests, underwent anonymization and were manually assessed by three raters using a grading rubric. The grading rubric utilized in

the study is provided in Table 2. Each question was assigned a maximum of 100 points and evaluated based on three criteria: structure (50 points), word usage (20 points), and communication (30 points). In cases where typing errors occurred (e.g., “我闷” instead of “我们”), these errors were considered as incorrect word usage. If an answer did not employ the designated structure at all, it received a score of 0. Likewise, if the answer utilized the structure correctly but lacked logical coherence with the question, indicating a lack of understanding of the question or grammar usage and a failure in communication, the answer also received a score of 0.

5.4 Results

5.4.1 Effects of feedback types on learning

First, to investigate the effects of the feedback types on learning, we compared the learning gains of students in each posttest between conditions. Robust ANOVA tests were conducted for each posttest learning gain. By employing robust ANOVA tests, specifically using the WRS2 package in R, we were able to overcome the limitations

associated with violations of normality and equal variance assumptions. Robust ANOVA tests are designed to handle non-normal and heteroscedastic data, making them suitable for analyzing our dataset. The analysis included the learning gain results from each of the three posttests, with an effective number of 8,000 bootstrap samples. The summary statistics of the three posttest learning gains in different conditions are included in Table 3.

In Posttest 1, the test statistic was 1.31, resulting in a p -value of 0.29. The effect size was 0.31. No significant difference was found between the conditions. For Posttest 2, the test statistic was 4.03, yielding a p -value of 0.03, and the effect size was 0.56, indicating a significant difference between the conditions. A robust *post hoc* test using trimmed means revealed a significant difference between the CR and EF conditions, 95% CI of $[-37.78, -2.68]$, test statistics value of -19.95 . Providing CR was more effective than providing EF. No significant difference was found between the AF and CR conditions (test statistics value of 13.14, CI $[-4.77, 33.46]$), suggesting that after CR is provided, providing EF or not does not make a difference. In Posttest 3, the test statistic was 3.76, resulting in a p -value of 0.05. The effect size was 0.50, indicating a marginal significant difference

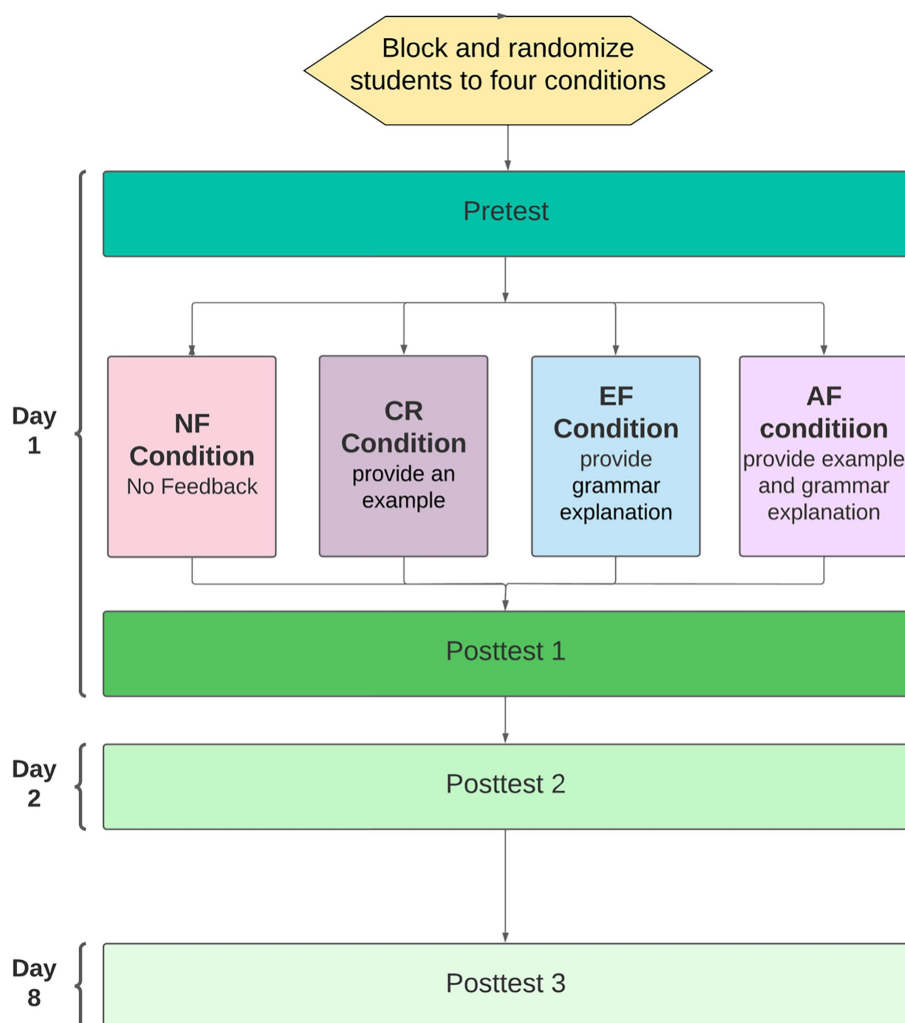


FIGURE 3
Experimental procedure and timeline.

TABLE 2 Grading rubric for evaluating student performance on all the tests.

Criteria	Unsatisfactory	Needs improvement	Good
Structure	0 Did not use the structure or used with wrong order (e.g., 把吃那碗汤) or major mistakes.	25 Used the structure correctly with a minor mistake, such as the wrong usage of or missing 了 in 把 structure, missing 要, or missing resultative complement.	50 Used the structure correctly.
Word usage	0 Three or more words not properly used in the sentence, such as using 桌 instead of 桌子.	10 One or two words in the sentence not properly used, such as using 桌 instead of 桌子, or with one or two typing errors, e.g., 把-吧.	20 All words properly used.
Communication/meaning delivery	0 Answer not logically connected to the question and cannot be used to answer the question.	15 Student showed that they understood the question, but the answer does not directly answer the question and leads to confusion.	30 Student successfully answered the question and reached the communication goal.

between the conditions. However, a robust post hoc test using trimmed means did not find significant difference between conditions.

In summary, robust ANOVA findings did not indicate that providing feedback to SOE questions resulted in better performance compared to the absence of feedback, contradicting Hypothesis 3 (H3). However, providing CR was more effective than providing EF in the Posttest 2 results. This finding differ from our Hypothesis 4 (H4), which proposed that providing AF would be the most effective in enhancing learning. They also contradict Hypothesis 5 (H5), which suggested that students receiving more complex feedback with EF will achieve greater learning gains than those receiving only CR. Figure 4 displays the boxplot of the three posttest learning gains for different conditions, which supported the observed differences between the conditions.

5.4.2 Effects of correct response and elaborated feedback on learning

To further investigate how different factors affect learning, we conducted an analysis to examine the impact of two factors, namely CR (with or without) and EF (with or without), on the learning outcome. We aimed to determine the extent to which these factors could predict students' learning gains. To achieve this, we employed robust linear regression models for each posttest result. Effect coding was employed in our analysis to enable a meaningful assessment of the effects of the predictor variables on the outcome variable.

The model used to analyze the learning gain for Posttest 1 did not account for a significant amount of the variance: $F(3, 73) = 1.63$, $p = 0.19$, $R^2 = 0.06$, adjusted $R^2 = 0.02$. The analysis further revealed that neither the two factors nor their interaction significantly predicted the Posttest 1 learning gain ($p > 0.05$ for both cases).

In the case of Posttest 2, the model yielded a significant explanation for a portion of the variance in the learning gain: $F(3, 63) = 3.73$, $p = 0.02$, $R^2 = 0.14$, adjusted $R^2 = 0.10$. These findings indicated that the model demonstrated a meaningful relationship between the predictor variables and the value of the Posttest 2

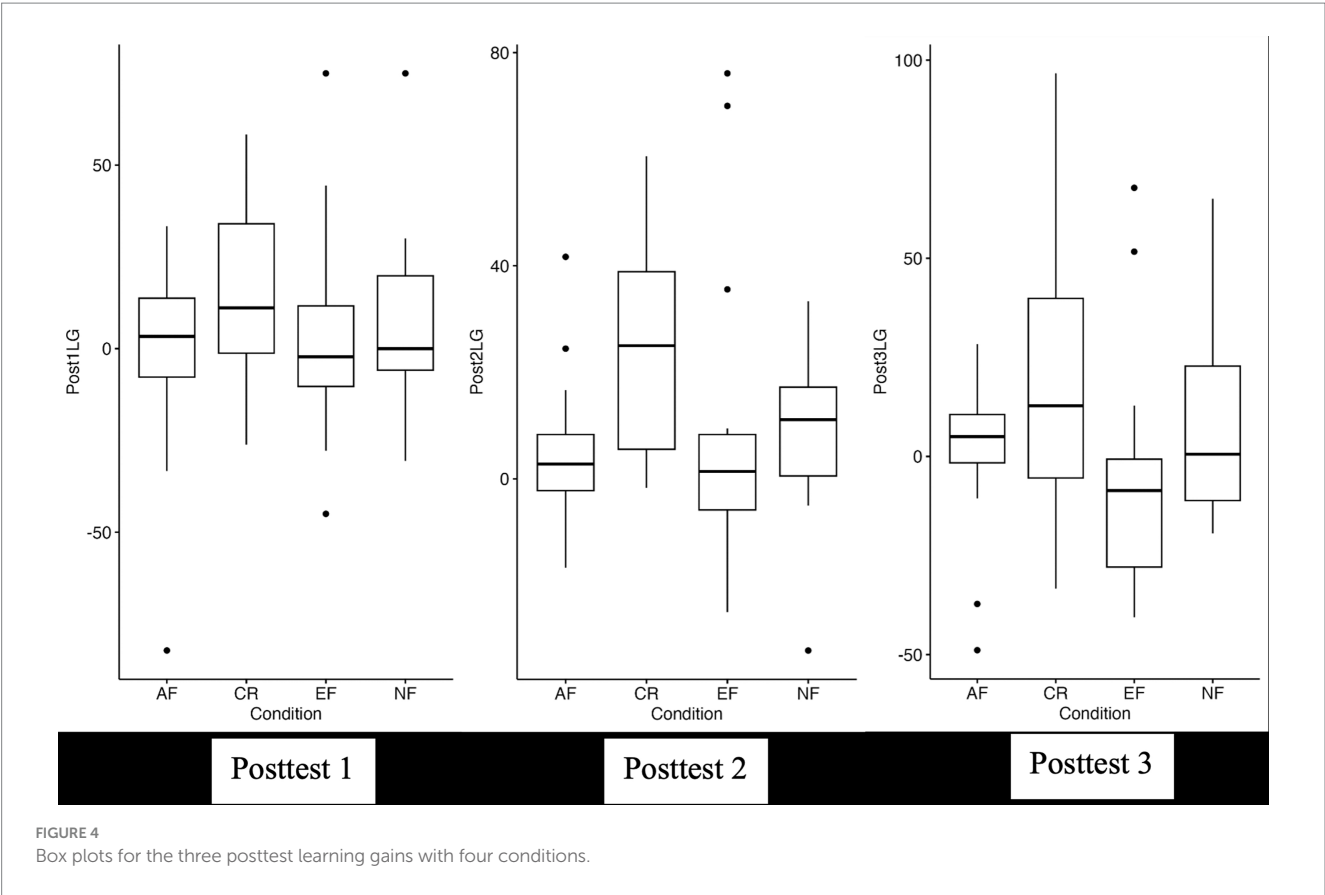
learning gain. Specifically, the analysis revealed that the EF factor significantly predicted the value of the Posttest 2 learning gain ($\beta = -10.25$, $t [63] = -2.12$, $p = 0.04$, 95% CI $[-19.89, -0.60]$). This suggests that the presence or absence of EF had a significant negative impact on the learning gain for Posttest 2, with negative values indicating a decrease in learning gain. CR, on the other hand, did not significantly predict the value of the Posttest 2 learning gain ($p > 0.05$). The interaction between the two factors of CR and EF did not significantly predict the value of the Posttest 2 learning gain ($p > 0.05$), indicating that the combined effect of the two factors did not have a statistically significant influence on the learning gain for Posttest 2.

The model for Posttest 3 did not find a significant variance either: $F(3, 61) = 2.15$, $p = 0.10$, $R^2 = 0.12$, adjusted $R^2 = 0.07$. However, the factor of EF significantly predicted the value of Posttest 3 learning gain in the negative way: $\beta = -16.20$, $t (61) = -2.43$, $p = 0.02$, 95% CI $[-29.53, -2.88]$. The factor CR and the interaction did not significantly predict the Posttest 3 learning gain ($p > 0.05$ for both cases). Table 4 summarizes the statistics of the three models.

The results revealed that providing EF or not played a significant role in the posttest learning gains. Regardless of whether CR was provided or not, the provision of EF negatively affected learning. The difference was not significant in Posttest 1, but it emerged in both Posttest 2 and Posttest 3. To better illustrate how providing EF affects learning gains under different conditions, a boxplot for the three posttest learning gains with the two factors is presented in Figure 5. As shown in the boxplots, across all three posttests, conditions without EF (marked in blue) consistently outperformed those with EF (marked in yellow), regardless of whether CR was provided. This finding contradicts our Hypothesis 5 (H5) and suggests that students' performance in response to different types of feedback does not align with instructors' perspectives on feedback effectiveness. While the earlier survey indicated that instructors believed high-complexity feedback, such as AF, would lead to greater learning gains, our experimental results revealed that more complex feedback actually had a negative impact on learning.

TABLE 3 Means and standard deviations of the three posttests’ learning gain scores of the four conditions as well as Robust ANOVA results for each posttest.

Posttest time point	Condition	Learning gain			Robust ANOVA		
		<i>n</i>	Mean	SD	Effective size (η_c^2)	Statistic	<i>p</i> -value
Posttest 1	No Feedback (NF)	16	7.64	23.90	0.31	1.31	0.29
	Correct Response (CR)	20	16.90	25.00			
	Elaborated Feedback (EF)	19	1.810	25.90			
	All Feedback (AF)	22	1.67	25.60			
Posttest 2	No Feedback (NF)	15	8.93	16.10	0.56	4.03	0.03
	Correct Response (CR)	17	25.00	20.80			
	Elaborated Feedback (EF)	18	8.27	26.80			
	All Feedback (AF)	17	5.16	13.40			
Posttest 3	No Feedback (NF)	15	8.15	24.10	0.50	3.76	0.05
	Correct Response (CR)	16	18.20	33.00			
	Elaborated Feedback (EF)	18	−7.19	28.70			
	All Feedback (AF)	16	1.13	20.00			



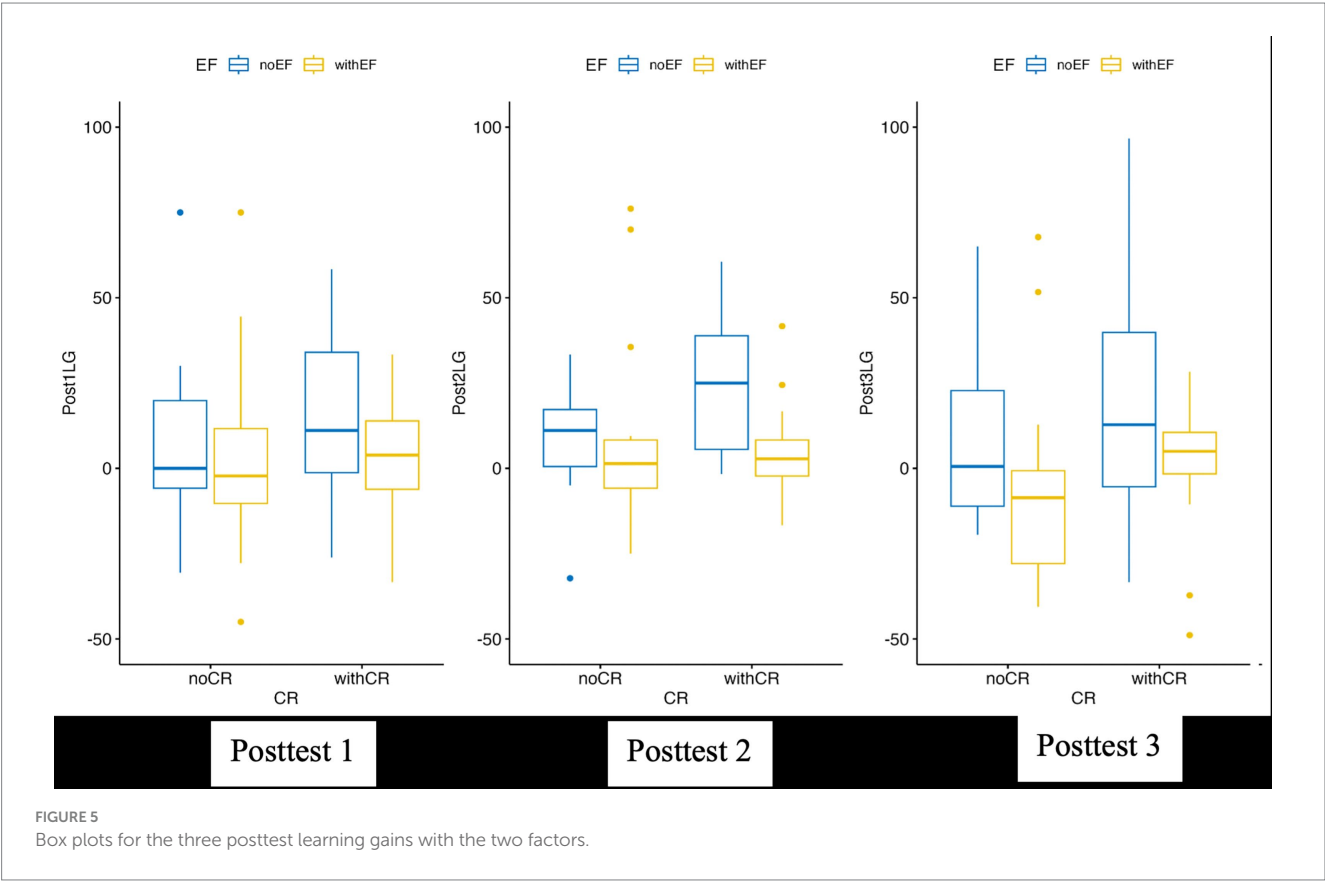
5.4.3 Time spent on learning

To further investigate why EF was less effective than instructors had expected, and to understand how students interacted with feedback under different conditions, we analyzed data collected

during the experiment. We first examined the time spent on problems as a potential factor. The time spent on each problem was determined by subtracting the problem start time from the problem end time. We constructed a robust linear model, clustering the data by student

TABLE 4 Robust linear regression models for the three posttests' learning gains regressed on two factors: correct response (CR) and elaborated feedback (EF).

Posttest time point	Factors	Coefficient (β)	Robust standard error	t -statistic	p -value	95% Confidence Interval for β		df
						Lower bound	Upper bound	
Posttest 1	(Intercept)	7.01	2.87	2.44	0.02	1.29	12.73	73
	CR	4.57	5.74	0.80	0.43	−6.88	16.01	73
	EF	−10.54	5.74	−1.83	0.07	−21.98	0.91	73
	CR * EF	−9.43	11.49	−0.82	0.41	−32.32	13.47	73
Posttest 2	(Intercept)	11.84	2.41	4.91	6.893 e-06	7.02	16.66	63
	CR	6.48	4.83	1.343	0.18	−3.16	16.13	63
	EF	−10.25	4.83	−2.12	0.04	−19.89	−0.60	63
	CR * EF	−19.18	9.66	−1.99	0.051	−38.48	0.11	63
Posttest 3	(Intercept)	5.07	3.33	1.52	0.13	−1.59	11.73	61
	CR	9.18	6.66	1.38	0.17	−4.14	22.51	61
	EF	−16.20	6.66	−2.43	0.02	−29.53	−2.88	61
	CR * EF	−1.73	13.33	−0.13	0.90	−28.37	24.92	61



ID and employing the CR² standard error type, to assess the relationship between the time spent on each problem and the condition. Twenty-three problems that took longer than 5 min were deemed errors and subsequently excluded from the data. The model did not yield significant results ($F[3, 95] = 1.32, p = 0.27$), indicating

no significant differences in time spent on problems were observed between the conditions.

We then examined the time spent on feedback as a potential factor. After each problem, students spent some time reading the provided feedback (if any). The time spent reading feedback was

calculated as the feedback end time minus the feedback start time. Five attempts with feedback durations longer than 4 min—significantly exceeding the average response time—were considered outliers and removed from the dataset. Summary statistics are presented in Table 5. Seven students who achieved a perfect score of 100 on the pretest were excluded from the analysis. This resulted in a slightly unbalanced number of participants across conditions, leading to different practice counts in the experiments.

We constructed a robust linear model to examine the time spent reading the content of three types of feedback (CR, EF, and AF), clustered by student ID, and employed the CR² standard error type. The model has a significant effect: $F(2, 76) = 4.11, p = 0.02$, Multiple $R^2 = 0.02$, Adjusted $R^2 = 0.02$. Compared to providing all feedback (condition AF), providing only CR took significantly less time, $\beta = -7.45, t = -2.59, p = 0.01279$. No significant difference was found between the EF and AF conditions, indicating that the EF condition required a similarly long amount of time as the AF condition. The regression results can be seen in Table 6.

This result appears reasonable, given that both the AF and EF conditions included a greater amount of information. However, it is noteworthy that despite the increased informational content and longer reading times in the AF and EF conditions, neither led to better learning performance than the CR condition on the subsequent posttests. Based on these findings, it can be concluded that providing CR may be the most efficient approach among the three feedback conditions. This outcome contrasts with Hypothesis 6 (H6), which predicted that the duration of time spent on feedback messages would positively correlate with learning gains.

5.4.4 Effects of feedback types on JoL

After receiving each feedback text (if applicable) in the learning practice questions, students were asked to provide a JoL score. To investigate the impact of different types of feedback on students' JoL scores, we analyzed the JoL scores across the three conditions where feedback was provided, taking students' learning gains into consideration. Figure 6 displays the plot depicting the relationship between students' JoL scores and learning gain scores across different feedback types.

A robust linear model was employed with the CR² standard error type, and data was clustered by student ID. The model yielded a significant effect, indicating that the type of feedback had a significant influence on students' JoL scores ($F[3, 621] = 35.34, p < 0.001$). The multiple R^2 value was 0.15. When holding students' problem scores constant, a significant difference in confidence levels was observed between the CR condition and the AF condition. Specifically, the CR condition received significantly higher confidence ratings compared to the AF condition ($\beta = 0.04, t = 2.15$, standard error = 0.02, $p = 0.03$). The regression results can be seen in Table 7.

TABLE 5 Summary statistics for in-experiment time spent on feedback by condition.

Condition	<i>n</i>	Mean (in seconds)	SD
AF	228	18.80	26.30
CR	191	11.40	10.70
EF	210	15.60	19.50

n indicates the number of feedback attempts collected for analysis in each condition.

These findings suggest that, when controlling for students' problem scores, the provision of only CR led to significantly higher levels of JoL scores compared to the AF condition, which combined CR and EF. Given that CR feedback represented the lowest level of complexity and AF the highest, the results indicate that simpler feedback may actually boost students' confidence more than complex feedback when responding to SOE questions. This finding contradicts our Hypothesis 7 (H7), which anticipated that more complex feedback would lead to higher learning confidence. When considering students' confidence in their learning, providing simpler feedback may be more effective.

6 General discussion

The purpose of this study is to gain a deeper understanding of instructors' beliefs and practices regarding the provision of immediate online feedback on assignments—particularly for open-ended and SOE questions—and to investigate the effectiveness of different types of immediate feedback on students' learning of SOE questions in FL classrooms. Our survey results reveal that most instructors hold positive attitudes toward providing immediate feedback to SOE questions; however, very few of them put it into practice. Most instructors believe that providing immediate feedback with both a correct response and elaboration would be the most effective response to SOE questions; however, this result is not supported by our experiment on student learning gains. Our experiment results indicate that CR and EF act differently on students' learning, with CR more effective than EF. Surprisingly, providing EF negatively affected SOE learning regardless of whether CR is provided or not. The difference was not significant in Posttest 1 but emerged in Posttest 2 and Posttest 3. Moreover, students spent significantly more time engaging with feedback in the AF condition compared to the CR condition. No significant difference was found in feedback reading time between the CR and EF conditions. Considering the learning outcomes, providing CR appears to be the most efficient approach to supporting student learning. Lastly, when controlling for students' performance on practice problems, the provision of CR resulted in significantly higher levels of JoL scores compared to the condition where both CR and EF were provided. However, providing only EF did not significantly raise students' JoL scores. Further elaboration will be provided in the subsequent sections.

6.1 Misalignment between instructor beliefs and student learning outcomes

The survey results address RQ1, showing that most instructors believe more complex feedback leads to better learning outcomes. However, the actual use of immediate online feedback remains limited in teaching practice, particularly for open-ended and SOE questions. These findings reveal several important trends in current Chinese language teaching practices regarding feedback strategies, particularly in the context of online assignments. First, while instructors regularly assign homework—most doing so at least weekly, and a significant portion doing so daily—their use of different question types suggests a notable emphasis on open-ended tasks. In fact, over 70% of assignment questions were reported to be open-ended, with 33.53% classified as SOE. This reflects a pedagogical orientation that values student expression, deeper processing, and language production over

TABLE 6 Robust linear model estimating the effect of condition on feedback reading time, with 95% confidence intervals.

Condition	Coefficient	Std. Error	t-value	Pr (> t)	CI Lower	CI Upper	df
(Intercept)	18.83	2.62	7.20	1.47e-07	13.44	24.21	25.18
Condition CR	−7.45	2.87	−2.59	0.01	−13.24	−1.67	44.44
Condition EF	−3.24	3.39	−0.96	0.34	−10.06	3.58	48.27

The model compares the Correct Response (CR) and Elaborated Feedback (EF) conditions against the All Feedback (AF) condition as the reference group.

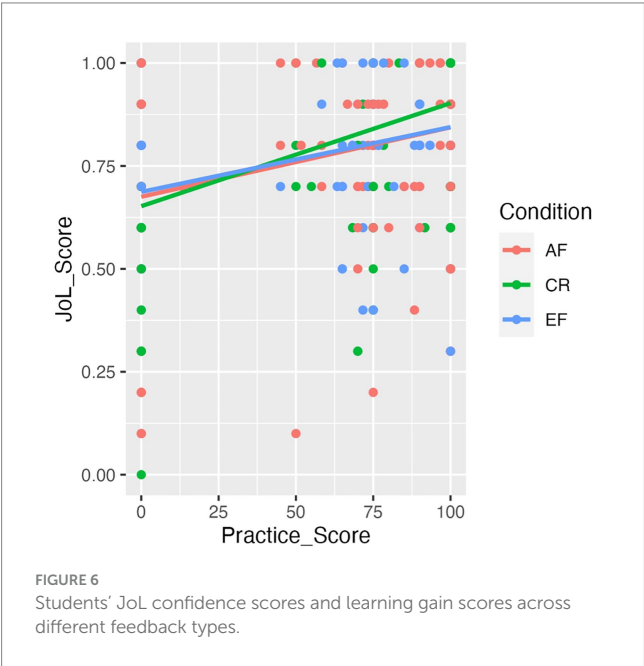


TABLE 7 Robust linear model of predictor score and condition on JoL scores with 95% confidence intervals.

Model term	Coefficient	Std. Error	t-value	Pr(> t)
(Intercept)	0.66	0.018	37.17	<2e-16
Score	0.002	0.0002	9.93	<2e-16
Condition CR	0.04	0.02	2.15	0.03
Condition EF	0.003	0.02	0.16	0.87

rote recall. However, this emphasis raises a critical instructional challenge: such question types are often not easily auto-gradable and thus require more nuanced feedback mechanisms.

Despite the reported accessibility of online platforms capable of delivering immediate feedback, and instructors’ general agreement about the benefits of doing so, the actual use of these online functions remains limited. While 82.14% of instructors have access to platforms that allow immediate feedback, only a small proportion consistently take advantage of this functionality. Most notably, only 9.76% always provided immediate feedback to SOE questions. This gap between belief and practice highlights a significant disconnect—what we might call a “feedback-action gap.” It suggests that barriers may exist not in awareness or attitude, but in implementation, workload concerns, platform usability, or a lack of targeted training.

The strong belief among instructors (85.71%) that AF is the most effective form of feedback for SOE questions is pedagogically intuitive,

as complex feedback is often assumed to help students not only correct their errors but also understand the underlying reasoning. However, this assumption may also contribute to instructors’ hesitation to provide immediate online feedback, as delivering high-complexity feedback can be time-consuming and increase their workload.

Interestingly, only a small fraction of instructors (3.57%) believed that simple corrective feedback alone is the most effective, indicating a general preference among educators for more elaborated feedback strategies. However, in the subsequent experiment, students who received simpler corrective feedback consistently outperformed those who received more complex feedback (EF or AF) across multiple posttests. The reduced learning gains associated with more complex immediate feedback may be attributed to the increased cognitive load placed on students, as well as the perceived appeal of the elaborations that instructors are capable of providing.

The discrepancy between instructors’ beliefs and students’ actual learning outcomes reveals a significant misalignment in expectations versus empirical effectiveness, addressing RQ4 by demonstrating that students’ performance in response to different types of feedback does not align with instructors’ perspectives on feedback effectiveness. This finding challenges the assumption that more feedback is inherently better, highlighting the need for instructors to reconsider feedback strategies and calibrate the complexity of their immediate feedback based on students’ cognitive readiness and task difficulty. Bridging this gap between instructor perception and learner response is essential for aligning pedagogical intentions with effective learning outcomes.

In examining the “feedback-action gap,” it is important to consider factors beyond instructor workload that may limit the use of immediate online feedback for SOE questions. Instructors may lack training in designing effective feedback tailored to SOE tasks, perceive that such feedback is primarily suited to closed-ended questions, or worry that providing immediate feedback could oversimplify complex language use. This signals a need for professional development that goes beyond advocating for immediate feedback in theory, and instead supports instructors with practical tools, examples, and workflows for delivering targeted feedback efficiently—especially for complex and open-ended question types.

6.2 The most efficient feedback for semi-open-ended questions

For RQ2, which examined how different types of immediate online feedback (CR, EF, AF, NF) affect students’ learning outcomes on SOE questions, our findings contradicted Hypotheses 3, 4, 5, and 6. One notable finding of our study is that the provision of both correct response and elaboration resulted in significantly more time spent on feedback compared to providing only correct response; however, the provision of elaborated feedback did not. This finding is in line with expectations as the AF condition contained the most

information. However, interestingly, we did not observe a significant increase in the time spent on EF condition despite it also containing significantly more information compared to CR feedback; this implies that simply including more information in the feedback does not necessarily lead to a longer learning time. On the other hand, the combination of CR and EF feedback may significantly increase the time students spend on feedback as CR helps students verify their answers in a certain way. Without verification, students may feel confused and may not invest additional time in examining detailed feedback. They may be unsure of what to focus on in the feedback or may become overconfident in their answers, leading them to spend less time on the feedback details. Hattie and Timperley (2007) emphasized that when feedback fails to establish a clear goal for students, they may not perceive the need to bridge the gap between the goal and their current status through feedback, thereby limiting their engagement with detailed feedback information.

A significant body of research has consistently shown that learning outcomes improve with increased time spent engaging with feedback (Kuklick et al., 2023). However, our findings do not support the assumption that spending more time on feedback necessarily enhances learning—at least in the context of SOE questions. Further research is needed to identify more effective strategies for providing immediate feedback on open-ended tasks.

When considering the efficiency of learning and examining learning gains in relation to the time spent reading feedback, the provision of only correct response proved to be the most efficient. Specifically, feedback with only correct response resulted in the highest learning gains while requiring the least amount of time. These findings suggest that, for SOE questions, providing correct response is the most time-efficient and effective feedback. Instructors aiming to optimize both student learning and time-on-task should take this into consideration when designing feedback strategies to open-ended questions.

6.3 Impact of two immediate feedback factors: presence or absence of correct responses and elaborated feedback

In contrast to the results of the instructors' survey, the provision of both correct response and elaborated feedback for SOE questions did not yield the best learning outcomes. Instead, providing only correct response proved to be a superior choice compared to elaborated feedback; providing elaborated feedback actually had a negative impact on learning for SOE questions. These findings contrast with H7 and previous research that generally supported the idea that feedback providing more information is more effective.

It is important to note that most of the research (Van der Kleij et al., 2015) supporting the increase in learning performance with feedback complexity is based on feedback for close-ended questions, in which all complex feedback (CR, EF, AF) includes verifications. One key difference between feedback for open-ended and close-ended questions is the lack of verification. Our study strongly indicates that without verification, learning outcomes may not improve with an increase in feedback complexity for SOE questions. Upon examining the CR and EF conditions in our experiment, we observe that while providing a correct response does not directly provide verification of students' responses, it still allows students to compare their answers

to a correct response, implicitly aiding them in verifying their answers. In contrast, the EF condition only provides bullet points of grammar explanations, which may make it more difficult for students to verify their answers. Feedback that includes verifications has been observed to outperform feedback without verifications (Kuklick and Lindner, 2021). Longer explanations without the provision of verification may lead to students' uncertainty regarding their performance on the task as well as uncertainty about how to respond to the feedback, potentially confusing students and creating learning difficulties (Weaver, 2006). Hattie and Timperley (2007) also assume that forms of feedback are "most useful when they assist students in rejecting erroneous hypotheses and provide direction for searching and strategizing" (pp. 91–92). When providing elaborated feedback for SOE questions, the answers provided by students are not verified, making it challenging to reject erroneous hypotheses through only elaborated feedback. In such cases, longer explanations may lead to confusion rather than provide clear guidance.

Support for this viewpoint can be substantiated by theories from the field of psychology. Cognitive load theory suggests that when learning materials cause confusion, cognitive resources are diverted toward tasks that are not directly relevant to the learning process (Sweller, 2010). EF might have provided overwhelming feedback to students compared to CR, especially when an example which can be used as verification is not provided. Moreover, scaffolded instructions were not provided in the EF in the present study. Unlike scaffolded instruction, where elaboration is introduced progressively and supported by teacher guidance, the EF provided in our study presented information in a static, text-based form without interactive scaffolding. This lack of step-by-step guidance may have reduced learners' ability to integrate the explanations into their existing knowledge, thereby hindering rather than supporting learning. Together, these factors suggest that elaborated feedback, when presented without scaffolding or verification, may not enhance learning and can even be detrimental by overloading cognitive resources.

One remaining question is why the AF condition did not outperform the CR condition in our experiment. Although the AF condition included both correct response and elaborated feedback, the addition of elaborated feedback did not lead to improved learning outcomes. Since no significant difference was observed between the AF and CR conditions, one possible explanation is that the correct response alone was sufficient to support students' learning, while the elaborated feedback did not contribute additional benefit. This lack of impact from elaborated feedback may be attributed to its length, cognitive load, or other contextual factors that warrant further investigation—such as the narrative style of elaborated feedback, the classroom learning environment, or student motivation. Another potential explanation is the relatively small sample size of our study, which may have limited the statistical power needed to detect subtle effects. Future research should explore these possibilities in greater depth to better understand when and how elaborated feedback may effectively enhance learning.

Another important consideration concerns the cross-linguistic implications of our findings. The present study was conducted in the context of Mandarin Chinese, a logographic language in which feedback often involves both semantic and orthographic elements. It remains an open question whether the effects we observed—particularly the greater efficiency of correct response feedback

compared to elaborated or combined feedback—would generalize to alphabetic languages, where orthographic-phonological mapping may reduce the cognitive load of processing written feedback. Research on alphabetic languages has shown that elaborated feedback can sometimes support deeper learning by clarifying morphological or syntactic rules (Varnosfadrani and Ansari, 2011), whereas in Chinese, elaboration may impose additional cognitive demands due to the complexity of characters and grammar patterns. Thus, it is possible that alphabetic and logographic systems interact differently with feedback types, with simpler verification-based feedback (e.g., correct responses) being especially efficient for Chinese learners. Future cross-linguistic studies are needed to investigate whether the balance between cognitive load and informativeness in feedback holds similarly across writing systems, or whether language-specific features shape the optimal design of immediate feedback.

6.4 Correct response improves judgment of learning outcomes

In addition to learning outcomes, we also found that the provision of only correct response led to higher judgment of learning compared to the condition in which both correct response and elaborated feedback were provided. This finding addresses RQ3, which examines how feedback complexity influences students' metacognitive JoL, and it aligns with previous research showing that variations in feedback content can significantly increase cognitive load (Taxipulati and Lu, 2021). One possible explanation is that the increased cognitive load associated with both the EF and AF conditions may have mediated the observed decrease in JoL scores.

An alternative explanation for the observed results draws on the idea that learners tend to base their judgments of learning on the fluency or ease of processing of the information presented (Castel et al., 2007; Yue et al., 2013). In this context, it is likely that participants perceived the concise, one-sentence correct examples provided in the CR condition as more fluent and easier to process than the more detailed grammar explanations found in the EF and AF conditions, even though the latter contained a greater amount of instructional information.

While the teacher survey emphasized instructors' belief that elaborated feedback would lead to better learning outcomes, our experimental findings reveal a potential disconnect between these beliefs and actual learner experiences. Specifically, the results suggest that more complex feedback may not enhance learners' perceived understanding and may even hinder it, highlighting the importance of aligning instructional strategies not only with pedagogical intentions but also with students' cognitive and metacognitive responses.

One more important consideration is the influence of learner characteristics on the effectiveness of feedback interventions, such as prior knowledge, motivation, and self-esteem. Studies have shown that these individual differences can act as moderators, shaping how learners respond to feedback and the learning outcomes they achieve. Although our participants were recruited from the same instructional level based on placement tests and a pretest was administered to ensure a degree of homogeneity, variation in prior knowledge, motivation, and emotional dispositions remained. Mihalca and Mengelkamp (2020) demonstrated that higher levels of prior

knowledge improve the accuracy of learners' metacognitive judgments, which in turn affects how learners utilize feedback and regulate their learning processes, ultimately leading to higher performance than those with lower prior knowledge. Glaser and Richter (2022) similarly found that practice testing effects depend on both cognitive factors (e.g., prior knowledge) and affective factors (e.g., motivation, test anxiety), which can modulate the extent of learning gains. Maier and Klotz (2022) further emphasized that self-esteem can influence the impact of feedback in digital environments, with higher self-esteem fostering greater receptivity. Since the present study focused primarily on the effects of feedback types and measured only JoL after each feedback instance, future research would benefit from incorporating broader assessments of learner characteristics prior to feedback. Such measures would allow for a more nuanced understanding of how feedback interacts with individual differences to shape learning outcomes.

6.5 Implications for practice

The idea for the present study originated from frontline teaching practice, and it is our intention that the findings be applied to real-world classroom teaching practices. These findings raise several important pedagogical implications.

First, this disconnect between instructional intentions and student outcomes highlights the need for instructors to rethink and restructure their teaching strategies and curriculum design to better align with how students actually process and benefit from immediate online feedback. While instructors may assume that providing more detailed feedback automatically improves learning outcomes, our results suggest that more is not always better. In the context of semi-open-ended questions, simpler feedback such as a correct response may be more effective in promoting both learning and learners' confidence. The gap between instructor beliefs and learner experience also underscores the importance of student-centered feedback design. Since students' judgments of learning affect their motivation, giving feedback that boosts confidence without overwhelming them may contribute to sustained engagement.

The study also found that although many instructors have access to online platforms capable of delivering immediate feedback, only a small proportion consistently utilize these tools, particularly for open-ended questions. It is therefore essential for instructors to critically evaluate their use of available technological resources and integrate them more effectively into their instructional practices to enhance student learning outcomes. This finding also highlights the potential need for professional development programs focused on the online assignment design and implementation of immediate online feedback.

What's more, the results of this study support the integration of more immediate feedback for open-ended questions. Specifically, we recommend providing immediate feedback in the form of a concise correct response, such as a working example, rather than relying solely on elaborated explanations. If elaborated feedback is provided, it should be accompanied by a clear correct response to ensure clarity and reduce cognitive load. Students who receive correct responses for SOE questions not only demonstrate improved learning outcomes but also report higher judgments of learning, which may positively influence their motivation and future learning behaviors.

It is important to note that the findings of this study are most directly applicable to SOE questions that assess the accurate use of a specific grammatical structure. These types of questions are different from fully open-ended tasks, such as essays or reflections, where the effectiveness of feedback may be shaped by different cognitive and pedagogical dynamics. For example, while correct responses proved most effective for SOE questions in our study, more elaborated forms of feedback may be better suited for tasks that require critical reasoning or extended writing. Clarifying this distinction underscores the pedagogical value of immediate online feedback for focused linguistic practice while also pointing to the need for future research that examines immediate online feedback effects in broader, more complex open-ended contexts.

6.6 Limitations and future research

First, the number of participants in both the survey and the experiment was relatively small, which may have limited the ability to detect differences between conditions in Posttest 1. As such, the findings of this study should be viewed as foundational rather than definitive. Future research should increase the number of participants to obtain more robust results, replicate this work with larger and more diverse samples across different languages, proficiency levels, and educational systems to obtain more robust results and strengthen the generalizability of the findings.

Since our research primarily focused on feedback, JoL questions were administered solely after the feedback conditions and not in the NF condition. This design choice precluded us from comparing JoL between the feedback and NF conditions. Future studies should consider incorporating JoL measures within the NF condition in order to complete the comparative framework and validate JoL trends across all feedback types. Of course, this should be done without unduly interfering with the structure of assessment. It might also be useful to survey these students about their perceptions of the feedback.

Extending the current experimental findings through the investigation of additional types of feedback with varying complexities would be beneficial. This could involve exploring EF with differing levels of information, or considering EF both with and without verification components. Although providing verification feedback to SOE questions poses challenges, it would be intriguing to investigate whether the inclusion of verification would alter students' learning behaviors and outcomes.

What's more, it is noteworthy that various types of SOE questions exist beyond the scope of the questions examined in this study. Subsequent research endeavors should consider including different question types to investigate potential variations in findings. Furthermore, future studies should expand to encompass a wider range of open-ended questions as they may also demonstrate the potential benefits of immediate online feedback. Considering the limited number of studies investigating feedback provided for open-ended questions, it is our aspiration that our research will offer valuable insights into this matter.

Additionally, the growing integration of artificial intelligence in education has significant implications for the design and delivery of online feedback, particularly for SOE and open-ended questions. AI-powered systems, such as natural language processing tools and intelligent tutoring platforms, are increasingly capable of analyzing

student responses and generating personalized, immediate feedback at scale. This technological advancement opens up new possibilities for implementing more sophisticated feedback strategies without increasing instructors' workload. As AI continues to evolve, it is critical to align the design of AI-generated feedback with pedagogical evidence—such as the value of concise correct responses demonstrated in this study—ensuring that automation enhances rather than hinders the learning experience. Future research could explore how AI and NLP tools can generate concise and accurate CR feedback for SOE and open-ended questions, providing timely guidance that addresses learners' specific errors without overwhelming time with excessive information. These systems have the potential to personalize feedback by adapting to individual learners' proficiency levels and error patterns, ensuring that each student receives support tailored to their unique needs. Investigating such models could also shed light on how to better include learners' engagement, motivation, and perceptions in the feedback-providing system.

Finally, an important direction for future research is the use of longitudinal designs to examine the sustained effects of feedback on foreign language learning. While our study focused on immediate posttest and short-term delayed (no longer than 1 week) posttest outcomes, it is crucial to understand how feedback influences longer-term retention and the transfer of skills in order to evaluate its enduring pedagogical value. Longitudinal studies could track learners over several weeks or months to determine whether the benefits of immediate feedback persist and facilitate the generalization of knowledge to new contexts.

7 Conclusion

This study represents an initial effort to examine the effects of immediate online feedback on open-ended questions in foreign language learning contexts. It contributes to a deeper understanding of how different types of immediate feedback influence both learning outcomes and learners' judgments of learning, what's instructors' beliefs and practices on immediate online feedback to open-ended questions, as well as the disconnect between instructional intentions and student outcomes.

The study highlights a gap between instructors' perceptions and actual student performance, pointing to a need for reflection and adjustment in online immediate feedback strategies and teacher training. Key findings reveal that providing only correct responses not only led to the most efficient learning outcomes but also resulted in higher JoL scores compared to more complex feedback formats. These results suggest that, contrary to common instructional beliefs, simpler feedback may be more effective and confidence-boosting for learners working with semi-open-ended questions.

Practically, this research underscores the importance of encouraging the use of immediate feedback tools already available in many classrooms. It also supports the integration of correct response immediate feedback as a pedagogically sound and cognitively efficient approach for open-ended language tasks.

While the generalizability of these findings requires further validation through larger and more diverse samples, this study lays the groundwork for continued exploration into effective immediate feedback design for open-ended language learning. It is our hope that these insights will inform future studies and instructional practices

aimed at enhancing both the efficiency and effectiveness of immediate online feedback in foreign language education.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors without undue reservation.

Ethics statement

The studies involving humans were approved by Brandeis Institutional Review Board and Worcester Polytechnic Institute's Institutional Review Board. Brandeis IRB approved the teacher survey and WPI IRB approved documents that govern the use of ASSISTments for running randomized controlled trials and analyzing anonymized data. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

XL: Methodology, Data curation, Investigation, Supervision, Writing – review & editing, Conceptualization, Writing – original draft, Formal analysis, Project administration. QY: Data curation, Writing – original draft. JW: Writing – original draft, Data curation. NH: Supervision, Writing – original draft, Funding acquisition, Software.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. We acknowledge funding from NSF (2118725, 2118904, 1950683, 1917808, 1931523, 1940236, 1917713, 1903304, 1822830, 1759229, 1724889, 1636782, & 1535428), IES (R305N210049, R305D210031, R305A170137, R305A170243, R305A180401, R305A120125 & R305R220012), GAANN (P200A180088 & P200A150306), EIR (U411B190024 S411B210024, & S411B220024), ONR (N00014-18-1-2768) and NHI (via a SBIR R44GM146483).

References

- Attali, Y., and Powers, D. (2010). Immediate feedback and opportunity to revise answers to open-ended questions. *Educ. Psychol. Meas.* 70, 22–35. doi: 10.1177/0013164409332231
- Bonotto, C. (2013). Artifacts as sources for problem-posing activities. *Educ. Stud. Math.* 83, 37–55. doi: 10.1007/s10649-012-9441-7
- Butler, A. C., Karpicke, J. D., and Roediger, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 918–928. doi: 10.1037/0278-7393.34.4.918
- Castel, A. D., McCabe, D. P., and Roediger, H. L. (2007). Illusions of competence and overestimation of associative memory for identical items: evidence from judgments of learning. *Psychon. Bull. Rev.* 14, 107–111. doi: 10.3758/BF03194036
- Enders, N., Gaschler, R., and Kubik, V. (2021). Online quizzes with closed questions in formal assessment: how elaborate feedback can promote learning. *Psychol. Learn. Teach.* 20, 91–106. doi: 10.1177/1475725720971205
- Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., et al. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *Psychol. Rec.* 52, 187–201. doi: 10.1007/BF03395423
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive-developmental inquiry. *Am. Psychol.* 34, 906–911. doi: 10.1037/0003-066X.34.10.906
- Glaser, J., and Richter, T. (2022). The testing effect in the lecture hall: does it depend on learner prerequisites? *Psychol. Learn. Teach.* 22, 159–178. doi: 10.1177/14757257221136660 (Original work published 2023)
- Golke, S., Dörfler, T., and Artelt, C. (2015). The impact of elaborated feedback on text comprehension within a computer-based assessment. *Learn. Instr.* 39, 123–136. doi: 10.1016/j.learninstruc.2015.05.009
- Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487
- Heffernan, N. T., and Heffernan, C. L. (2014). The ASSISTments ecosystem: building a platform that brings scientists and teachers together for minimally invasive research

Acknowledgments

This study is part of my doctoral dissertation research conducted at Worcester Polytechnic Institute. I would like to sincerely thank my dissertation committee, Neil Heffernan, Adam Sales, Stacy Shaw, and Shijuan Liu, for their invaluable guidance and constructive feedback throughout the research and writing process.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that Gen AI was used in the creation of this manuscript. Generative AI was used to polish the language.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2025.1654809/full#supplementary-material>

- on human learning and teaching. *Int. J. Artif. Intell. Educ.* 24, 470–497. doi: 10.1007/s40593-014-0024-x
- Jaehng, W., and Miller, M. L. (2007). Feedback types in programmed instruction: a systematic review. *Psychol. Rec.* 57, 219–232. doi: 10.1007/BF03395573
- Kuechler, W. L., and Simkin, M. G. (2010). Why is performance on multiple-choice tests and constructed-response tests not more closely related? Theory and an empirical test. *Decis. Sci. J. Innov. Educ.* 8, 55–73. doi: 10.1111/j.1540-4609.2009.00243.x
- Kuklick, L., Greiff, S., and Lindner, M. A. (2023). Computer-based performance feedback: effects of error message complexity on cognitive, metacognitive, and motivational outcomes. *Comput. Educ.* 200:104785. doi: 10.1016/j.compedu.2023.104785
- Kuklick, L., and Lindner, M. A. (2021). Computer-based knowledge of results feedback in different delivery modes: effects on performance, motivation, and achievement emotions. *Contemp. Educ. Psychol.* 67:102001. doi: 10.1016/j.cedpsych.2021.102001
- Kulhavy, R. W., White, M. T., Topp, B. W., Chan, A. L., and Adams, J. (1985). Feedback complexity and corrective efficiency. *Contemp. Educ. Psychol.* 10, 285–291. doi: 10.1016/0361-476X(85)90025-6
- Labuhn, A. S., Zimmerman, B. J., and Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: the influence of feedback and self-evaluative standards. *Metacogn. Learn.* 5, 173–194. doi: 10.1007/s11409-010-9056-2
- Liu, S., Wang, Y., and Zhan, H. (2022). "Perspectives of instructors and students on online Chinese teaching and learning in 2020: preliminary findings" in *Teaching the Chinese language remotely*. ed. S. Liu (Cham: Palgrave Macmillan), 349–372.
- Livingston, S. A. (2009). Constructed-response test questions: why we use them; how we score them. R&D connections. *Educ. Test. Serv.* 11. Available at: <https://files.eric.ed.gov/fulltext/ED507802.pdf>
- Ma, W., Adesope, O. O., Nesbit, J. C., and Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: a meta-analysis. *J. Educ. Psychol.* 106, 901–918. doi: 10.1037/a0037123
- Maier, U., and Klotz, C. (2022). Personalized feedback in digital learning environments: classification framework and literature review. *Comput. Educ. Artif. Intell.* 3:100080. doi: 10.1016/j.caeai.2022.100080
- Mihalca, L., and Mengelkamp, C. (2020). Effects of induced levels of prior knowledge on monitoring accuracy and performance when learning from self-regulated problem solving. *J. Educ. Psychol.* 112, 795–810. doi: 10.1037/edu0000389
- Myers, S. J., Rhodes, M. G., and Hausman, H. E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Mem. Cogn.* 48, 745–758. doi: 10.3758/s13421-020-01025-5
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory Into Pract.* 41, 219–225. doi: 10.1207/s15430421tip4104_3
- Rhodes, M. G. (2016). "Judgments of learning: methods, data, and theory" in *Oxford handbook of metacognition*. eds. J. Dunlosky and S. K. L. Tauber (Oxford University Press), 65–80.
- Rüth, M., Breuer, J., Zimmermann, D., and Kaspar, K. (2021). The effects of different feedback types on learning with mobile quiz apps. *Front. Psychol.* 12:665144. doi: 10.3389/fpsyg.2021.665144
- Shute, V. J. (2008). Focus on formative feedback. *Rev. Educ. Res.* 78, 153–189. doi: 10.3102/0034654307313795
- Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educ. Psychol. Rev.* 12, 437–475. doi: 10.1023/A:1009084430926
- Swart, E. K., Nielen, T. M., and de Sikkema-Jong, M. T. (2019). Supporting learning from text: a meta-analysis on the timing and content of effective feedback. *Educ. Res. Rev.* 28:100296. doi: 10.1016/j.edurev.2019.100296
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learn. Instr.* 4, 295–312. doi: 10.1016/0959-4752(94)90003-5
- Sweller, J. (2010). "Cognitive load theory: recent theoretical advances" in *Cognitive load theory*. eds. J. L. Plass, R. Moreno and R. Brünken (Cambridge: Cambridge University Press), 29–47. doi: 10.1017/CBO9780511844744
- Taxipulati, S., and Lu, H. D. (2021). The influence of feedback content and feedback time on multimedia learning achievement of college students and its mechanism. *Front. Psychol.* 12:706821. doi: 10.3389/fpsyg.2021.706821
- Urban, K., and Urban, M. (2021). Anchoring effect of performance feedback on accuracy of metacognitive monitoring in preschool children. *Eur. J. Psychol.* 17, 104–118. doi: 10.5964/ejop.2397
- Van der Kleij, F. M., Feskens, R. C. W., and Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: a meta-analysis. *Rev. Educ. Res.* 85, 475–511. doi: 10.3102/0034654314564881
- Varnosfadrani, A. D., and Ansari, D. N. (2011). The effectiveness of error correction on the learning of morphological and syntactic features. *World J. English Lang.* 1. doi: 10.5430/wjel.v1n1p29
- Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assess. Eval. High. Educ.* 31, 379–394. doi: 10.1080/02602930500353061
- Wisniewski, B., Zierer, K., and Hattie, J. (2020). The power of feedback revisited: a meta-analysis of educational feedback research. *Front. Psychol.* 10:3087. doi: 10.3389/fpsyg.2019.03087
- Ye, X., and Manoharan, S. (2018). Machine learning techniques to automate scoring of constructed-response type assessments, Proceedings of 28th EAAEIE annual conference, Hafnarfjörður, Iceland.
- Yue, C. L., Castel, A. D., and Bjork, R. A. (2013). When disfluency is--and is not--a desirable difficulty: the influence of typeface clarity on metacognitive judgments and memory. *Mem. Cogn.* 41, 229–241. doi: 10.3758/s13421-012-0255-8
- Zhang, L., Huang, Y., Yang, X., Yu, S., and Zhuang, F. (2022). An automatic short-answer grading model for semi-open-ended questions. *Interact. Learn. Environ.* 30, 177–190. doi: 10.1080/10494820.2019.1648300
- Zimmerman, B. J. (2000). "Attaining self-regulation: a social cognitive perspective" in *Handbook of self-regulation*. eds. M. Boekaerts, P. R. Pintrich and M. Zeidner (San Diego: Academic Press), 13–39. doi: 10.1016/B978-012109890-2/50031-7