Chunked Feedback in Teacher-Student Interaction Promotes Long-Term Learning Transfer: A Behavioral and Near-Infrared Hyperscanning Study

Authors: Zhu Yi, Hu Yi, Zhu Yi, Hu Yi

Date: 2024-02-24T12:18:29+00:00

Abstract

Detailed content feedback promotes deep-level learning, such as transfer. However, how the presentation mode of feedback in teacher-student interactions influences long-term learning transfer and its interpersonal neural basis remains unclear. This study employed a face-to-face teacher-student question-and-answer feedback task to investigate the long-term promoting effect of chunked presentation of feedback on learning transfer, the underlying cognitive processes, and its interpersonal neural basis through two dyadic experiments (behavioral, fNIRS hyperscanning). The results revealed that chunked feedback facilitated long-term transfer in students with low prior knowledge. Chunked error correction mediated the relationship between feedback presentation mode and long-term transfer. During the process of providing and receiving chunked feedback, teachers and students exhibited greater interpersonal brain synchronization in frontal and parietal regions, and frontal interpersonal brain synchronization predicted long-term transfer and chunked error correction. These findings provide novel insights into the cognitive neural basis of pedagogical feedback as it authentically occurs in classrooms from an interpersonal perspective, and offer practical implications for enhancing the effectiveness and efficiency of teaching feedback.

Full Text

Chunking Feedback in Instructor-Learner Interaction Facilitates Long-Term Learning Transfer: Behavioral and fNIRS Hyperscanning Studies

ZHU Yi[1], **HU Yi**[2]
([1] Department of Psychology, Ningbo University, Ningbo 315211)

([2] School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062)

Abstract

Elaborated content feedback promotes deep learning, such as transfer. However, how the presentation format of feedback in instructor-learner interaction influences long-term learning transfer and its interpersonal neural basis remains unclear. This study employed a face-to-face question-and-answer feedback task and conducted two dyadic experiments (behavioral, fNIRS hyperscanning) to investigate the long-term facilitative effects of chunked feedback on learning transfer, the underlying cognitive processes, and their interpersonal neural foundations. The results revealed that chunked feedback enhanced long-term transfer in students with low knowledge foundations. Chunked error correction mediated the relationship between feedback presentation format and long-term transfer. During the provision and reception of chunked feedback, instructors and learners exhibited greater brain-to-brain synchrony in frontal and parietal regions, with frontal synchrony predicting both long-term transfer and chunked error correction. These findings offer new insights from an interpersonal perspective into the cognitive-neural basis of pedagogical feedback as it naturally occurs in classrooms and provide practical implications for enhancing the effectiveness and efficiency of instructional feedback.

Keywords: feedback, chunking, long-term transfer, instructor-learner interaction, fNIRS hyperscanning, brain-to-brain synchrony

Received: 2023-09-04
Funding: Ministry of Education Humanities and Social Sciences Youth Project (23YJC190043); Zhejiang Provincial Universities Basic Research Fund (SJWZ2024016); Ningbo University Teaching Research Project (JYXM2024037)
Corresponding Authors: ZHU Yi, HU Yi
Email: zhuyi@nbu.edu.cn; yhu@psy.ecnu.edu.cn

1 Introduction

Teaching is inseparable from instructor-learner interaction, which constitutes the reciprocal influence and dynamic exchange between teachers and students (Watanabe, 2013). Instructional feedback represents a typical form of instructor-learner interaction in authentic classrooms, typically involving teachers providing information about the gap between students' current performance and learning goals. Feedback effectively drives attitudinal and behavioral development as well as knowledge and skill acquisition (Hattie & Timperley, 2007). Knowledge acquisition manifests as both recognition and transfer, with the latter representing deeper learning built upon the former—involving the application of knowledge gained in one context to another situation. Learning transfer is widely established as an instructional objective and is closely related to students' problem-solving abilities in novel contexts (Prenzel & Mandl, 1993). Previous research has demonstrated that elaborated content feedback facilitates learning transfer; for instance, providing correct answers supplemented with explanatory reasons or illustrative examples deepens learners' understanding and promotes knowledge application in new situations (Butler et al., 2013; Finn et al., 2018; Zhu et al., 2022). However, in real instructor-learner interactions, if we maintain identical feedback content but alter its presentation format—for example, by chunking according to conceptual relationships—how might this affect learning transfer? Beyond immediate learning gains, can chunked feedback produce long-term benefits? Instructional feedback constitutes a bidirectional process involving dynamic knowledge transmission through continuous interaction between instructors and learners. To uncover the cognitive-neural processes underlying this dynamic, continuous, and bidirectional pedagogical process, research must move beyond static learning materials, discontinuous learning procedures, and single-person non-interactive or pseudo-interactive paradigms, instead adopting an interpersonal perspective that simultaneously records and analyzes brain signals from both interaction partners (Tan et al., 2023).

1.1 Effects of Feedback Presentation Format on Learning

Prior research indicates that maintaining identical information content while altering presentation format influences information processing. For example, presenting memory materials in a chunked format enhances both short-term and long-term memory (Gobet et al., 2001). A chunk represents a collection of elements with strong internal associations but weak connections to elements in other chunks (Chase & Simon, 1973). In teaching and learning, emphasizing relationships among learning content helps learners acquire advanced knowledge in ill-structured domains and facilitates problem-solving in novel situations (Spiro et al., 1991). Presenting complex action sequences in chunked formats can alter learners' recall strategies and promote transfer (Cohen & Sekuler, 2010), while grouping feedback information by valence (positive or negative) influences implicit perceptual category learning (Smith et al., 2014).

Notably, chunking feedback information may simultaneously alter feedback timing, introducing feedback delay (Smith et al., 2014). Previous research on feedback timing has yielded inconsistent results. Some researchers argue that immediate feedback more effectively prevents error encoding in memory, thereby enhancing language, procedural knowledge, and motor skill learning (Anderson et al., 2001). Others support delayed feedback, suggesting it reduces proactive interference and facilitates forgetting of earlier errors while processing subsequent corrective information (Kulhavy & Anderson, 1972). Subsequent studies found that delayed feedback's facilitative effects manifest only after extended intervals (e.g., 7 days) on tests of recognition, recall, and transfer (Butler et al., 2007; Mullet et al., 2014; Smith & Kimball, 2010).

1.2 Cognitive-Neural Basis of Feedback-Promoted Learning

Two primary cognitive mechanisms may underlie feedback's learning benefits. First, according to cognitivist learning theory, feedback's main function is error correction. Specific, targeted, or information-rich feedback not only corrects knowledge errors but also helps adjust strategic processing errors (Kulhavy & Stock, 1989; Narciss & Huth, 2004; Bangert-Drowns et al., 1991). Second, metacognitivist theory posits that feedback effectively identifies gaps between current performance and goals, prompting learners to invest additional cognitive effort to narrow these gaps and thereby improve performance (Nicol & McFarlane-Dick, 2006; Sadler, 1989). Feedback information that more specifically, proximally, or appropriately identifies performance-goal gaps better stimulates learning engagement and cognitive effort (Song & Keller, 2001; Krijgsman et al., 2019).

Previous research has found that feedback processing activates frontal and parietal regions in recipients, including the anterior cingulate cortex (ACC), dorsolateral prefrontal cortex (DLPFC), and parietal cortex. The ACC primarily supports basic feedback functions such as error detection and expectancy violation (Cavanagh et al., 2012; Luft et al., 2013; Mars et al., 2005), while DLPFC and superior parietal lobule participate in more complex feedback processing, including error correction and performance adjustment (Crone et al., 2008; van Duijvenvoorde et al., 2008; Zanolie et al., 2008). Frontal and parietal activity can predict feedback-based learning performance across various domains, including paired-associate memory (Arbel et al., 2013), motor learning efficiency (van der Helden et al., 2010), and reading and mathematics performance (Peters et al., 2017). Feedback processing activates not only recipients' brains but also providers' brains. For example, when teachers provide correctness feedback and monitor students' (confederates') associative learning performance, teachers' ACC activity correlates with students' error expectancy, while teachers' insula and ventromedial prefrontal activity correlate with students' value expectancy (Apps et al., 2015). Therefore, to understand how feedback presentation format influences learning's cognitive-neural processes, simultaneously investigating both parties' brain activity during authentic instructor-learner interactions is essential, though currently understudied.

1.3 Interpersonal Neural Basis of Instructor-Learner Interaction

Recent social interaction neuroscience has shifted from third-person to second-person perspectives. This research logic holds that neural activity during authentic second-person social interaction fundamentally differs from neural responses elicited by observing social stimuli from a third-person perspective (Schilbach et al., 2013). Hyperscanning technology, which simultaneously records brain activity from two or more individuals during tasks, offers possibilities for revealing the neural basis of real social interaction (Redcay & Schilbach, 2019). Research indicates that brain-to-brain synchrony supports successful interpersonal communication and likely constitutes the interpersonal neural foundation of social interaction (Hasson et al., 2012; Jiang et al., 2012). Although brain-to-brain synchrony's cognitive significance remains debated, studies consistently indicate it serves as a key mechanism for achieving behavioral, emotional, and cognitive alignment among interaction partners, involving behavioral coordination, emotional empathy, social conformity, language comprehension, and interpersonal bonding, thereby reflecting dynamic cognitive-neural mechanisms in complex social interactions (Kelsen et al., 2022; Shamay-Tsoory et al., 2019; Tan et al., 2023).

Teaching, as a form of social interaction, typically involves dynamic, continuous information transmission and reception between instructors and learners. Single-brain metrics generally reflect individual information processing and have limited power to reveal the neural basis of instructor-learner interaction (Tan et al., 2023). Hyperscanning studies have revealed that instructor-learner brain-to-brain synchrony indexes effective teaching and learning (Bevilacqua et al., 2019; Holper et al., 2013; Nguyen et al., 2022). Due to fMRI's temporal resolution limitations and spatial constraints on task types, it has rarely been applied to authentic teaching interaction studies for simultaneous data collection. EEG and functional near-infrared spectroscopy (fNIRS) more easily enable simultaneous brain activity recording from both parties during instructor-learner interaction. Compared to EEG, fNIRS offers higher spatial resolution and greater tolerance for motion artifacts (Lloyd-Fox et al., 2010), making it more suitable for studying authentic teaching contexts. fNIRS hyperscanning studies have found that frontal or temporoparietal brain-to-brain synchrony during instructor-learner interaction can indicate effective teaching strategies, such as scaffolding and high-frequency interactive approaches (Pan et al., 2018, 2020; Zheng et al., 2018), and predict learning performance in domains including vocal performance (Pan et al., 2018), mathematics (Liu et al., 2019; Zheng et al., 2018), and conceptual knowledge recognition and transfer (Pan et al., 2020; Zhu et al., 2022). During elaborated feedback-based instructor-learner interaction, parietal brain-to-brain synchrony reflects students' deep conceptual understanding and predicts knowledge transfer (Zhu et al., 2022). Furthermore, instructor-learner brain-to-brain synchrony can provide process-level information about students' knowledge understanding, such as the time point of achieving understanding and temporal lag patterns between instructors and learners (Liu et al., 2019; Zheng et al., 2018; Zhu et al., 2022), thereby enhancing understanding of the dynamic continuity in authentic classroom teaching. However, most previous research has focused on associations between brain-to-brain synchrony and immediate learning gains, rarely examining whether brain-to-brain synchrony continues to predict long-term learning performance.

1.4 Current Study

This study employed a conceptual teaching task based on instructor-learner Q&A feedback (Zhu et al., 2022) and conducted two sequential experiments. Experiment 1 was a dyadic behavioral experiment investigating whether maintaining identical feedback content while altering only presentation format (chunking) could further promote deep learning (transfer) and its long-term effects, while exploring the underlying cognitive processes. Based on previous research on chunked feedback information and chunked learning materials, chunked feedback was defined as simultaneously presenting answers and examples for two related concepts (Smith et al., 2014; Cohen & Sekuler, 2010; Gobet et al., 2001; Spiro et al., 1991). Because fNIRS offers higher spatial resolution and greater tolerance for motion artifacts (Lloyd-Fox et al., 2010), making it more suitable for studying dynamic, continuous multi-person teaching activities, Experiment 2 incorporated fNIRS hyperscanning to simultaneously record instructors' and learners' brain activity, aiming to further investigate the neural basis of chunked feedback presentation on students' long-term learning transfer during instructor-learner interaction.

In Experiment 1, students learned psychology concepts and received either chunked (two related concepts) or separate (one concept) elaborated feedback (correct answer and example) from instructors. Given that prior knowledge background represents a crucial factor influencing knowledge comprehension and transfer (Gick & Holyoak, 1987), and that students' pre-feedback knowledge level affects feedback effectiveness (Fyfe et al., 2012; Krause et al., 2009), a learning introduction phase was implemented to manipulate students' knowledge foundation before receiving feedback, thereby examining potential moderating effects. Experiment 1 employed a fully between-subjects design: Feedback Presentation (Chunked vs. Separate) × Knowledge Foundation (High vs. Low). Following the learning session, student performance was measured across recognition and transfer dimensions, and cognitive effort was assessed. To investigate long-term gains from chunked feedback, a second knowledge test was administered after 7 days, following previous research (Butler et al., 2007; Smith & Kimball, 2010). Additionally, chunked error correction was quantified as the number of concept pairs that changed from incorrect to correct between pre- and post-tests. Experiment 1 hypotheses were: (1) Novices or low-knowledge students benefit more from supportive feedback (Paas et al., 2003; Sweller et al., 1998), while experienced or high-knowledge students depend less on additional supportive information (Renkl & Atkinson, 2003; Sweller et al., 1998). Elaborated feedback promotes knowledge transfer (Butler et al., 2013; Finn et al., 2018; Zhu et al., 2022), and presenting learning content in chunks rather than separately provides more support and opportunities for concept discrimination (Spiro et al., 1991; Chase & Simon, 1973). Therefore, we hypothesized that for low-knowledge students, elaborated feedback presented in chunks would more effectively support learning and promote knowledge transfer. (2) Because chunked information benefits long-term memory (Gobet et al., 2001) and chunked presentation-induced feedback delay may produce delayed-retention effects (Kulhavy & Anderson, 1972), we hypothesized that chunked feedback would also promote long-term learning performance, particularly transfer after 7 days. (3) Based on potential cognitive mechanisms of feedback-promoted learning (Bangert-Drowns et al., 1991; Nicol & McFarlane-Dick, 2006), we hypothesized that feedback presentation format would influence long-term knowledge transfer through chunked error correction or cognitive effort.

In Experiment 2, fNIRS hyperscanning was added to simultaneously record instructors' and learners' brain activity during interaction. To exclude potential confounding between feedback presentation format and timing, a pseudo-chunk feedback condition (two less-related concepts) was added. Experiment 2 hypotheses were: (1) Chunked feedback would produce better long-term transfer performance than non-chunked feedback (including pseudo-chunk and separate feedback). (2) Feedback presentation format would influence long-term knowledge transfer through chunked error correction or cognitive effort. (3) Human feedback processing activates frontal and parietal regions in both providers and recipients (Apps et al., 2015; Crone et al., 2008; Luft et al., 2013), while chunk processing relies on DLPFC and posterior parietal cortex (Alamia et al., 2016; Bor et al., 2003; Jin et al., 2020; Pammi et al., 2012). Authentic teaching interaction elicits synchronized brain activity in frontal and temporoparietal regions (Tan et al., 2023; Zhu et al., 2022). Therefore, we hypothesized that chunked feedback would elicit greater instructor-learner brain-to-brain synchrony in frontal and parietal regions associated with both feedback and chunk processing. (4) Because chunking of linguistic information depends more on frontal cortex (Jin et al., 2020; Grodzinsky & Santi, 2008) and frontal activity relates to long-term memory and learning performance (Sakai & Passingham, 2003; Squire et al., 1993), and instructor-learner frontal brain-to-brain synchrony indicates effective teaching strategies (Pan et al., 2018, 2020) while chunked presentation represents a more effective instructional approach (Spiro et al., 1991; Cohen & Sekuler, 2010), we hypothesized that frontal brain-to-brain synchrony during chunked feedback would positively correlate with students' long-term transfer performance. (5) Because social interaction-induced brain-to-brain synchrony reflects cognitive alignment between interaction partners (Shamay-Tsoory et al., 2019) and emerges in the mentalizing network supporting mutual understanding, including frontal cortex (Kelsen et al., 2022), and error correction reflects students' understanding converging with instructors' understanding, we hypothesized that frontal brain-to-brain synchrony would positively correlate with chunked error correction.

2 Experiment 1

2.1 Participants

Referencing previous instructor-learner interaction studies, 20-24 instructor-learner dyads per condition were recruited to achieve effect sizes of 0.20-0.25 (Liu et al., 2019; Pan et al., 2020; Zheng et al., 2018; Zhu et al., 2022). Power analysis using G*Power 3.1 (Effect size = 0.20, $\alpha = 0.05$, 1 - $\beta = 0.95$) yielded a planned sample size of 81. Experiment 1 recruited 127 East China Normal University students: 47 served as instructors, required to major in psychology or sociology and have completed at least one education course or have teaching experience (M age = 21.80, SD = 2.12, 18 males); 80 served as learners, required to major in non-psychology and non-sociology fields (M age = 20.67, SD = 1.96, 12 males) and score below passing (<0.6) on a pre-test matching concepts with examples. Twelve instructors (6 males), 10 instructors (6 males), 13 instructors (3 males), and 11 instructors (3 males) were randomly assigned to high-knowledge-chunked feedback, high-knowledge-separate feedback, low-knowledge-chunked feedback, and low-knowledge-separate feedback groups, respectively. On the same day, each instructor conducted one-on-one teaching with one or two randomly assigned unfamiliar students using the same method. For instructors, knowledge foundation and feedback presentation were between-subjects variables to prevent psychology majors from guessing the experimental purpose and to enhance teaching consistency. This yielded 80 instructor-learner dyads: 20 in high-knowledge-chunked feedback, 19 in high-knowledge-separate feedback, 21 in low-knowledge-chunked feedback, and 20 in low-knowledge-separate feedback. All participants had normal or corrected-to-normal vision and no neurological disorders. Each participant read and signed an informed consent form before the experiment. The study was approved by the university's human subjects ethics committee.

2.2 Materials and Assessment

A separate group of psychology majors (N = 20, 4 males, M age = 24.45, SD = 2.89) was recruited to pair 12 judgment and decision-making psychology concepts (see Appendix 1, adapted from Rawson et al., 2015) based on their relationships. Results showed that 20 people paired foot-in-the-door effect with door-in-the-face effect, 16 paired availability heuristic with representativeness heuristic, 15 paired fundamental attribution error with self-serving bias, 14 paired hindsight bias with counterfactual thinking, 14 paired deindividuation with social facilitation, and 9 paired observer effect with exposure effect. The first five pairings were made by \$ \$70% of participants, while the sixth pairing was made by only 45%.

These participants subsequently rated the relatedness of the first five concept pairs on a 1-7 scale (from extremely low to extremely high). Wilcoxon signed-rank tests revealed that the first four pairs' relatedness ratings were significantly above the midpoint (Ms > 4.80, ps < 0.026). However, the fifth pair (deindividuation and social facilitation) showed no significant difference from the midpoint (M = 4.00, SD = 1.00, p = 0.839). Considering these results, we selected the first five concept pairs as experimental materials and chunked them accordingly.

Teaching materials included concept terms, definitions, and two examples (see Appendix 1). Examples represented real-life manifestations or applications of concepts, adapted from previous research and textbooks (Finn et al., 2018; Rawson et al., 2015; Zimbardo et al., 2012). The same psychology majors rated the situational similarity of each concept's two examples on a 7-point scale (1 = extremely low similarity, 7 = extremely high similarity). Ratings for the 10 concepts averaged M = 4.28, SD = 0.88, ranging from 2.85 to 5.5. A Wilcoxon signed-rank test showed no significant difference from the midpoint

($p = 0.447$). Kendall's W coefficient for inter-rater reliability was 0.43 ($p < 0.001$), indicating significant consistency. The experimenter then pre-selected examples for feedback and transfer measurement, which remained fixed across all participants.

2.3 Procedure

The experimental task occurred in two sessions separated by approximately 4 days (Figure 1). The first session was conducted in the laboratory or online via conferencing software, while the second session took place in the laboratory. During the first session, instructor participants received standardized training on teaching content and procedures (approximately 30 minutes). Afterward, instructors took home printed teaching materials to study and memorize concept definitions and examples. Before the second session, experimenters required instructors to recall the teaching procedure and randomly checked two concept definitions and examples; only after correct recall could the formal teaching task begin.

Figure 1. Experiment 1 task and procedure

During the first session, learner participants completed a pre-learning test (15-minute limit) measuring their prior knowledge of the 10 psychology concepts. The test had two parts: Part 1 required matching 10 definitions to corresponding terms; Part 2 required matching 10 examples to corresponding terms (with 12 judgment and decision-making terms as options, see Appendix 1; test format adapted from Finn et al., 2018). Knowledge foundation was quantified by pre-test accuracy. Pre-test accuracy did not differ significantly between feedback presentation groups (chunked vs. separate: Part 1, 0.55 ± 0.20 vs. 0.57 ± 0.19, $F(1, 56) = 0.20$, $p = 0.659$; Part 2, 0.23 ± 0.13 vs. 0.27 ± 0.15, $F(1, 56) = 0.85$, $p = 0.361$) or between knowledge foundation levels (high vs. low: Part 1, 0.52 ± 0.17 vs. 0.60 ± 0.21, $F(1, 56) = 2.09$, $p = 0.154$; Part 2, 0.27 ± 0.14 vs. 0.24 ± 0.14, $F(1, 56) = 0.66$, $p = 0.420$), with no significant interaction effects (Part 1: $F(1, 56) = 1.08$, $p = 0.304$; Part 2: $F(1, 56) = 1.27$, $p = 0.264$).

During the second session, instructors and learners sat face-to-face approximately one meter apart. High-knowledge foundation groups included both an introduction phase and a Q&A feedback phase, while low-knowledge foundation groups included only the Q&A feedback phase. In the introduction phase, instructors presented the 10 concept terms and definitions twice consecutively, with order predetermined by instructors but requiring that chunked concepts not appear consecutively. In the Q&A feedback phase, the separate feedback condition comprised 10 trials, each with three stages: instructor question (stating a concept definition and asking for the term), learner response, and instructor feedback (providing correct term and definition plus one example). Question order was predetermined but differed from the introduction phase (if present), with chunked concepts not appearing consecutively. The chunked feedback condition comprised 5 trials, each with five stages: instructor question 1, learner response 1, instructor question 2, learner response 2, instructor feedback for both concepts (1 and 2 were pre-paired concepts). An example chunked feedback trial:

Instructor: "What is the psychological term for the definition: 'When an event has already occurred, people tend to overestimate their ability to have predicted the outcome'?"
Learner: "Hindsight bias."
Instructor: "Good. Let's look at the next question. What is the psychological term for the definition: 'After an event occurs, people tend to imagine alternatives to reality that could have happened but did not'?"
Learner: "Counterfactual thinking."
Instructor: "These two terms are hindsight bias and counterfactual thinking. Hindsight bias is defined as 'When an event has already occurred, people tend to overestimate their ability to have predicted the outcome.' For example, some students might slap their thighs after the teacher announces the correct answer and say, 'I knew it was that one!' Counterfactual thinking is defined as 'After an event occurs, people tend to imagine alternatives to reality that could have happened but did not.' For example, right after an Olympic competition ends and athletes receive their medals, silver medalists are often less happy than bronze medalists. In interviews, silver medalists sometimes say, 'I almost could have won; that's terrible.' "

The entire experiment was recorded by a digital video camera (HDR-XR100, Sony Corporation, Tokyo, Japan). After the Q&A feedback phase, learners completed a cognitive load assessment scale (Hart, 2006; see Appendix 2) measuring mental, physical, temporal, effort, performance, and frustration dimensions. Learners then completed a post-learning test (15-minute limit) measuring knowledge recognition and transfer. For recognition measurement, learners matched 10 definitions to corresponding terms; for transfer measurement, learners matched 10 new examples to corresponding terms (test format adapted from Finn et al., 2018). Post-test content matched the pre-test. Following previous research measuring long-term learning effects with 7-day intervals (Butler et al., 2007; Smith & Kimball, 2010), learners completed a second post-test online via conferencing software after 7 days (15-minute limit) with identical content to assess long-term effects of feedback presentation format.

2.4 Data Analysis

Because each instructor was randomly paired with 1-2 students, creating a nested structure, all analyses used linear mixed models via the R package lme4 unless otherwise specified. For multiple comparisons, the FDR method was applied (Benjamini & Hochberg, 1995).

The analytical approach included: (1) Confirming the effectiveness of the introduction phase manipulation on students' pre-feedback knowledge foundation; (2) Confirming that feedback-based instructor-learner interaction (regardless of presentation format or knowledge foundation) increased students' conceptual knowledge with long-term retention; (3) Investigating how students' knowledge foundation and feedback presentation format influenced long-term conceptual knowledge gains; (4) Exploring relevant cognitive processes (e.g., promoting error correction, increasing cognitive effort) and conducting mediation analysis using the R package mediation.

2.5 Results

2.5.1 Manipulation Check of Pre-Feedback Knowledge Foundation During the learner response stage of the Q&A feedback phase, learners received 2 points for complete correct terms, 1 point for key terms, and 0 points otherwise, with a maximum of 20 points. Response accuracy quantified pre-feedback knowledge foundation in a linear mixed model with fixed effects of feedback presentation (chunked vs. separate) and knowledge foundation (high vs. low) and random effects of instructor ID, gender, and teaching 次数, plus learner ID and gender. Results showed a significant knowledge foundation effect ($F(1, 54.98) = 194.27$, $p < 0.001$): high-knowledge foundation groups showed significantly higher accuracy ($M = 0.60$, $SD = 0.19$) than low-knowledge groups ($M = 0.10$, $SD = 0.09$). Feedback presentation effect was non-significant ($F(1, 54.19) = 2.45$, $p = 0.123$), as was the interaction ($F(1, 54.41) = 2.28$, $p = 0.137$). These results confirm the effectiveness of the introduction phase manipulation.

2.5.2 Feedback-Based Instructor-Learner Interaction Enhanced Recognition and Transfer with Long-Term Retention Separate linear mixed models were constructed for recognition and transfer test accuracy with fixed effects of test time (pre-learning vs. immediate post-learning vs. 7-day post-learning), feedback presentation, and knowledge foundation, plus random effects of instructor ID, gender, and teaching 次数, and learner ID and gender. Transfer and recognition accuracy were included in respective models to control for potential effects. The primary focus was the test time effect.

For recognition, test time effect was significant ($F(2, 133.94) = 22.17$, $p < 0.001$, Figure 2a): immediate post-learning accuracy ($M = 0.95$, $SD = 0.12$, $t(156) = 6.35$, corrected $p < 0.001$, $\beta = 0.22$, $SE = 0.03$) and 7-day post-learning accuracy ($M = 0.88$, $SD = 0.15$, $t(151) = 5.57$, corrected $p < 0.001$, $\beta = 0.18$, $SE = 0.04$) were both significantly higher than pre-learning accuracy ($M = 0.56$, $SD = 0.20$). Accuracy decay over 7 days was non-significant ($t(114) = 1.65$, corrected $p = 0.31$, $\beta = 0.04$, $SE = 0.02$).

For transfer, test time effect was significant ($F(2, 119.44) = 15.15$, $p < 0.001$, Figure 2b): immediate post-learning accuracy ($M = 0.74$, $SD = 0.25$, $t(141) = 5.19$, corrected $p < 0.001$, $\beta = 0.24$, $SE = 0.05$) and 7-day post-learning accuracy ($M = 0.66$, $SD = 0.26$, $t(135) = 4.89$, corrected $p < 0.001$, $\beta = 0.21$, $SE = 0.04$) were both significantly higher than pre-learning accuracy ($M = 0.25$, $SD = 0.14$). Accuracy decay over 7 days was non-significant ($t(107) = 1.29$, corrected $p = 0.198$, $\beta = 0.04$, $SE = 0.03$).

For recognition, feedback presentation main effect was non-significant ($F(1, 54.76) = 0.02$, $p = 0.890$), knowledge foundation main effect was significant ($F(1, 55.51) = 5.52$, $p = 0.022$), all two-way interactions were non-significant ($Fs < 1.14$, $ps > 0.292$), and the three-way interaction was non-significant ($F(2, 106.69) = 0.96$, $p = 0.385$). For transfer, feedback presentation main effect was non-significant ($F(1, 61.04) = 0.01$, $p = 0.910$), knowledge foundation main effect was non-significant ($F(1, 60.90) = 2.49$, $p = 0.120$), all two-way interactions were non-significant ($Fs < 0.14$, $ps > 0.711$), and the three-way interaction was marginally significant ($F(2, 100.38) = 2.68$, $p = 0.074$).

2.5.3 Chunked Feedback Benefited Long-Term Transfer More Than Separate Feedback for Low-Knowledge Students To control for individual differences in knowledge foundation and examine differential effects of feedback presentation and knowledge foundation across test times, subsequent analyses used linear mixed models for post-learning conceptual knowledge gains (accuracy increment relative to pre-test). For immediate post-learning recognition gain (immediate post-learning minus pre-learning), the model with fixed effects of knowledge foundation and feedback presentation showed non-significant effects of feedback presentation ($F(1, 41.49) = 0.07$, $p = 0.797$), knowledge foundation ($F(1, 41.23) = 1.20$, $p = 0.279$), and their interaction ($F(1, 40.70) = 0.06$, $p = 0.804$). For immediate post-learning transfer gain, the model showed non-significant effects of feedback presentation ($F(1, 39.84) = 1.55$, $p = 0.220$), knowledge foundation ($F(1, 40.66) = 0.07$, $p = 0.786$), and their interaction ($F(1, 39.79) = 0.81$, $p = 0.372$).

For 7-day post-learning recognition and transfer gains, parallel linear mixed models revealed non-significant effects of feedback presentation ($F(1, 39.32) = 0.05$, $p = 0.827$), knowledge foundation ($F(1, 40.82) = 0.02$, $p = 0.890$), and their interaction ($F(1, 40.53) = 0.01$, $p = 0.904$, Figure 3a) for recognition gain. For transfer gain, feedback presentation effect was non-significant ($F(1, 73) = 1.26$, $p = 0.266$), knowledge foundation effect was non-significant ($F(1, 73) = 0.94$, $p = 0.335$), but the interaction was significant ($F(1, 73) = 4.79$, $p = 0.032$, Figure 3b). Simple effects analysis revealed that for low-knowledge students, chunked feedback produced significantly greater 7-day transfer gain (0.49 ± 0.25) than separate feedback (0.31 ± 0.24, $t(34.90) = 2.17$, $p = 0.037$, $\beta = 0.15$, $SE = 0.07$). For high-knowledge students, no significant difference emerged between feedback types (0.43 ± 0.21 vs. 0.49 ± 0.21, $t(37.60) = -0.73$, $p = 0.469$, $\beta = -0.05$, $SE = 0.07$). In the separate feedback condition, high-knowledge students showed marginally greater 7-day transfer gain than low-knowledge students (0.49 ± 0.21 vs. 0.31 ± 0.24, $t(33.50) = 1.99$, $p = 0.054$, $\beta = 0.15$, $SE = 0.07$). In the chunked feedback condition, no significant difference existed between knowledge levels (0.43 ± 0.21 vs. 0.49 ± 0.25, $t(42.20) = -0.82$, $p = 0.417$, $\beta = -0.06$, $SE = 0.07$).

Figure 2. Conceptual knowledge levels
Figure 3. Long-term learning gains

2.5.4 Feedback Presentation Influenced Long-Term Transfer in Low-Knowledge Students Through Chunked Error Correction To explore cognitive processes underlying how feedback presentation influenced low-knowledge students' long-term transfer, we examined chunked error correction and cognitive effort. For each student, chunked error correction was quantified as the number of concept pairs that changed from both incorrect pre-learning to both correct post-learning. For rigorous hypothesis testing, three additional chunked maintenance/change patterns were included: both incorrect → both incorrect, both correct → both correct, and both correct → both incorrect. A linear mixed model was constructed for concept pair maintenance/change patterns on the 7-day transfer test, with fixed effects of maintenance/change pattern (incorrect→correct vs. incorrect→incorrect vs. correct→correct vs. correct→incorrect) and feedback presentation (chunked vs. separate), random effects of instructor ID, gender, and teaching 次数, and learner ID and gender, controlling for initial concept pair correctness on pre-test measure 2 and maintenance/change patterns on 7-day recognition measure. Results showed a non-significant maintenance/change pattern effect ($F(3, 166) = 0.77$, $p = 0.513$) and feedback presentation effect ($F(1, 166) = 0.07$, $p = 0.791$), but a significant interaction ($F(3, 166) = 3.59$, $p = 0.015$). Simple effects analysis revealed that students receiving chunked feedback showed more incorrect→correct concept pairs (1.62 ± 2.49) than those receiving separate feedback (0.95 ± 1.05, $t(133) = 2.47$, $p = 0.015$, $\beta = 0.61$, $SE = 0.25$) and fewer incorrect→incorrect concept pairs (0.69 ± 0.88 vs. 1.14 ± 1.08, $t(127) = -2.01$, $p = 0.047$, $\beta = -0.46$, $SE = 0.23$). Additionally, chunked feedback students showed more incorrect→correct pairs (1.62 ± 2.49) than incorrect→incorrect pairs (0.69 ± 0.88, $t(129) = 2.92$, corrected $p = 0.025$, $\beta = 0.69$, $SE = 0.24$), with no other significant pairwise differences (ts < 2.13, corrected ps > 0.210).

Figure 4. Feedback presentation influences long-term transfer in low-knowledge students through chunked error correction

Furthermore, 7-day transfer gain correlated positively with chunked error correction ($r = 0.89$, $R^2 = 79.21\%$, $p < 0.001$) and negatively with chunked error maintenance ($r = 0.69$, $R^2 = 47.61\%$, $p < 0.001$). Mediation analysis examined whether chunked error correction and maintenance mediated the relationship between feedback presentation and 7-day transfer gain, with chunked feedback coded as 1 and separate feedback as 0. Results showed a marginally significant indirect effect of chunked error correction, $ab = 0.12$, bootstrap 95% CI = [-0.02, 0.24], $p = 0.072$; $c' = 0.07$, $p = 0.063$ (Figure 4b). The indirect effect of chunked error maintenance was non-significant, $ab = 0.07$, bootstrap 95% CI = [-0.03, 0.18], $p = 0.184$; $c' = 0.12$, $p = 0.050$.

Next, a linear mixed model for cognitive effort with fixed effect of feedback presentation (chunked vs. separate) revealed no significant effect (11.52 ± 3.10 vs. 12.70 ± 5.35, $F(1, 18.54) = 0.86$, $p = 0.366$). Additionally, cognitive effort did not correlate significantly with 7-day transfer gain ($r = -0.13$, $R^2 = 1.69\%$, $p = 0.434$). Thus, no further mediation analysis of cognitive effort was conducted.

3 Experiment 2

3.1 Participants

Referencing previous instructor-learner fNIRS hyperscanning studies and using G*Power 3.1 for power analysis (same parameters as Experiment 1), the planned sample size was 68. Experiment 2 recruited 108 East China Normal University students: 40 served as instructors, required to major in psychology or sociology and have completed at least one education course or have teaching experience (M age = 22.75, SD = 2.34, 13 males); 68 served as learners, required to major in non-psychology and non-sociology fields (M age = 21.22, SD = 2.45, 17 males). Fifteen instructors (4 males), 13 instructors (5 males), and 12 instructors (4 males) were randomly assigned to chunked feedback, pseudo-chunked feedback, and separate feedback groups, respectively. Each instructor conducted one-on-one teaching with 1-3 randomly assigned unfamiliar students, with only one experimental session scheduled per day. This yielded 68 instructor-learner dyads: 23 in chunked feedback, 23 in separate feedback, and 22 in pseudo-chunked feedback. All participants had normal or corrected-to-normal vision and no neurological disorders. Each participant read and signed an informed consent form before the experiment. The study was approved by the university's human subjects ethics committee.

3.2 Materials

Test item order was randomized for both immediate and 7-day post-tests to reduce memory effects from fixed item sequences. Other materials remained identical to Experiment 1.

3.3 Procedure

The experimental task occurred in two sessions separated by approximately 5 days (Figure 5). The first session activities matched Experiment 1. Learner pre-test accuracy showed no between-group differences in feedback presentation (chunked vs. pseudo-chunked vs. separate: Part 1, 0.49 ± 0.15 vs. 0.52 ± 0.20 vs. 0.44 ± 0.17, $F(2, 65) = 1.18$, $p = 0.314$; Part 2, 0.23 ± 0.12 vs. 0.25 ± 0.14 vs. 0.26 ± 0.12, $F(2, 65) = 0.32$, $p = 0.731$).

Figure 5. Experiment 2 task and procedure, optode positions, and experimental scenario

The second session consisted of fNIRS hyperscanning and post-scan phases. During hyperscanning, instructors and learners sat face-to-face approximately one meter apart wearing fNIRS equipment. This phase included two stages: rest and Q&A feedback. To ensure low knowledge foundation when receiving feedback, Experiment 2 omitted the introduction phase for all conditions. During the rest stage (300s), participants closed their eyes, cleared their minds, and minimized head and body movements. During the Q&A feedback stage, the separate and chunked feedback procedures matched Experiment 1. In the pseudo-chunked feedback condition, each trial included five stages: instructor question A, learner response A, instructor question B, learner response B, instructor feedback for A and B, where A and B were unrelated concepts. Question order was predetermined with the constraint that chunked concepts did not appear consecutively. The hyperscanning phase was recorded by a digital video camera (HDR-XR100, Sony Corporation, Tokyo, Japan). Post-scan procedures matched Experiment 1.

3.4 fNIRS Data Acquisition and Preprocessing

During hyperscanning, brain activity was simultaneously recorded using fNIRS (ETG-7100, Hitachi Medical Corporation, Japan). Based on previous research (Alamia et al., 2016; Bor et al., 2003; Crone et al., 2008; Jin et al., 2020; Luft, 2014; Moore et al., 2006; Olesen et al., 2004; Pammi et al., 2012; van Duijvenvoorde et al., 2008; Zhu et al., 2022), frontal and parietal regions were selected as regions of interest. Two optode arrays covered these regions: a 3$×5*arrayoverfrontalcortex*(8*emittersand*7*detectorsforming*22*channels*, *numbered*1−22)*anda*3×$3 array over left parietal cortex (5 emitters and 5 detectors forming 12 channels, numbered 23-34), with specific positions shown in Figure 5b. fNIRS channel locations were determined using a 3D virtual positioning system (Singh et al., 2005). Only left parietal cortex was targeted due to left hemisphere language lateralization (Vigneau et al., 2006) and the close relationship between conceptual knowledge learning and language function.

fNIRS recorded optical signals at wavelengths of 695 and 830 nm with a sampling rate of 10 Hz. Data preprocessing used MATLAB (R2018a, MathWorks, Natick, MA, U.S.A) functions and the Homer2 toolbox (v2.2, Huppert et al., 2009). Raw light intensity signals were converted to optical density (OD). Channels with OD signals exceeding 5 standard deviations were excluded. OD signals were inspected and motion artifacts were corrected using a channel-wise wavelet-based method with Daubechies 5 wavelet and adjustment parameter set to 0.1 (Cooper et al., 2012; Molavi & Dumont, 2012). OD signals were band-pass filtered at 0.01-1 Hz to remove low-frequency drift and high-frequency noise. Modified Beer-Lambert law converted OD signals to oxyhemoglobin (HbO) and deoxyhemoglobin concentration changes (Cope & Delpy, 1988). This study focused primarily on HbO concentration changes because they reflect cerebral blood flow changes during brain activity, offer higher signal-to-noise ratio (Hoshi, 2007), and are more widely used in fNIRS hyperscanning studies of social interaction (Cheng et al., 2015; Jiang et al., 2015; Yang et al., 2020).

3.5 Data Analysis

3.5.1 Behavioral Data Analysis Because each instructor was randomly paired with 1-3 students, creating a nested structure, all analyses used linear mixed models via the R package lme4 unless otherwise specified. For multiple comparisons, the FDR method was applied (Benjamini & Hochberg, 1995).

The analytical approach included: (1) Confirming that feedback-based Q&A interaction (regardless of presentation format) increased students' conceptual knowledge with long-term retention; (2) Investigating long-term effects of feedback presentation format on conceptual knowledge gains; (3) Exploring relevant cognitive processes (e.g., promoting error correction or increasing cognitive effort) and conducting mediation analysis using the R package mediation.

3.5.2 fNIRS Data Analysis **Wavelet Transform Coherence (WTC).** Instructor-learner brain-to-brain synchrony was computed using WTC algorithm, yielding frequency- and time-based correlation values between two time series (Grinsted et al., 2004). Preprocessed HbO time series from the same brain location (channel) were extracted from both instructor and learner (e.g., signals i and j from channel 15), then WTC was calculated using the formula:

$$\mathrm{WTC}(t, s) = |S(s^{-1}W\ \ (t, s))|^2 \ /\ (S(s^{-1}|W\ (t, s)|^2)\ \cdot S(s^{-1}|W\ (t, s)|^2))$$

where t represents time points, s represents wavelet scales, $<\cdot>$ denotes smoothing across time and scale, and W represents continuous wavelet transform. This yielded a two-dimensional (time × frequency) WTC matrix. This study included only WTC analysis of corresponding channels.

We focused on feedback-related instructor-learner brain-to-brain synchrony. Based on experimental videos, feedback onset and offset times were marked and adjusted for the 6-second delay-to-peak effect (Cui et al., 2009; Jiang et al., 2015).

Cluster-Permutation Test. For each dyad and channel pair, WTC time series during feedback and rest stages (excluding first and last 30s for stability) were averaged separately and converted to Fisher-z values. Cluster-permutation testing identified frequency-channel clusters comprising at least two adjacent channels and two adjacent frequency points where feedback-stage WTC exceeded rest-stage WTC. This non-parametric method is suitable for multi-channel, multi-frequency neural data, offering greater sensitivity to individual differences than single-channel/frequency multiple comparison corrections (Maris & Oostenveld, 2007). Step 1: Linear mixed models for WTC difference (feedback minus rest) were constructed at each frequency point and channel with fixed effect of presentation format (chunked vs. pseudo-chunked vs. separate) and random effects of instructor ID, gender, and teaching 次数, and learner ID and gender. Step 2: Channels and frequency points (0.01-1 Hz) showing significant presentation format effects with values satisfying chunked > pseudo-chunked, chunked > separate, and chunked > rest were identified. Based on Experiment 1 results showing better transfer under chunked feedback, we focused on these directional effects. Frequency points related to respiration (0.15-0.30 Hz) and heartbeat (>0.70 Hz) were excluded (Nozawa et al., 2016). Step 3: Clusters comprising at least two adjacent channels and frequency points were constructed, with cluster statistics calculated as the sum of F-values within each cluster. Step 4: Data were randomly permuted by pairing each learner's data with a different instructor's data (Jiang et al., 2012; Long et al., 2021). To accommodate different data lengths, longer data were truncated to match shorter data (Reindl et al., 2018). WTC computation and Step 1 analysis were repeated on permuted data 1000 times to generate a distribution of null cluster statistics. Step 5: Real cluster statistics were compared to the permuted distribution (square-root transformed for normalization) to compute p-values (Theiler et al., 1991). Step 6: For significant clusters, average WTC differences were calculated and linear mixed models with fixed effect of presentation format and random effects as above were constructed. If presentation format effects were significant, post-hoc pairwise comparisons were conducted. Clusters satisfying chunked > pseudo-chunked and chunked > separate with FDR-corrected $p < 0.05$ were considered related to chunked feedback processing.

Brain-to-Brain Synchrony Confirmation. Because Step 4 used feedback-rest WTC differences, significant results could theoretically occur if feedback-stage synchrony was lower than rest-stage synchrony. Given that interpersonal interaction should produce greater synchrony than rest (Cui et al., 2012; Jiang et al., 2012), confirmatory analyses were conducted in the chunked feedback group. Step 1: Linear mixed models for average WTC values in each cluster were constructed with fixed effect of task (chunked feedback vs. rest) and random effects as above. Step 2: Chunked feedback WTC values were compared to 1000 permuted null WTC values to compute p-values. Step 3: WTC differences (chunked feedback minus rest) were compared to 1000 permuted null WTC differences to compute p-values. Clusters satisfying significant fixed effects in Step 1 and $p < 0.05$ in Steps 2-3 were considered to reflect authentic chunked feedback-based teaching interaction rather than shared task or environmental factors.

3.5.3 Brain-Behavior Correlation Analysis

Relationships between chunked feedback-related brain-to-brain synchrony and learning performance were analyzed. To control for pre-learning knowledge foundation, relative accuracy (post-test accuracy minus pre-test accuracy) was used to quantify learning performance. Pearson correlations were computed between instructor-learner brain-to-brain synchrony during chunked feedback and students' recognition and transfer accuracy, as well as chunked error correction.

3.6 Results

3.6.1 Pre-Feedback Knowledge Foundation

A linear mixed model for pre-feedback response accuracy with fixed effect of feedback presentation (chunked vs. pseudo-chunked vs. separate) showed no significant effect (0.07 ± 0.08 vs. 0.07 ± 0.08 vs. 0.08 ± 0.10, $F(2, 36.58) = 0.10$, $p = 0.907$), confirming low and equivalent knowledge foundation across groups.

3.6.2 Feedback-Based Instructor-Learner Interaction Enhanced Recognition and Transfer with Long-Term Retention

Linear mixed

models for recognition and transfer accuracy included fixed effects of test time (pre-learning vs. immediate post-learning vs. 7-day post-learning) and feedback presentation, with random effects as above, controlling for respective transfer/recognition accuracy.

For recognition, test time effect was significant ($F(2, 156.20) = 22.50$, $p < 0.001$, Figure 6a): immediate post-learning accuracy ($M = 0.91$, $SD = 0.13$, $t(185) = 6.43$, corrected $p < 0.001$, $\beta = 0.22$, $SE = 0.03$) and 7-day post-learning accuracy ($M = 0.81$, $SD = 0.20$, $t(183) = 4.92$, corrected $p < 0.001$, $\beta = 0.15$, $SE = 0.03$) were both significantly higher than pre-learning accuracy ($M = 0.48$, $SD = 0.17$). Accuracy significantly decayed over 7 days ($t(139) = -2.94$, corrected $p = 0.011$, $\beta = -0.06$, $SE = 0.02$).

For transfer, test time effect was significant ($F(2, 153.14) = 18.55$, $p < 0.001$, Figure 6b): immediate post-learning accuracy ($M = 0.71$, $SD = 0.23$, $t(185) = 5.34$, corrected $p < 0.001$, $\beta = 0.21$, $SE = 0.04$) and 7-day post-learning accuracy ($M = 0.64$, $SD = 0.25$, $t(177) = 5.84$, corrected $p < 0.001$, $\beta = 0.20$, $SE = 0.03$) were both significantly higher than pre-learning accuracy ($M = 0.25$, $SD = 0.13$). Accuracy did not significantly decay over 7 days ($t(141) = -0.44$, corrected $p = 0.661$, $\beta = -0.01$, $SE = 0.03$).

For recognition, feedback presentation main effect was non-significant ($F(2, 24.43) = 2.20$, $p = 0.132$), and test time × feedback presentation interaction was non-significant ($F(4, 135.30) = 0.44$, $p = 0.778$). For transfer, feedback presentation main effect was non-significant ($F(2, 27.43) = 1.74$, $p = 0.195$), while the interaction was marginally significant ($F(4, 130.73) = 2.22$, $p = 0.070$).

Figure 6. Conceptual knowledge levels

3.6.3 Chunked Feedback Benefited Long-Term Transfer More Than Non-Chunked Feedback To control for individual knowledge foundation differences and examine differential effects of feedback presentation across test times, subsequent analyses used linear mixed models for post-learning conceptual knowledge gains (accuracy increment relative to pre-test). For immediate post-learning recognition gain, the model with fixed effect of feedback presentation (chunked vs. pseudo-chunked vs. separate) showed no significant effect (0.44 ± 0.19 vs. 0.38 ± 0.21 vs. 0.45 ± 0.16, $F(2, 38.57) = 0.74$, $p = 0.483$). For immediate post-learning transfer gain, feedback presentation effect was marginally significant (0.53 ± 0.16 vs. 0.41 ± 0.20 vs. 0.44 ± 0.23, $F(2, 62.67) = 2.64$, $p = 0.079$).

For 7-day post-learning recognition gain, feedback presentation effect was non-significant (0.37 ± 0.21 vs. 0.30 ± 0.19 vs. 0.34 ± 0.21, $F(2, 30.95) = 0.20$, $p = 0.823$, Figure 7a). For 7-day post-learning transfer gain, feedback presentation effect was significant (0.53 ± 0.17 vs. 0.30 ± 0.25 vs. 0.37 ± 0.22, $F(2, 61.70) = 6.11$, $p = 0.004$, Figure 7b): chunked feedback produced significantly greater 7-day transfer gain than separate feedback ($t(29.9) = 2.49$, $p = 0.019$, corrected $p = 0.028$, $\beta = 0.15$, $SE = 0.06$) and pseudo-chunked feedback ($t(32) = 3.20$,

$p = 0.0031$, corrected $p = 0.009$, $\beta = 0.20$, SE = 0.06), while pseudo-chunked and separate feedback did not differ ($t(26.7) = -0.76$, $p = 0.455$, corrected $p = 0.455$, $\beta = -0.05$, SE = 0.06).

Figure 7. Long-term learning gains

3.6.4 Feedback Presentation Influenced Long-Term Transfer Through Chunked Error Correction To explore cognitive processes underlying feedback presentation effects on long-term transfer, we examined chunked error correction and cognitive effort. Given that pseudo-chunked and separate feedback showed no significant difference in long-term transfer, these were combined into a non-chunked feedback group for comparison with the chunked feedback group. A linear mixed model was constructed for maintenance/change patterns on 7-day transfer measure with fixed effects of maintenance/change (incorrect→correct vs. incorrect→incorrect vs. correct→correct vs. correct→incorrect) and feedback presentation (chunked vs. non-chunked), random effects as above, controlling for initial concept pair correctness and recognition maintenance/change patterns. Results (Figure 8a) showed a significant change pattern effect ($F(3, 261.15) = 9.37$, $p < 0.001$): incorrect→correct concept pairs (1.26 ± 0.26) were more numerous than incorrect→incorrect pairs (0.78 ± 0.83, $t(197) = 3.71$, corrected $p = 0.002$, $\beta = 0.42$, SE = 0.11), with no other significant comparisons (ts < 2.16, corrected ps > 0.096). Feedback presentation effect was non-significant ($F(1, 249.65) = 0.36$, $p = 0.549$). The change pattern × presentation interaction was significant ($F(3, 260.06) = 8.23$, $p < 0.001$). Simple effects analysis revealed that chunked feedback students showed more incorrect→correct concept pairs (1.74 ± 1.05) than non-chunked feedback students (1.02 ± 0.99, $t(214) = 4.13$, $p < 0.001$, $\beta = 0.69$, SE = 0.17) and fewer incorrect→incorrect pairs (0.48 ± 0.73 vs. 0.93 ± 0.84, $t(213) = -2.71$, $p = 0.007$, $\beta = -0.45$, SE = 0.17). Additionally, chunked feedback students showed more incorrect→correct pairs (1.74 ± 1.05) than correct→incorrect pairs (0.00 ± 0.00, $t(240) = 3.66$, corrected $p < 0.001$, $\beta = 0.87$, SE = 0.24), correct→correct pairs (0.22 ± 0.42, $t(241) = 3.16$, corrected $p = 0.004$, $\beta = 0.75$, SE = 0.24), and incorrect→incorrect pairs (0.48 ± 0.73, $t(198) = 6.15$, corrected $p < 0.001$, $\beta = 1.19$, SE = 0.19). No other comparisons were significant (ts < 1.36, corrected ps > 0.209).

Figure 8. Feedback presentation influences long-term transfer through chunked error correction

Furthermore, 7-day transfer gain correlated positively with chunked error correction ($r = 0.81$, $R^2 = 65.61\%$, $p < 0.001$) and negatively with chunked error maintenance ($r = -0.52$, $R^2 = 27.04\%$, $p < 0.001$). Mediation analysis examined whether chunked error correction and maintenance mediated the relationship between feedback presentation and 7-day transfer gain, with chunked feedback coded as 1 and non-chunked as 0. Results showed significant partial mediation by chunked error correction, ab = 0.12, bootstrap 95% CI = [0.03, 0.21], $p = 0.008$; c' = 0.07, $p = 0.046$ (Figure 8b). Chunked error maintenance also showed significant partial mediation, ab = 0.06, bootstrap 95% CI = [0.01, 0.11], $p = 0.024$; c' = 0.14, $p = 0.009$.

A linear mixed model for cognitive effort with fixed effect of feedback presentation (chunked vs. non-chunked) revealed that chunked feedback students reported higher cognitive effort (12.19 ± 4.84) than non-chunked feedback students (10.58 ± 3.80, $F(1, 61.08) = 4.36$, $p = 0.041$). Cognitive effort correlated significantly with the frustration dimension of cognitive load ($r = 0.32$, $R^2 = 10.24\%$, $p = 0.011$), and chunked feedback students reported greater learning frustration (7.24 ± 4.53) than non-chunked feedback students (5.45 ± 4.92, $F(1, 32.94) = 4.32$, $p = 0.045$). However, cognitive effort did not correlate significantly with 7-day transfer gain ($r = -0.05$, $R^2 = 0.25\%$, $p = 0.700$), so no further mediation analysis was conducted.

3.6.5 Chunked Feedback Elicited Stronger Instructor-Learner Brain-to-Brain Synchrony in Frontal and Parietal Regions Analysis examined whether chunked feedback elicited significantly greater instructor-learner brain-to-brain synchrony than non-chunked feedback (including pseudo-chunked and separate). Three significant frequency-channel clusters were identified: two in frontal cortex and one in left parietal cortex.

Cluster 1 included channels 21 and 22, located in right superior and middle frontal gyri, frequency 0.019-0.028 Hz (Figure 9a). Cluster statistic = 5.16, permutation test $p < 0.001$ (Figure 9b). Presentation format effect on Cluster 1 synchrony increase (feedback minus rest) was significant ($F(2, 65) = 3.83$, $p = 0.027$, Figure 9c). Post-hoc comparisons revealed that chunked feedback synchrony increase (0.07 ± 0.16) was significantly greater than pseudo-chunked feedback (-0.05 ± 0.16, $t(32.6) = 2.32$, $p = 0.027$, corrected $p = 0.045$, $\beta = 0.12$, SE = 0.05) and separate feedback (-0.05 ± 0.17, $t(31.1) = 2.28$, $p = 0.030$, corrected $p = 0.045$, $\beta = 0.12$, SE = 0.05), with no difference between pseudo-chunked and separate feedback ($t(27.1) = -0.04$, $p = 0.966$, corrected $p = 0.966$, $\beta = -0.00$, SE = 0.05). Confirmatory analyses verified that chunked feedback synchrony (0.40 ± 0.10) was significantly greater than rest synchrony (0.33 ± 0.13, $F(1, 29.52) = 5.00$, $p = 0.033$, Figure 9d), greater than permuted null synchrony values ($p = 0.022$, Figure 9e), and the synchrony increase was greater than permuted null increases ($p = 0.032$, Figure 9f).

Cluster 2 included channels 7, 11, 12, and 17, located in bilateral superior frontal gyri, frequency 0.010-0.015 Hz (Figure 9g). Cluster statistic = 7.62, permutation test $p < 0.001$ (Figure 9h). Presentation format effect on Cluster 2 synchrony increase was significant ($F(2, 62.51) = 8.05$, $p < 0.001$, Figure 9i). Chunked feedback synchrony increase (0.11 ± 0.14) was significantly greater than pseudo-chunked feedback (-0.07 ± 0.20, $t(31.5) = 2.79$, $p = 0.009$, corrected $p = 0.013$, $\beta = 0.18$, SE = 0.06) and separate feedback (-0.14 ± 0.28, $t(30.2) = 3.76$, $p < 0.001$, corrected $p = 0.002$, $\beta = 0.24$, SE = 0.06), with no difference between pseudo-chunked and separate feedback ($t(26.0) = 0.98$, $p = 0.337$, corrected $p = 0.337$, $\beta = 0.06$, SE = 0.06). Confirmatory analyses verified that chunked feedback synchrony (0.48 ± 0.12) was significantly greater than rest synchrony (0.37 ± 0.12, $F(1, 22) = 15.47$, $p < 0.001$, Figure 9j), greater than permuted null synchrony values ($p = 0.008$, Figure 9k), and the synchrony increase was greater than permuted null increases ($p < 0.001$, Figure 9l).

Cluster 3 included channels 25 and 28, located in left inferior parietal lobule, frequency 0.027-0.034 Hz (Figure 9m). Cluster statistic = 4.72, permutation test $p < 0.001$ (Figure 9n). Presentation format effect on Cluster 3 synchrony increase was significant ($F(2, 36.43) = 6.29$, $p = 0.005$, Figure 9o). Chunked feedback synchrony increase (0.08 ± 0.11) was significantly greater than pseudo-chunked feedback (-0.03 ± 0.12, $t(36.5) = 2.88$, $p = 0.007$, corrected $p = 0.010$, $\beta = 0.12$, SE = 0.04) and separate feedback (-0.04 ± 0.13, $t(34.5) = 3.04$, $p = 0.005$, corrected $p = 0.010$, $\beta = 0.13$, SE = 0.04), with no difference between pseudo-chunked and separate feedback ($t(32.6) = 0.20$, $p = 0.842$, corrected $p = 0.842$, $\beta = 0.01$, SE = 0.04). Confirmatory analyses verified that chunked feedback synchrony (0.37 ± 0.08) was significantly greater than rest synchrony (0.29 ± 0.10, $F(1, 22) = 12.87$, $p = 0.002$, Figure 9p), greater than permuted null synchrony values ($p = 0.006$, Figure 9q), and the synchrony increase was greater than permuted null increases ($p < 0.001$, Figure 9r).

Figure 9. Stronger instructor-learner brain-to-brain synchrony in frontal and parietal regions during chunked feedback

3.6.6 Frontal Brain-to-Brain Synchrony During Chunked Feedback Predicted Long-Term Transfer and Chunked Error Correction We examined whether instructor-learner brain-to-brain synchrony during chunked feedback predicted students' long-term learning performance. Cluster 1 synchrony difference (chunked feedback minus rest) correlated marginally positively with immediate post-learning recognition gain ($r = 0.37$, $R^2 = 13.69\%$, $p = 0.082$) and 7-day post-learning recognition gain ($r = 0.38$, $R^2 = 14.44\%$, $p = 0.070$), and correlated significantly positively with immediate post-learning transfer gain ($r = 0.64$, $R^2 = 40.96\%$, $p = 0.001$) and 7-day post-learning transfer gain ($r = 0.55$, $R^2 = 30.25\%$, $p = 0.006$, Figure 10a). No other clusters correlated significantly with learning performance (rs < 0.37, ps > 0.078).

Additionally, Cluster 1 synchrony difference correlated positively with immediate post-learning error correction concept pairs ($r = 0.69$, $R^2 = 47.61\%$, $p < 0.001$), marginally negatively with immediate post-learning error maintenance pairs ($r = -0.24$, $R^2 = 5.76\%$, $p = 0.052$), positively with 7-day post-learning error correction pairs ($r = 0.54$, $R^2 = 29.16\%$, $p = 0.007$, Figure 10b), and non-significantly with immediate error maintenance pairs ($r = -0.14$, $R^2 = 1.96\%$, $p = 0.260$).

Figure 10. Frontal brain-to-brain synchrony during chunked feedback predicts long-term transfer and chunked error correction

4 General Discussion

Experiment 1 compared chunked versus separate feedback in instructor-learner interaction, revealing long-term effects of feedback presentation format: chunked feedback was more beneficial than separate feedback for students' long-term transfer, but only for those with lower prior knowledge. To exclude an alternative explanation that delayed feedback timing alone produced these benefits, Experiment 2 added a pseudo-chunked feedback (non-chunked but delayed) condition. Results showed chunked feedback produced significantly better long-term transfer than both non-chunked conditions, while pseudo-chunked and separate feedback did not differ, ruling out pure delay effects. These findings align with previous research showing that emphasizing relationships among learning content or presenting materials in chunks promotes knowledge acquisition and transfer to new situations (Cohen & Sekuler, 2010; Spiro et al., 1991). Notably, chunked feedback promoted knowledge transfer rather than recognition, the transfer benefits were long-term rather than immediate, and effects occurred only in low-knowledge students.

First, different presentation formats of elaborated content feedback affected only transfer, not recognition. Previous research found elaborated feedback promoted both recognition and transfer (Finn et al., 2018) or only transfer (Butler et al., 2013), possibly depending on test type or difficulty. Notably, recognition performance showed ceiling effects, likely reflecting low measurement difficulty or masking condition differences. Future research could employ cued recall instead of associative recognition to increase difficulty and examine feedback presentation effects on recognition. Additionally, ceiling effects and lack of condition differences in recognition help rule out alternative explanations that feedback presentation effects on transfer were driven by recognition failures or differences.

Second, long-term effects likely emerged because chunked feedback promoted deeper cognitive processing of conceptual knowledge, thereby slowing transfer decay over time. According to Chunking Theory (Chase & Simon, 1973) and Template Theory (Gobet & Simon, 1996), learning occurs through the development of discrimination networks that are continuously influenced by the system's current state and environmental input. Theoretically, chunked environmental input helps discrimination networks develop higher-level structures that likely index procedural or semantic information in long-term memory. In this study, chunked versus separate feedback input helped low-level learners better discriminate similarities and differences between related concepts, promoting advanced network structures. These higher-level network structures created more durable semantic memory indexes that supported long-term retrieval and application of conceptual knowledge in novel contexts, producing long-term transfer gains.

Third, the restriction of chunked feedback's long-term transfer effects to low-knowledge students can be explained by the Expertise Reversal Effect (Kalyuga & Sweller, 2004). For novices lacking relevant background knowledge, novel tasks easily overload working memory, making their performance more dependent on external supportive guidance. In contrast, more experienced learners can draw on existing schemas to complete tasks without working memory overload, rendering external guidance redundant (Renkl & Atkinson, 2003; Sweller et al., 1998). In this study, chunked feedback served as supportive external guidance that more effectively helped low-prior-knowledge students process feedback, correct errors, and promote transfer to new contexts. For high-knowledge learners, chunked feedback may have been redundant, providing no additional learning gains.

Furthermore, Experiment 1 manipulated pre-feedback knowledge foundation through the presence or absence of an introduction phase, yet found no significant differences between knowledge levels in recognition or transfer performance. This suggests the introduction phase may have provided redundant instructional support, while Q&A feedback without introduction actually improved learning efficiency—achieving equivalent outcomes with streamlined instruction and reduced time. However, the introduction phase involved presenting concept terms and definitions twice consecutively, which may have excessively elevated knowledge foundation, obscuring potential feedback presentation effects. Future research could revise the introduction phase to produce smaller knowledge gains, such as reducing presentation repetitions from two to one or inserting longer intervals (e.g., one day) between introduction and Q&A phases, to further examine whether feedback presentation affects learning in relatively higher-knowledge students.

4.2 Feedback Presentation Promotes Long-Term Transfer in Low-Knowledge Students Through Chunked Error Correction

This study found an indirect pathway from feedback presentation to long-term transfer in low-knowledge students through whole-concept-pair error correction, highlighting the importance of chunked error correction. This result aligns with cognitivist learning theory's proposal that specific, targeted, or information-rich feedback promotes learning through more effective error correction (Kulhavy & Stock, 1989; Narciss & Huth, 2004; Bangert-Drowns et al., 1991), while extending it by demonstrating that organizing feedback content through presentation format promotes organized correction of conceptual knowledge in low-prior-knowledge students, yielding long-term transfer gains. Integrating Chunking and Template Theories, chunked versus separate feedback input likely helped low-level learners more effectively identify and correct misunderstandings or errors in conceptual knowledge, enabling discrimination networks to develop advanced structures. These advanced structures created more durable semantic memory indexes that supported effective retrieval and application of conceptual knowledge in novel contexts after extended time intervals.

Unlike Experiment 1, Experiment 2 found that chunked feedback students reported greater cognitive effort and learning frustration than non-chunked feedback students. This may occur because chunked feedback more clearly identifies gaps between current performance and goals, increasing learning frustration and stimulating cognitive effort. This finding partially aligns with metacognitivist theory, which posits that feedback promotes learning only when it effectively indicates performance-goal gaps that motivate learners to invest cognitive effort to reduce them (Nicol & McFarlane-Dick, 2006; Sadler, 1989). However, no relationship was found between cognitive effort and long-term transfer, suggesting feedback presentation cannot achieve long-term transfer gains through cognitive effort alone. Although this study does not support cognitive effort as the mechanism, more precise measurement of cognitive effort (Laufer & Hulstijn, 2001; Golonka et al., 2015) is needed in future research.

4.3 Interpersonal Neural Basis of Chunked Feedback in Instructor-Learner Interaction

Experiment 2 used fNIRS hyperscanning to investigate the interpersonal neural basis of chunked feedback during instructor-learner interaction. Results showed that providing and receiving chunked feedback elicited greater brain-to-brain synchrony between instructors and learners in frontal and parietal regions, including superior frontal gyrus, middle frontal gyrus, and inferior parietal lobule. These regions are spatially close to those previously implicated in feedback processing (ACC, DLPFC, and parietal cortex; Cavanagh et al., 2012; Crone et al., 2008; Luft et al., 2013; Mars et al., 2005; van Duijvenvoorde et al., 2008; Zanolie et al., 2008) and chunk processing (DLPFC and posterior parietal cortex; Alamia et al., 2016; Bor et al., 2003; Jin et al., 2020; Pammi et al., 2012). They also align with regions showing brain-to-brain synchrony during general or elaborated feedback-based instructor-learner interaction (frontal or temporoparietal regions; Pan et al., 2020; Zheng et al., 2018; Zhu et al., 2022). This suggests that frontal and parietal brain-to-brain synchrony supports feedback processing during instructor-learner interaction and is sensitive to changes in feedback presentation format.

Furthermore, Experiment 2 found that frontal (but not parietal) brain-to-brain synchrony related to chunked feedback positively correlated with learners' long-term transfer performance and error correction. This suggests frontal brain-to-brain synchrony may constitute the interpersonal neural basis through which chunked feedback promotes error correction and facilitates long-term deep learning like transfer. Previous research has shown that chunking linguistic information depends on frontal cortex (Jin et al., 2020; Grodzinsky & Santi, 2008), frontal activity relates to long-term memory retention and delayed task performance (Sakai & Passingham, 2003; Squire et al., 1993), and abstract knowledge structures or schemas relate to medial prefrontal function (Gilboa & Marlatte, 2017). Thus, our findings support a critical role for frontal cortex in deep knowledge representation, possibly supporting the development of higher-level, more abstract knowledge structures in discrimination networks that promote long-term transfer gains. Additionally, previous research found that frontal brain-to-brain synchrony indicates effective teaching strategies such as scaffolding and high-frequency interaction (Pan et al., 2018, 2020), and chunked presentation represents an effective instructional strategy (Spiro et al., 1991; Cohen & Sekuler, 2010). Our results thus support frontal brain-to-brain synchrony's role in distinguishing teaching strategy effectiveness. However, previous research also indicates that adults rely more on parietal cortex than ACC for processing effective feedback to adjust performance or correct errors (Crone et al., 2008; van Duijvenvoorde et al., 2008; Zanolie et al., 2008), that parietal brain-to-brain synchrony during separate feedback predicts immediate transfer (Zhu et al., 2022), and that temporoparietal synchrony relates to interactive teaching strategy selection (Zheng et al., 2018). Future research should further investigate distinct roles and relationships of frontal versus parietal brain-to-brain synchrony in supporting feedback processing, effective teaching, and promoting immediate versus long-term deep learning.

Notably, frontal brain-to-brain synchrony correlated positively with chunked error correction, which reflects students' understanding gradually aligning with instructors' understanding during feedback-based teaching interaction, indicating enhanced cognitive alignment. This supports brain-to-brain synchrony as a neural basis for cognitive alignment and mutual understanding during interaction (Shamay-Tsoory et al., 2019; Kelsen et al., 2022). In contrast, single-brain metrics reflect individual information processing and have limited power to reveal neural mechanisms underlying complex, continuous, naturalistic stimuli without predefined events (Hasson et al., 2004). Authentic classroom teaching involves dynamic, continuous, bidirectional information exchange, making instructor-learner brain-to-brain synchrony more effective than single-brain metrics for reflecting whether mutual understanding is achieved—an essential foundation for effective teaching (Tan et al., 2023). Additionally, brain-to-brain synchrony can reflect the dynamic process of achieving alignment, including time points of understanding and temporal lag patterns (Liu et al., 2018; Zheng et al., 2018; Zhu et al., 2022). Thus, instructor-learner brain-to-brain synchrony can provide timely, objective evidence for adjusting and optimizing classroom instruction; for example, persistently low synchrony may indicate student comprehension problems preventing shared understanding.

4.4 Limitations and Future Directions

Several issues warrant attention. First, feedback's metacognitive effects represent another important aspect of feedback's impact on learning. Research shows feedback helps correct high-confidence errors (Butterfield & Metcalfe, 2001) and calibrate metacognitive errors for low-confidence correct answers (Butler et al., 2008). Thus, chunked feedback may trigger deeper conceptual understanding that updates or revises initial metacognitive assessments. Future research could measure students' confidence in their answers to examine metacognitive effects of feedback presentation format.

Second, social factors play crucial roles in feedback-based social interaction, including interaction partners (e.g., instructor-learner, peer-peer) and interpersonal relationships (e.g., trust, rapport). Future research should investigate these social factors and reveal broader interpersonal neural mechanisms of feedback in social interaction.

Third, limited fNIRS channels restricted brain activity recording to frontal and left parietal regions, leaving other brain areas unexplored. Given the asymmetrical roles of instructors and learners in feedback-based interaction, future research should examine cross-brain, temporally-lagged interpersonal neural synchrony between different brain regions in both parties.

Fourth, instructor-learner brain-to-brain synchrony related to chunked feedback occurred primarily in the 0.01-0.03 Hz frequency range. While this overlaps with frequencies found in previous fNIRS hyperscanning studies using communication paradigms (Jiang et al., 2012, 2015) and teaching tasks (Zheng et al., 2018; Zhu et al., 2022), the functional significance of brain-to-brain synchrony in the frequency domain requires further investigation.

Fifth, chunked feedback in this study simultaneously introduced feedback delay. Although such delayed feedback better promoted transfer, previous research found learners subjectively prefer more immediate feedback (Lefevre & Cox, 2017; Mullet et al., 2014). Whether learners also prefer immediate separate feedback over chunked feedback, and the underlying cognitive-neural mechanisms, warrant future investigation.

Finally, although both experiments attempted to include diverse gender participants with some opposite-sex dyads, female participants and female-female dyads outnumbered males. Therefore, caution is needed when generalizing findings across genders.

5 Conclusion

This study conducted two dyadic experiments (behavioral, fNIRS hyperscanning) to investigate how maintaining identical feedback content while altering presentation format influences long-term learning transfer, cognitive processes, and interpersonal neural basis during authentic instructor-learner interaction. Conclusions are: (1) Chunked feedback promoted long-term transfer in low-knowledge students, excluding the possibility that these gains resulted simply from altered feedback timing; (2) Feedback presentation format influenced long-term transfer through chunked error correction; (3) Providing and receiving chunked feedback elicited greater brain-to-brain synchrony in frontal and parietal regions between instructors and learners; (4) Frontal brain-to-brain synchrony during chunked feedback predicted long-term transfer and chunked error correction.

References

Alamia, A., Solopchuk, O., D'Ausilio, A., Van Bever, V., Fadiga, L., Olivier, E., & Zénon, A. (2016). Disruption of Broca's area alters higher-order chunking processing during perceptual sequence learning. *Journal of Cognitive Neuroscience, 28*(3), 402–417.

Anderson, D. I., Magill, R. A., & Sekiya, H. (2001). Motor learning as a function of KR schedule and characteristics of task-intrinsic feedback. *Journal of Motor Behavior, 33*(1), 59–66.

Apps, M. A., Lesage, E., & Ramnani, N. (2015). Vicarious reinforcement learning signals when instructing others. *Journal of Neuroscience, 35*(7), 2904–2913.

Arbel, Y., Goforth, K., & Donchin, E. (2013). The good, the bad, or the useful? The examination of the relationship between the feedback-related negativity (FRN) and long-term learning outcomes. *Journal of Cognitive Neuroscience, 25*(8), 1249–1260.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 218–238.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289–300.

Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., Poeppel, D., & Dikker, S. (2019). Brain-to-brain synchrony and learning outcomes vary by student–teacher dynamics: Evidence from a real-world classroom electroencephalography study. *Journal of Cognitive Neuroscience, 31*(3), 401–411.

Bor, D., Duncan, J., Wiseman, R. J., & Owen, A. M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron, 37*(2), 361–367.

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology, 105*(2), 290–298.

Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13*, 273–281.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 918–928.

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1491–1494.

Cavanagh, J. F., Figueroa, C. M., Cohen, M. X., & Frank, M. J. (2012). Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cerebral Cortex, 22*(11), 2575–2586.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–81.

Cheng, X., Li, X., & Hu, Y. (2015). Synchronous brain activity during cooperative exchange depends on gender of partner: A fNIRS-based hyperscanning study. *Human Brain Mapping, 36*(6), 2039–2048.

Cohen, N. R., & Sekuler, R. (2010). Chunking and compound cueing of movement sequences: Learning, retention, and transfer. *Perceptual and Motor Skills, 110*(3), 736–750.

Cooper, R. J., Selb, J., Gagnon, L., Phillip, D., Schytz, H. W., Iversen, H. K., Ashina, M., & Boas, D. A. (2012). A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy. *Frontiers in Neuroscience, 6*, 147.

Cope, M., & Delpy, D. T. (1988). System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infrared transillumination. *Medical & Biological Engineering & Computing, 26*(3), 289–294.

Crone, E. A., Zanolie, K., Van Leijenhorst, L., Westenberg, P. M., & Rombouts, S. A. (2008). Neural mechanisms supporting flexible performance adjustment during development. *Cognitive, Affective & Behavioral Neuroscience, 8*(2), 165–177.

Cui, X., Bryant, D. M., & Reiss, A. L. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *NeuroImage, 59*(3), 2430–2437.

Cui, X., Stetson, C., Montague, P. R., & Eagleman, D. M. (2009). Ready…go: Amplitude of the FMRI signal encodes expectation of cue arrival time. *PLoS Biology, 7*(8), e1000167.

Finn, B., Thomas, R., & Rawson, K. A. (2018). Learning more from feedback: Elaborating feedback with examples enhances concept learning. *Learning and Instruction, 54*, 104–113.

Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology, 104*, 1094–1108.

Gick, M. L., & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In *Transfer of Learning* (pp. 9-46). Academic Press.

Gilboa, A., & Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends in Cognitive Sciences, 21*(8), 618–631.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive*

Sciences, 5(6), 236–243.

Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology, 31*(1), 1–40.

Golonka, E., Bowles, A., Silbert, N., Kramasz, D., Blake, C., & Buckwalter, T. (2015). The role of context and cognitive effort in vocabulary learning: A study of intermediate-level learners of Arabic. *The Modern Language Journal, 99*(1), 19–39.

Grinsted, A., Moore, J. C., & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics, 11*(5/6), 561–566.

Grodzinsky, Y., & Santi, A. (2008). The battle for Broca's region. *Trends in Cognitive Sciences, 12*, 474–480.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904–908.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences, 16*(2), 114–121.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science, 303*(5664), 1634–1640.

Holper, L., Goldin, A. P., Shalóm, D. E., Battro, A. M., Wolf, M., & Sigman, M. (2013). The teaching and the learning brain: A cortical hemodynamic marker of teacher–student interactions in the Socratic dialog. *International Journal of Educational Research, 59*, 1–10.

Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). HomER: A review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics, 48*(10), D280–D298.

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience, 32*(45), 16064–16069.

Jiang, J., Chen, C., Dai, B., Shi, G., Ding, G., Liu, L., & Lu, C. (2015). Leader emergence through interpersonal neural synchronization. *Proceedings of the National Academy of Sciences of the United States of America, 112*(14), 4274–4279.

Jin, P., Lu, Y., & Ding, N. (2020). Low-frequency neural activity reflects rule-based chunking during speech listening. *Elife, 9*, e55613.

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96*(3), 558–568.

Kelsen, B. A., Sumich, A., Kasabov, N., Liang, S. H., & Wang, G. Y. (2022). What has social neuroscience learned from hyperscanning studies of spoken communication? A systematic review. *Neuroscience & Biobehavioral Reviews, 132*, 1249–1262.

Krause, U. M., Stark, R., & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction, 19*(2), 158–170.

Krijgsman, C., Mainhard, T., van Tartwijk, J., Borghouts, L., Vansteenkiste, M., Aelterman, N., & Haerens, L. (2019). Where to go and how to get there: Goal clarification, process feedback and students' need satisfaction and frustration from lesson to lesson. *Learning and Instruction, 61*, 1–11.

Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology, 63*(5), 505–512.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*(4), 279–308.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1–26.

Lefevre, D., & Cox, B. (2017). Delayed instructional feedback may be more effective, but is this contrary to learners' preferences? *British Journal of Educational Technology, 48*(6), 1357–1367.

Liu, J., Zhang, R., Geng, B., Zhang, T., Yuan, D., Otani, S., & Li, X. (2019). Interplay between prior knowledge and communication mode on teaching effectiveness: Interpersonal neural synchronization as a neural marker. *NeuroImage, 193*, 93–102.

Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy. *Neuroscience & Biobehavioral Reviews, 34*(3), 269–284.

Long, Y., Zheng, L., Zhao, H., Zhou, S., Zhai, Y., & Lu, C. (2021). Interpersonal neural synchronization during interpersonal touch underlies affiliative pair bonding between romantic couples. *Cerebral Cortex, 31*(3), 1647–1659.

Luft, C. D. B., Nolte, G., & Bhattacharya, J. (2013). High-learners present larger mid-frontal theta power and connectivity in response to incorrect performance feedback. *The Journal of Neuroscience, 33*(5), 2029–2038.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190.

Molavi, B., & Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiological Measurement, 33*(2), 259–270.

Mullet, H. G., Butler, A. C., Verdin, B., von Borries, R., & Marsh, E. J. (2014). Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately. *Journal of Applied Research in Memory and Cognition, 3*(3), 222–229.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, D. Leutner, & R. Brunken (Eds.), *Instructional Design for Multimedia Learning* (pp. 181–195). Munster, NY: Waxmann.

Nicol, D. J., & McFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218.

Nguyen, M., Chang, A., Micciche, E., Meshulam, M., Nastase, S. A., & Hasson, U. (2022). Teacher–student neural coupling during teaching and learning. *Social Cognitive and Affective Neuroscience, 17*(4), 367-376.

Nozawa, T., Sasaki, Y., Sakaki, K., Yokoyama, R., & Kawashima, R. (2016). Interpersonal frontopolar neural synchronization in group communication: An exploration toward fNIRS hyperscanning of natural interactions. *NeuroImage, 133*, 484–497.

Pammi, V. C., Miyapuram, K. P., Samejima, K., Bapi, R. S., & Doya, K. (2012). Changing the structure of complex visuo-motor sequences selectively activates the fronto-parietal network. *Neuroimage, 59*(2), 1180–1189.

Pan, Y., Novembre, G., Song, B., Li, X., & Hu, Y. (2018). Interpersonal synchronization of inferior frontal cortices tracks social interactive learning of a song. *NeuroImage, 183*, 280–290.

Pan, Y., Dikker, S., Goldstein, P., Zhu, Y., Yang, C., & Hu, Y. (2020). Instructor-learner brain coupling discriminates between instructional approaches and predicts learning. *NeuroImage, 211*, 116657.

Peters, S., Van der Meulen, M., Zanolie, K., & Crone, E. A. (2017). Predicting reading and mathematics from neural activity for feedback learning. *Developmental Psychology, 53*(1), 149–159.

Rawson, K. A., Thomas, R. C., & Jacoby, L. L. (2015). The power of examples: Illustrative examples enhance conceptual learning of declarative concepts. *Educational Psychology Review, 27*(3), 483–504.

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience, 20*(8), 495–505.

Reindl, V., Gerloff, C., Scharke, W., & Konrad, K. (2018). Brain-to-brain synchrony in parent-child dyads and the relationship with emotion regulation revealed by fNIRS-based hyperscanning. *NeuroImage, 178*, 493–502.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*, 15–22.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144.

Sakai, K., & Passingham, R. E. (2003). Prefrontal interactions reflect future task operations. *Nature Neuroscience, 6*(1), 75–81.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences, 36*(4), 393–414.

Shamay-Tsoory, S. G., Saporta, N., Marton-Alper, I. Z., & Gvirts, H. Z. (2019). Herding brains: A core neural mechanism for social alignment. *Trends in Cognitive Sciences, 23*(3), 174–186.

Singh, A. K., Okamoto, M., Dan, H., Jurcak, V., & Dan, I. (2005). Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI. *NeuroImage, 27*(4), 842–851.

Smith, J. D., Boomer, J., Zakrzewski, A. C., Roeder, J. L., Church, B. A., & Ashby, F. G. (2014). Deferred feedback sharply dissociates implicit and explicit category learning. *Psychological Science, 25*(2), 447–457.

Smith, T. A., & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay-retention effect. *Journal of Experimental Psychology: Learning, Memory and Cognition, 36*, 80–95.

Song, S. H., & Keller, J. M. (2001). Effectiveness of motivationally adaptive computer-assisted instruction on the dynamic aspects of motivation. *Educational Technology Research and Development, 49*(2), 5–22.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology, 24–33.*

Squire, L. R., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annual Review of Psychology, 44*(1), 453–495.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Tan, S. J., Wong, J. N., & Teo, W. P. (2023). Is neuroimaging ready for the classroom? A systematic review of hyperscanning studies in learning. *NeuroImage, 120367.*

Theiler, J., Galdrikian, B., Longtin, A., Eubank, S., & Farmer, J. D. (1991). Testing for nonlinearity in time series: The method of surrogate data (No. LA-UR-91-3343; CONF-9108181-1). Los Alamos National Lab.

van der Helden, J., Boksem, M. A., & Blom, J. H. (2010). The importance of failure: Feedback-related negativity predicts motor learning efficiency. *Cerebral Cortex, 20*(7), 1596–1603.

Van Duijvenvoorde, A. C., Zanolie, K., Rombouts, S. A., Raijmakers, M. E., & Crone, E. A. (2008). Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *Journal of Neuroscience, 28*(38), 9495–9503.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage, 30*(4), 1414–1432.

Watanabe, K. (2013). Teaching as a dynamic phenomenon with interpersonal interactions. *Mind, Brain, and Education, 7*(2), 91–100.

Yang, J., Zhang, H., Ni, J., De Dreu, C. K. W., & Ma, Y. (2020). Within-group synchronization in the prefrontal cortex associates with intergroup conflict. *Nature Neuroscience, 23*(6), 754–760.

Zanolie, K., Van Leijenhorst, L., Rombouts, S. A. R. B., & Crone, E. A. (2008). Separable neural mechanisms contribute to feedback processing in a rule-learning task. *Neuropsychologia, 46*(1), 117–126.

Zheng, L., Liu, W., Long, Y., Zhai, Y., Zhao, H., Bai, X., … & Lu, C. (2020). Affiliative bonding between teachers and students through interpersonal synchronisation in brain activity. *Social Cognitive and Affective Neuroscience, 15*(1), 97–109.

Zimbardo, P. G., Johnson, R. L., & McCann, V. (2012). *Psychology: Core concepts* (7th ed.). Pearson.

Zhu, Y., Leong, V., Hou, Y., Zhang, D., Pan, Y., & Hu, Y. (2022). Instructor–learner neural synchronization during elaborated feedback predicts learning transfer. *Journal of Educational Psychology, 114*(6), 1427–1441.

Appendix 1: Concept Definitions and Examples

Term	Definition	Example 1 (for feedback)	Example 2 (for transfer)
Availability Heuristic	When judging the likelihood of an event, people tend to base their judgment on how easily specific examples come to mind.	1. In a study estimating causes of death, people overestimated murders, car accidents, and fires (which appear frequently in news) despite more deaths from lung disease.	2. When deciding post-task rewards, partners who invested equal time/effort both believe they contributed more because they more easily recall their own actions.
Representativeness Heuristic	When judging whether something belongs to a category, people assess similarity to typical category examples.	1. "If it looks like a duck, walks like a duck, and quacks like a duck, it's probably a duck."	2. People judge "HTHTH" as more likely than "HHHHT" when expecting random sequences, as the former appears more "random."
Foot-in-the-Door Effect	After agreeing to a small request, people become more likely to agree to a subsequent larger request.	1. A friend first asks you to accompany them to a nearby store, then to a farther mall; you're more likely to comply with the second request.	2. Animal shelter workers first ask people to wear "Adopt Don't Shop" badges, then later ask them to adopt; compliance increases.
Door-in-the-Face Effect	After refusing an extreme large request, people become more likely to agree to a subsequent more reasonable request.	1. After refusing a 1000 RMB alumni donation request, a person agrees to donate 100 RMB when asked.	2. Students asking for assignment extensions first request one week, but are more likely to receive 2 days.

Term	Definition	Example 1 (for feedback)	Example 2 (for transfer)
Hindsight Bias	After an event occurs, people tend to overestimate their ability to have predicted the outcome.	1. After celebrity breakups, people claim they "saw it coming."	2. After stock market crashes, investors say "the market was obviously due for correction."
Counterfactual Thinking	After an event, people imagine alternatives to reality that could have occurred but didn't.	1. Olympic silver medalists are less happy than bronze medalists, thinking "I almost won."	2. After a minor scooter accident, someone thinks how much worse it could have been without a helmet.
Deindividuation	In groups, behavioral constraints relax, increasing impulsive and deviant behavior.	1. Sports fans throw bottles after their team loses.	2. In 1967, 200 Oklahoma students chanted "Jump!" to a suicidal peer, who ultimately jumped.
Social Facilitation	Presence of others improves performance on simple tasks but impairs complex task performance.	1. Bosses should have employees work publicly for simple tasks but privately for challenging ones.	2. New postal workers sort mail slower with colleagues present, but experienced workers sort faster.
Fundamental Attribution Error	When explaining others' behavior, people overestimate internal dispositional factors and underestimate external situational factors.	1. We judge an admissions clerk as unfriendly when they're actually stressed from handling complaints.	2. In a quiz game, questioners are rated as smarter, ignoring their role advantage.

Term	Definition	Example 1 (for feedback)	Example 2 (for transfer)
Self-Serving Bias	For positive outcomes, people attribute to internal factors; for negative outcomes, to external factors.	1. Students attribute good grades to ability/effort but poor grades to test difficulty or poor teaching.	2. Drivers blame weather/roads for accidents but credit their own alertness for near-misses.

Note: The first 10 concepts were used for teaching. Adjacent concepts formed chunks. Example 1 was used for feedback; Example 2 for transfer measurement. Concepts 11-12 provided additional options for the 12-alternative test.

Appendix 2: Cognitive Load Rating Scale

This scale has six dimensions. Please read each dimension's description carefully, then mark the line corresponding to your actual experience during the task.

Mental Demand: How much mental activity (observing, remembering, thinking, searching) was required? Was the task easy or difficult mentally? Simple or complex? Relaxed or tense muscles?

Physical Demand: How much physical activity (pushing, pulling, turning, controlling movements) was required? Was the task easy or difficult physically?

Temporal Demand: Was the task pace slow or fast? Leisurely or rushed?

Effort: How much effort (mental and physical) was required to achieve your performance level?

Performance: How successful were you in achieving goals? How satisfied were you with your performance?

Frustration: How much frustration, annoyance, or stress did you experience during the task?

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.