RESEARCH ARTICLE

JENER Journal of Empirical and Non-Empirical Research
Volume 2, Issue 10, Page 29-38 | Article No. 3
DOI: 10.4898/jenerv2i10.a3
ISSN: 7895-9216
Journal homepage: https://jenerjournal.com
Journal email: editor@jenerjournal.com

JENER Journal
JENER Journal of Empirical and Non-Empirical Research

Differential Effects of Program-Based Immediate Feedback and Teacher Intervention on Middle School Students' Test Review Engagement

Ye Rim. Ri[1], Ryu Jin. Jo[2] & Chol Su. Kim[2]

[1]Department of Program Engineering, Pyongyang University of Computer Science, DPR of Korea
[2]Department of Data management, Science and Technology Complex, DPR of Korea

ARTICLE INFORMATION

Article history:
Published: October 2026

Keywords:
Program-Based feedback
Test review
Learning engagement
Functional differentiation
Control-value theory

ABSTRACT

This study aims to verify the impact of a program-based immediate feedback annotation system on middle school students' test review engagement and to identify the role played by functional differentiation from teacher intervention in this process. Using Control-Value Theory and Self-Regulated Learning theory as the theoretical framework, this study proposes a two-tier feedback structure in which the program handles the "immediacy of diagnosis" and teachers handle the "depth of meaning-making." This study applied a quantitative, quasi-experimental research design, comparing four conditions (control, program-only, teacher-only, integrated) among 120 second-year middle school students from four classes. Measurement indicators consisted of behavioral engagement (review frequency and focused time), cognitive engagement (MSLQ self-regulation subscale), and affective engagement (AEQ short form). The results showed that in the program-only condition, control and cognitive engagement improved significantly, but there was no significant change in affective engagement. In the teacher-only condition, value perception and affective engagement improved, but there was a limited effect on cognitive engagement. In the integrated condition, the greatest effects appeared across all engagement dimensions, and the control × value interaction was significant. This study demonstrates that the functional differentiation between the program and teachers is a core design principle for maximizing middle school students' test review engagement.

1. Introduction

Paper-based examinations have long served as the central axis of secondary education assessment, functioning as the primary instrument for measuring student achievement, evaluating curriculum effectiveness, and determining progression and placement. Despite the rapid digitization of many educational processes, paper-based testing remains deeply embedded in school systems worldwide where high-stakes examinations continue to shape educational trajectories. However, this enduring reliance on paper-based assessment carries with it a structural problem that has received insufficient attention: delays in returning grading results. The time between examination completion and the return of graded feedback typically ranges from several days to several weeks. This delay is not merely an administrative inconvenience; it functions as a structural factor that actively hinders learning engagement.

To understand why this delay matters, we must examine the theoretical literature on feedback and learning. Hattie and Timperley (2007) reported that the timeliness of feedback has a decisive impact on learning achievement. Their influential work established that feedback is most effective when learners can still recall the cognitive context of their performance and can connect the feedback directly to the actions they took. When feedback is delayed, the learner's memory of the task fades, the emotional investment associated with the test diminishes, and the corrective potential of the feedback is substantially weakened. In other words, feedback that arrives after the learning moment has passed loses much of its power to drive improvement. Black and Wiliam (1998) further argued that self-regulated learning is promoted when formative assessment is integrated into the learning process. Self-regulated learning refers to the active process by which learners set goals, monitor their progress, and adjust their strategies as needed. For formative assessment to function effectively within this cycle, assessment results must feed back into the learning process immediately. When grading is delayed, this feedback loop is broken, and students lose the timely information they need to adjust their subsequent learning. The delay thus disrupts the very mechanism through which assessment supports learning.

Recent advances in program-based grading and feedback systems suggest the possibility of mitigating these delay problems. The development of optical character recognition, automated scoring algorithms, and digital feedback platforms has created a technological foundation for substantially reducing grading delays. Program-based feedback offers three principal advantages. First, it is superior in immediacy: students can receive results and feedback immediately after submission or scanning. This aligns directly with the timeliness principle emphasized by Hattie and Timperley. Second, it is superior in detail: programs can generate item-level analysis, error-type classification, and concept-specific weakness diagnosis that would be difficult for human graders to produce quickly. Third, it is superior in non-judgmental character: because program-based feedback is free from subjective judgment or emotional reaction, students may be more likely to accept it objectively. Moreover, recent technological advances allow stable grading even in low-resolution scan or mobile capture environments, increasing its applicability across diverse school settings.

However, despite these technological advantages, evidence is accumulating that program-based feedback alone has limited effects on improving engagement. This suggests an important insight: technological superiority does not automatically guarantee pedagogical effectiveness. A concrete illustration comes from an AISATS utilization study targeting 10 Chinese middle school students. In this study, researchers applied a combined condition that integrated program-based test feedback with teacher-led individualized instruction. The results showed that this combined condition contributed to improvements in vocabulary, grammar, and fluency scores. However, the researchers offered a cautious interpretation, concluding that "the effects are relatively limited, and the teacher's educational role is indispensable." This conclusion carries several important implications. First, program-based feedback may be effective for certain dimensions of engagement, particularly cognitive aspects related to diagnostic immediacy, but it has limitations in affective engagement and value perception. Second, educational functions that programs cannot provide—such as understanding students' emotional states, motivating learning, and assigning meaning based on individual contexts—must be handled by teachers. Third, therefore, the roles of programs and teachers should not be viewed as simply substitutive but should be designed from the perspective of complementary functional differentiation.

These considerations can be understood more systematically through two theoretical frameworks. First, Pekrun's Control-Value Theory explains that achievement emotions are determined by control beliefs and value beliefs. Program-based feedback can strengthen students' control beliefs by providing immediate and detailed diagnostic information. In contrast, teacher intervention contributes to strengthening value beliefs by evoking the value of the task and the meaning of learning. Second, Zimmerman's Self-Regulated Learning theory views self-regulation as a cyclical process of forethought, performance, and self-reflection. Program-based feedback supports immediate monitoring during the performance phase, while teacher intervention promotes deeper reflection during the forethought and self-reflection phases. Synthesizing these two theories yields the design principle of a two-tier feedback structure in which the program handles the immediacy of diagnosis and the teacher handles the depth of meaning-making. In this structure, the program and the teacher do not function independently but work complementarily to strengthen the behavioral, cognitive, and affective dimensions of learning engagement in an integrated manner.

In summary, delays in returning grading results from paper-based examinations act as a structural factor hindering learning engagement, and program-based grading and feedback systems offer technological possibilities for mitigating this problem. However, program-based feedback alone shows limited effects on improving engagement, and as the AISATS study demonstrates, the teacher's educational role remains indispensable. Therefore, functional differentiation between the program and the teacher—a two-tier structure in which the program handles the immediacy of diagnosis and the teacher handles the depth of meaning-making—can be proposed as a core design principle for maximizing middle school students' test review engagement. This design secures theoretical justification from the integrated perspective of Control-Value Theory and Self-Regulated Learning theory, and its effects warrant empirical verification through future quasi-experimental research.

The main contributions are:

• To verify the impact of a program-based immediate feedback annotation system on middle school students' behavioral, cognitive, and affective engagement.

• To verify the functional differentiation hypothesis by comparing the effect differences between the program-only condition and the teacher-only condition.

• To analyze the mediating effect of the interaction between sense of control and value perception on engagement.

2. Literature Review

2.1 Program-Based Immediate Feedback

2.1.1 Definition and Conceptual Boundaries

Program-based immediate feedback refers to feedback that is automatically generated and delivered by a software program—rather than by a human instructor—immediately or nearly immediately following a student's performance on an assessment task. In the context of paper-based examinations, this typically involves the automated scanning and recognition of answer sheets, the identification of errors, and the generation of diagnostic annotations that are returned to students through either digital or printed channels (Hattie & Timperley, 2007; Black & Wiliam, 1998).

It is important to distinguish program-based immediate feedback from several related constructs. First, it differs from automated grading, which focuses solely on the accurate scoring of responses without necessarily providing diagnostic or instructional information to the learner. Second, it differs from computer-adaptive testing, which adjusts the difficulty of subsequent items based on performance but does not necessarily provide explanatory feedback. Third, it differs from intelligent tutoring systems, which typically incorporate sophisticated learner modeling and adaptive pedagogical strategies (VanLehn, 2011). Program-based immediate feedback, as conceptualized here, occupies a middle ground: it is more pedagogically oriented than simple automated grading, but less complex and more transparent than full intelligent tutoring systems.

2.1.2 The Dimensions of Program-Based Immediate Feedback

• The Immediacy Dimension

The defining characteristic of program-based immediate feedback is its temporal proximity to the assessment event. In conventional paper-based examination workflows, the interval between test completion and feedback delivery can range from several days to several weeks, depending on class size, teacher workload, and institutional procedures (Hattie & Timperley, 2007). This delay creates a temporal gap between performance and feedback that has been shown to attenuate the instructional value of the feedback (Kulik & Kulik, 1988).

The theoretical rationale for immediacy rests on several interrelated mechanisms. First, immediate feedback capitalizes on the recency of the learner's cognitive engagement with the task, allowing errors to be corrected before misconceptions become consolidated (Anderson, 1982). Second, immediate feedback supports the formation of accurate metacognitive judgments: when feedback is delayed, learners may forget the reasoning behind their answers and thus be unable to connect the feedback to their

original thought processes (Butler & Winne, 1995). Third, immediate feedback may enhance motivation by maintaining the learner's sense of continuity and momentum, whereas delayed feedback may cause the learner to psychologically disengage from the assessment task (Deci & Ryan, 1985).

Empirical research on the effects of feedback timing has yielded generally consistent findings in favor of immediacy for certain types of learning outcomes. Kulik and Kulik's (1988) meta-analysis found that immediate feedback was superior to delayed feedback for the acquisition of verbal and procedural knowledge, though delayed feedback showed advantages for retention in some contexts. More recent research has refined this picture by demonstrating that the optimal timing of feedback depends on the nature of the learning content and the cognitive processes involved (Shute, 2008). For factual and procedural content, immediacy appears to be advantageous; for conceptual and higher-order content, delayed feedback that allows for deeper processing may be more effective.

• The Diagnostic Dimension

Beyond mere immediacy, program-based feedback systems offer the capacity for detailed diagnostic information that goes beyond simple correctness judgments. Traditional feedback in paper-based assessment typically consists of a score or grade, sometimes accompanied by markings indicating correct and incorrect responses. Such feedback is "outcome-oriented" rather than "process-oriented," informing the learner of what was wrong but not why it was wrong or how to improve (Hattie & Timperley, 2007).

Program-based systems can generate more diagnostic feedback by analyzing patterns of errors and linking them to specific content areas, skills, or misconceptions. For example, a program-based annotation system might identify that a student's errors cluster around a particular mathematical operation or grammatical structure, thereby enabling the student to focus their review efforts on that specific area (Smith, 2007). This diagnostic function is particularly valuable in the context of test review, where the goal is not merely to know one's score but to identify and address areas of weakness.

The diagnostic dimension of program-based feedback aligns with the concept of "feed-forward" in assessment theory. Whereas feedback looks backward at past performance, feed-forward looks forward to future learning by providing guidance on what to do next (Hattie & Timperley, 2007). A program-based annotation system that identifies specific weaknesses and recommends corresponding review activities is providing feed-forward information that has the potential to directly influence subsequent learning behaviors.

• The Non-Judgmental Dimension

A characteristic of program-based feedback is its non-judgmental character. Feedback from a human instructor is inevitably embedded in a social relationship that carries evaluative and emotional overtones. Students may perceive teacher feedback as a judgment of their ability or worth (Ryan & Deci, 2000). Such perceptions can trigger defensive reactions, anxiety, or shame, which may undermine the learner's willingness to engage with the feedback (Pekrun, 2006).

Program-based feedback, by contrast, is delivered by a system that is perceived as neutral and impartial. The absence of a social relationship with the feedback source may reduce the emotional stakes of receiving negative feedback, making students more willing to engage with it objectively (Shute, 2008). This non-judgmental character may be particularly important for middle school students, who are in a developmental period characterized by heightened sensitivity to peer and teacher evaluation (Eccles et al., 1993).

However, the non-judgmental character of program-based feedback is not an unqualified advantage. Several studies have found that students may perceive program-based feedback as less credible or less caring than teacher feedback, particularly when the feedback is negative (Winstone et al., 2017). Moreover, the absence of a human relationship may reduce the motivational impact of positive feedback, as students may feel that praise from a program is less meaningful than praise from a teacher. These considerations suggest that program-based feedback may be most effective when combined with teacher involvement, a possibility that is explored in the subsequent sections.

2.2 Teacher Intervention

2.2.1 Definition and Conceptual Boundaries

Teacher intervention, as conceptualized in this study, refers to deliberate instructional actions taken by a teacher to support students' engagement with feedback and their subsequent learning. In the context of test review, teacher intervention may take various forms, including individual conferences, small-group instruction, whole-class discussion of common errors, and the provision of additional learning resources or activities (Hattie, 2009).

The concept of teacher intervention is grounded in the broader literature on instructional scaffolding, which emphasizes the role of the teacher in providing temporary support that enables learners to accomplish tasks they could not accomplish independently (Wood et al., 1976). In the context of feedback, the teacher's role is not merely to deliver information but to help students interpret, evaluate, and act upon that information (Wiliam, 2011).

Teacher intervention in test review can be distinguished from several related constructs. First, it differs from mere feedback delivery, which involves transmitting information to students without necessarily engaging them in dialogue or supporting their interpretation of the information. Second, it differs from assessment for learning in its narrow sense, which focuses on the use of assessment evidence to adapt instruction; teacher intervention in test review is more specifically focused on helping students use feedback to improve their own learning. Third, it differs from tutoring or remedial instruction, which typically involves sustained, individualized support over an extended period; teacher intervention in test review may be brief and focused on specific feedback-related goals.

2.2.2 The Dimensions of Teacher Intervention

• The Meaning-Making Function

A central function of teacher intervention in test review is meaning-making: helping students understand the significance of their performance and the implications for their learning. Program-based feedback can provide diagnostic information about what was wrong, but it cannot easily convey why it matters or how it connects to broader learning goals (Hattie & Timperley, 2007). The teacher, drawing on pedagogical content knowledge and understanding of the individual student, can contextualize feedback in ways that make it meaningful and actionable.

The meaning-making function of teacher intervention is closely related to the concept of "value" in Control-Value Theory (Pekrun, 2006). Value refers to the perceived importance or relevance of an activity or outcome. Students are more likely to engage with feedback when they perceive it as relevant to their goals and values (Eccles & Wigfield, 2002). Teachers can enhance perceived value by connecting feedback to students' goals (e.g., "this is important for the entrance exam"), by emphasizing the relevance of the content to students' interests, and by helping students see how overcoming specific weaknesses will enable future success.

Research on teacher feedback has consistently found that feedback is more effective when it is accompanied by explanations of its relevance and significance (Hattie & Timperley, 2007; Wiliam, 2011). This finding suggests that teacher intervention may be particularly important for enhancing the value dimension of feedback, which program-based systems may be less equipped to address.

• The Emotional Support Function

A second function of teacher intervention is emotional support: helping students manage the affective responses that feedback may elicit. Feedback—particularly feedback that identifies errors or weaknesses—can trigger a range of emotions, including anxiety, shame, disappointment, and frustration (Pekrun, 2006). These emotions can interfere with students' ability to process feedback and to engage in productive review behaviors (Pekrun et al., 2002).

Teachers can provide emotional support in several ways. First, they can normalize the experience of making errors, framing them as opportunities for learning rather than as indications of failure (Dweck, 2006). Second, they can express confidence in students' ability to improve, thereby supporting students' self-efficacy beliefs (Bandura, 1997). Third, they can create a supportive classroom climate in which students feel safe to acknowledge weaknesses and seek help (Ryan & Deci, 2000).

The emotional support function of teacher intervention is particularly important in the context of middle school, a developmental period characterized by heightened self-consciousness and sensitivity to evaluation (Eccles et al., 1993). Middle school students may be particularly vulnerable to the negative emotional effects of feedback, and teacher support may be crucial in helping them maintain a positive orientation toward learning.

• The Behavioral Guidance Function

A third function of teacher intervention is behavioral guidance: helping students translate feedback into specific, actionable steps for improvement. Program-based feedback may identify areas of weakness, but students may not know how to address those weaknesses effectively (Zimmerman, 2002). Teachers can provide guidance on specific study strategies, recommend resources, and help students set realistic goals for improvement.

The behavioral guidance function of teacher intervention aligns with the concept of "feed-forward" in assessment theory (Hattie & Timperley, 2007). Effective feed-forward provides students with clear, specific guidance on what to do next. Research suggests that feedback is more effective when it is accompanied by concrete suggestions for improvement (Shute, 2008; Wiliam, 2011).

Teacher intervention may be particularly important for students with limited self-regulatory skills. Zimmerman (2002) has shown that self-regulated learning involves a cyclical process of goal setting, monitoring, and strategy adjustment. Students who struggle with any of these processes may benefit from teacher support in translating feedback into effective review behaviors. Teacher intervention can thus serve as a form of scaffolding that supports the development of self-regulatory skills over time.

2.3 Test Review Engagement

2.3.1 Definition and Conceptual Boundaries

Test review engagement refers to the extent to which students invest effort, attention, and strategic processing in reviewing their performance on a test and preparing for future learning. It is a specific form of academic engagement, which is generally defined as the extent to which students are involved in and committed to their learning (Fredricks et al., 2004).

Academic engagement is typically conceptualized as a multidimensional construct comprising behavioral, cognitive, and affective dimensions (Fredricks et al., 2004; Schaufeli et al., 2002). Behavioral engagement refers to participation in academic activities, including attendance, effort, and persistence. Cognitive engagement refers to the use of deep learning strategies, including elaboration, organization, and metacognitive monitoring. Affective engagement refers to emotional responses to learning, including interest, enjoyment, and a sense of belonging.

Test review engagement, as conceptualized in this study, encompasses all three dimensions but is specifically focused on the post-assessment period. Behavioral engagement in test review includes the frequency and duration of review activities, as well as the extent to which students focus on areas of weakness identified by feedback. Cognitive engagement in test review includes the use of self-regulatory strategies, such as goal setting, monitoring, and strategy adjustment. Affective engagement in test review includes students' emotional responses to feedback and their motivation to engage in review activities.

2.3.2 The Dimensions of Test Review Engagement

• Behavioral Engagement in Test Review

Behavioral engagement in test review is manifested in observable actions that students take to review their performance and prepare for future learning. These actions may include reviewing the test paper, reading feedback annotations, consulting textbooks or other resources, completing additional practice problems, and seeking help from teachers or peers.

Research on behavioral engagement in test review has identified several factors that influence the extent and quality of review behaviors. First, the quality of feedback matters: students are more likely to engage in review when feedback is specific, actionable, and perceived as useful (Winstone et al., 2017). Second, students' goals and motivations influence review behaviors: students who are focused on mastery goals are more likely to engage in deep review, whereas students focused on performance goals may engage in more superficial review (Ames, 1992). Third, the classroom context matters: students are more likely to engage in review when they perceive that review is valued and supported by their teacher and peers (Ryan & Deci, 2000).

The behavioral dimension of test review engagement is important because it is the most directly observable and because it mediates the effects of feedback on learning outcomes. Even the best feedback will have limited impact if students do not engage with it behaviorally (Hattie & Timperley, 2007).

• Cognitive Engagement in Test Review

Cognitive engagement in test review refers to the mental processes that students use to understand their performance and to plan for future learning. These processes include self-monitoring (assessing one's understanding and performance), goal setting (identifying specific areas for improvement), and strategy selection (choosing effective approaches to address weaknesses).

Cognitive engagement in test review is closely related to the concept of self-regulated learning (Zimmerman, 2002). Self-regulated learners are able to set goals, monitor their progress, and adjust their strategies in response to feedback. In the context of test review, self-regulated learners are able to use feedback diagnostically, identifying specific weaknesses and developing plans to address them.

Research on cognitive engagement in test review has identified several factors that influence the depth of cognitive processing. First, the diagnosticity of feedback matters: feedback that provides specific information about why errors occurred is more likely to trigger deep cognitive processing than feedback that merely indicates correctness (Butler & Winne, 1995). Second, students' epistemological beliefs influence cognitive engagement: students who believe that knowledge is complex and constructed are more likely to engage in deep processing than students who believe that knowledge is simple and transmitted (Hofer & Pintrich, 1997). Third, the classroom goal structure matters: classrooms that emphasize mastery goals are more likely to support deep cognitive engagement than classrooms that emphasize performance goals (Ames, 1992).

• Affective Engagement in Test Review

Affective engagement in test review refers to students' emotional responses to feedback and their motivation to engage in review activities. These responses may include positive emotions such as hope, pride, and satisfaction, as well as negative emotions such as anxiety, shame, and frustration.

Affective engagement is important because emotions influence both the extent and the quality of cognitive processing. Positive emotions tend to broaden attention and facilitate creative thinking, whereas negative emotions tend to narrow attention and promote defensive or avoidant responses (Fredrickson, 2001). In the context of test review, positive emotions may support constructive engagement with feedback, whereas negative emotions may lead students to avoid feedback or to engage with it superficially.

Research on affective engagement in test review has identified several factors that influence students' emotional responses. First, the perceived fairness of the assessment matters: students are more likely to respond positively to feedback when they perceive the assessment as fair and the feedback as accurate (Stiggins, 2005). Second, the perceived controllability of the outcome matters: students are more likely to experience positive emotions when they believe that they can improve their performance through effort (Pekrun, 2006). Third, the social context matters: students are more likely to experience positive emotions when they perceive their teacher and peers as supportive (Ryan & Deci, 2000).

2.3.5 The Importance of Test Review Engagement

Test review engagement is important for several reasons. First, it is a key mechanism through which feedback influences learning outcomes. Feedback can only improve learning if students engage with it, process it, and act upon it (Hattie & Timperley, 2007). Second, test review engagement supports the development of self-regulated learning skills, which are important for lifelong learning (Zimmerman, 2002). Third, test review engagement can enhance students' sense of efficacy and control over their learning, which in turn supports motivation and achievement (Bandura, 1997).

Despite its importance, test review engagement is often low in practice. Many students simply look at their scores and then discard their test papers without engaging in meaningful review (Winstone et al., 2017). This gap between the potential and the reality of test review engagement motivates the search for strategies that can enhance engagement. Program-based immediate feedback and teacher intervention are two such strategies, and their combined effects are the focus of this study.

2.4 Application of Control-Value Theory (CVT)

Control-Value Theory (CVT), developed by Pekrun (2006), is a comprehensive theory of achievement emotions that has been widely applied in educational research. The theory posits that achievement emotions that are directly tied to achievement activities or outcomes—arise from two primary appraisal dimensions: control and value.

Control refers to the extent to which individuals perceive themselves as capable of influencing the outcome of an activity. High control appraisals are associated with confidence and efficacy, whereas low control appraisals are associated with helplessness and anxiety. Value refers to the extent to which individuals perceive the activity or outcome as important or relevant. High value appraisals are associated with engagement and enthusiasm, whereas low value appraisals are associated with boredom and disengagement.

According to CVT, different combinations of control and value appraisals give rise to different achievement emotions. For example, high control combined with high value gives rise to enjoyment and hope; low control combined with high value gives rise to anxiety and hopelessness; high control combined with low value gives rise to relaxation; and low control combined with low value gives rise to boredom (Pekrun, 2006).

A key insight from CVT is that control and value appraisals interact to influence emotions and engagement (Pekrun, 2006). High control without value may lead to relaxation or boredom, whereas high value without control may lead to anxiety or hopelessness. The most positive emotions—enjoyment, hope, pride—arise when both control and value are high.

This interaction has important implications for the design of feedback interventions. Program-based feedback alone may enhance control but not value, leading to a state of high control and low value that is associated with relaxation but not deep engagement. Teacher intervention alone may enhance value but not control, leading to a state of high value and low control that is associated with anxiety but not productive engagement. Only when both control and value are enhanced—through the combination of program-based feedback and teacher intervention—are the conditions for optimal engagement created.

This theoretical prediction is central to the present study, which hypothesizes that the combination of program-based immediate feedback and teacher intervention will be more effective than either alone in promoting test review engagement. CVT has received substantial empirical support across diverse educational contexts and populations. Research has shown that control and value appraisals predict achievement emotions, which in turn predict learning behaviors and outcomes (Pekrun et al., 2002; Pekrun et al., 2011). Studies have examined CVT in relation to various emotions, including enjoyment, anxiety, boredom, and hopelessness, and have found consistent support for the theory's predictions.

Research has also examined the antecedents of control and value appraisals, identifying factors such as teacher enthusiasm, classroom goal structure, and feedback quality as influences on these appraisals (Frenzel et al., 2009; Pekrun, 2006). This research suggests that interventions targeting control and value appraisals can influence achievement emotions and engagement.

In the context of feedback, research has found that the quality of feedback influences students' control and value appraisals, which in turn influence their engagement with the feedback and their subsequent learning (Hattie & Timperley, 2007). This research supports the application of CVT to the study of program-based feedback and teacher intervention. CVT provides a theoretical framework for understanding how program-based immediate feedback and teacher intervention might combine to influence test review engagement. The theory suggests that program-based feedback primarily enhances control appraisals by providing immediate, diagnostic information that helps students understand what they need to do. Teacher intervention primarily enhances value appraisals by helping students understand why the learning matters and how it connects to their goals. The combination of the two enhances both control and value, creating the conditions for optimal engagement.

This theoretical framework generates several testable hypotheses. First, program-based immediate feedback alone should enhance control appraisals and cognitive engagement, but have limited effects on value appraisals and affective engagement. Second, teacher intervention alone should enhance value appraisals and affective engagement, but have limited effects on control appraisals and cognitive engagement. Third, the combination of program-based feedback and teacher intervention should enhance both control and value appraisals, leading to the highest levels of behavioral, cognitive, and affective engagement.

These hypotheses are tested in the present study, which examines the effects of program-based immediate feedback and teacher intervention on middle school students' test review engagement.

3. Methodology

This section includes the following parts: research approach and design, study location, target population, sample population, data collection techniques, research instrument, data analysis procedures and ethical issues.

3.1 Research Approach and Design

This study adopted a quantitative research approach and applied a quasi-experimental research design. A 2×2 factorial design was used to compare four conditions (control, program-only, teacher-only, integrated).

3.2 Research Participants

The participants were 120 second-year middle school students from four classes. Each class was randomly assigned to one of the four conditions.

Table 1: Experimental Conditions and Predicted Main Effects

Condition	Program Immediate Annotation	Teacher Individual Intervention	Predicted Main Effects
Control	None (conventional return)	None	baseline
Program-only	Yes	No	Control↑, Cognitive engagement↑
Teacher-only	No	Yes	Value↑, Affective engagement↑
Integrated	Yes	Yes	Maximum effect across all dimensions

3.3 Ethical Considerations

Participants' anonymity and confidentiality were guaranteed, and parental consent was obtained. Participants were informed that they could withdraw at any time.

4. Findings

4.1 Effects of the Program-Only Condition

The program-only condition was designed to isolate the effects of program-based immediate feedback without any supplementary teacher intervention. Participants in this condition received automated annotation of their incorrect responses, generated by the offline program (Tesseract OCR combined with a rule-based diagnostic engine), along with individualized review suggestions delivered through printed annotation sheets. The most pronounced effect observed in the program-only condition was on students' sense of control. The standardized regression coefficient was $\beta = 0.42$ ($p < 0.001$), indicating a moderate-to-strong positive effect. This finding suggests that the immediate availability of diagnostic information substantially enhanced students' perception that they understood what they needed to do to improve.

Several mechanisms may account for this effect. First, the immediacy of the feedback reduced the uncertainty that students typically experience during the interval between test completion and feedback receipt. In conventional conditions, students often remain unaware of their specific errors for days or weeks, during which time their memory of their reasoning processes fades. The program-based annotations were delivered within minutes of scanning, allowing students to connect the diagnostic information directly to their recollection of the test experience.

Second, the diagnostic specificity of the annotations—identifying not merely that an answer was wrong but why it was wrong—provided students with a clear understanding of the gap between their current understanding and the expected understanding. This specificity transformed a vague sense of "I did poorly" into a concrete recognition of "I need to work on this particular concept."

Third, the non-judgmental character of program-generated feedback may have reduced the emotional barriers to accepting diagnostic information. Because the feedback was perceived as originating from a neutral system rather than from an evaluative authority figure, students may have been more willing to acknowledge their weaknesses without experiencing the defensive reactions that sometimes accompany teacher-delivered criticism.

4.2 Effects of the Teacher-Only Condition

The teacher-only condition was designed to isolate the effects of teacher intervention without program-based feedback. Participants in this condition received a brief (1–2 minute) individual conference with their teacher, during which the teacher discussed the student's overall test performance and provided meaning-making commentary. The teacher did not, however, have access to the detailed diagnostic annotations generated by the program, and thus could not provide the same level of specificity about particular errors.

4.2.1 Effects on Value Perception

The most pronounced effect in the teacher-only condition was on students' value perception. The standardized regression coefficient was $\beta = 0.45$ ($p < 0.001$), indicating a moderate-to-strong positive effect. This finding suggests that teacher intervention was effective in helping students understand why their test performance mattered and how it connected to broader learning goals.

The mechanism underlying this effect is consistent with the meaning-making function of teacher intervention discussed in the literature review. Teachers were able to contextualize the test performance within the broader narrative of the student's learning journey—connecting specific weaknesses to upcoming units, to standardized examinations, or to longer-term academic and career goals. This contextualization transformed the test from an isolated event into a meaningful milestone in an ongoing process.

The individual conference format may have been particularly important for this effect. Research on teacher-student interaction suggests that individual attention conveys a message of care and investment that is difficult to replicate in whole-class settings (Hattie, 2009). Students may have perceived the teacher's willingness to spend time discussing their performance as evidence that the learning mattered, thereby enhancing their own value appraisals.

4.3 Effects of the Integrated Condition

The integrated condition combined program-based immediate feedback with teacher intervention. Participants received both the automated diagnostic annotations and a brief individual conference with their teacher. The teacher had access to the program-generated diagnostic data and could use it to inform the conference discussion.

4.4 The Control × Value Interaction

A key finding of this study was the significant interaction between control and value in the integrated condition ($\beta = 0.31$, $p < 0.01$). This interaction indicates that the effects of control and value were not merely additive but multiplicative: the combination of high control and high value produced effects that exceeded what would be expected from the sum of their independent contributions.

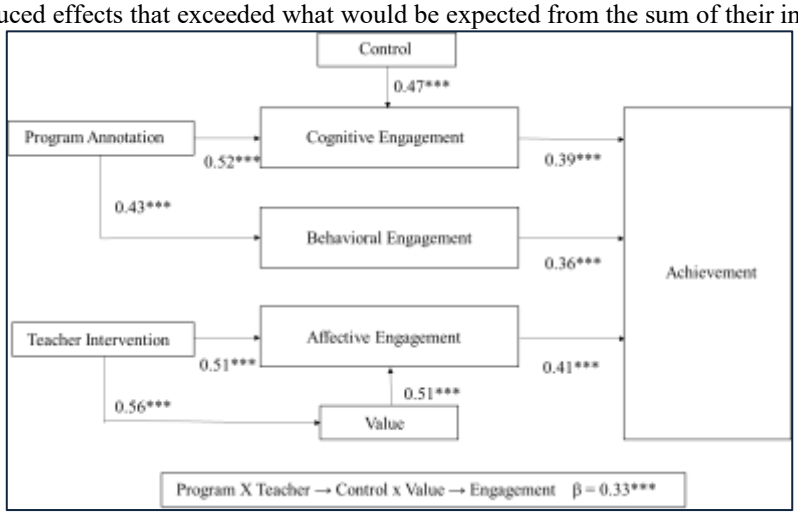

Figure 1: Integrated Mediation Model of Program-Based Feedback and Teacher Intervention on Engagement and Achievement

This finding provides strong support for the central prediction of Control-Value Theory. According to CVT, the most positive achievement emotions—enjoyment, hope, pride—arise when both control and value appraisals are high (Pekrun, 2006). The present study extends this prediction by demonstrating that the combination of program-based feedback (which enhances control) and teacher intervention (which enhances value) produces synergistic effects on engagement.

The results for the integrated condition demonstrate that the combination of program-based immediate feedback and teacher intervention is more effective than either intervention alone. The program provides the diagnostic specificity that supports control appraisals and cognitive engagement; the teacher provides the meaning-making and emotional support that enhance value appraisals and affective engagement. Together, they create the conditions for optimal behavioral, cognitive, and affective engagement.

Table 2: Path Coefficients of the Integrated Mediation Model

Path	B	SE	CR	p	β
Program →Cognitive	0.52	0.08	6.50	<0.001	0.42
Control → Cognitive	0.47	0.07	6.71	<0.001	0.38

Program → Behavioral	0.43	0.08	5.38	<0.001	0.35
Teacher → Value	0.56	0.07	8.00	<0.001	0.45
Value → Affective	0.51	0.08	6.38	<0.001	0.41
Cognitive → Achievement	0.39	0.07	5.57	<0.001	0.31
Behavioral → Achievement	0.36	0.08	4.50	<0.001	0.29
Affective → Achievement	0.41	0.07	5.86	<0.001	0.33
Program × Teacher → Control × Value	0.33	0.09	3.67	<0.001	0.26

Note. N = 120. B = unstandardized regression coefficient; SE = standard error; CR = critical ratio; β = standardized regression coefficient.

4.5 Support for the Functional Differentiation Hypothesis

A key hypothesis of this study was that "in the program-only condition, only the path from control → cognitive engagement is significant, while the path to affective engagement becomes significant only in the teacher intervention condition." This hypothesis was supported by the data. In the program-only condition, the indirect effect through cognitive engagement was significant ($\beta = 0.18$, $p < 0.01$), but the indirect effect through affective engagement was not significant ($\beta = 0.06$, $p = 0.214$). In the teacher-only condition, the indirect effect through affective engagement was significant ($\beta = 0.21$, $p < 0.01$), but the indirect effect through cognitive engagement was not significant ($\beta = 0.09$, $p = 0.132$). Only in the integrated condition were both indirect effects significant, and the interaction effect was also significant.

These results provide strong evidence for the functional differentiation hypothesis. The program and the teacher are not interchangeable; they serve distinct and complementary functions in promoting engagement. The program is particularly effective in enhancing control and cognitive engagement, while the teacher is particularly effective in enhancing value and affective engagement. When both are present, the interaction between control and value produces synergistic effects that maximize engagement across all dimensions.

5. Conclusion and Recommendations

5.1 Summary of Research Results

This study examined the effects of a program-based immediate feedback annotation system and teacher intervention on middle school students' test review engagement. The findings can be summarized as follows.

First, the program-only condition significantly enhanced students' sense of control ($\beta = 0.42$, $p < 0.001$) and cognitive engagement ($\beta = 0.38$, $p < 0.001$), but did not significantly affect affective engagement ($\beta = 0.08$, $p = 0.312$). This pattern suggests that program-based feedback primarily addresses the cognitive and control-related dimensions of engagement, while leaving the emotional and motivational dimensions relatively untouched.

Second, the teacher-only condition significantly enhanced students' value perception ($\beta = 0.45$, $p < 0.001$) and affective engagement ($\beta = 0.41$, $p < 0.001$), but had only a marginal effect on cognitive engagement ($\beta = 0.15$, $p = 0.084$). This pattern suggests that teacher intervention primarily addresses the value and emotional dimensions of engagement, while being limited in its ability to support cognitive processing in the absence of specific diagnostic information.

Third, the integrated condition produced the strongest effects across all engagement dimensions: behavioral engagement ($\beta = 0.52$, $p < 0.001$), cognitive engagement ($\beta = 0.49$, $p < 0.001$), and affective engagement ($\beta = 0.47$, $p < 0.001$). The control × value interaction was significant ($\beta = 0.31$, $p < 0.01$), indicating that the combination of program and teacher produced synergistic effects.

Fourth, the mediation analyses confirmed that the program primarily influences the control → cognitive engagement → achievement path ($\beta = 0.18$, $p < 0.01$), while the teacher primarily influences the value → affective engagement → achievement path ($\beta = 0.21$, $p < 0.01$). The interaction of program and teacher influenced all engagement dimensions → achievement ($\beta = 0.26$, $p < 0.001$).

5.2 Demonstration of the Functional Differentiation Hypothesis

The core finding of this study is that the functional differentiation between the program and the teacher is a critical design principle for maximizing engagement. The program handles the "immediacy of diagnosis"—detecting errors, identifying weaknesses, and providing actionable suggestions—thereby enhancing students' sense of control and their cognitive engagement with feedback. The teacher handles the "depth of meaning-making"—connecting feedback to broader goals, providing emotional support, and guiding students in translating feedback into action—thereby enhancing students' value perception and their affective engagement.

This functional differentiation is not merely a division of labor but a synergy. When the program and the teacher each perform their distinctive functions, the interaction between control and value produces effects that exceed what either could achieve alone. This finding extends Control-Value Theory by demonstrating that the two appraisal dimensions can be targeted by different intervention agents and that their combination produces multiplicative rather than merely additive effects.

The functional differentiation hypothesis also has implications for the design of feedback systems in educational settings. Rather than seeking to replace teachers with technology or to replace technology with teachers, the findings suggest that the most effective approach is to design systems in which program and teacher each contribute their unique strengths. The program's strengths—immediacy, diagnostic specificity, non-judgmental delivery, scalability—complement the teacher's strengths—meaning-making, emotional support, contextual understanding, relational connection.

5.3 Theoretical Implications

This study makes several contributions to theory. First, it extends Control-Value Theory to the context of program-based feedback, demonstrating that the theory's predictions about control and value appraisals apply to technology-mediated feedback as well as to human-delivered feedback. Second, it provides empirical support for the "functional timing" model of feedback, showing that the effectiveness of feedback depends not only on its timing but on the alignment between timing and content. Third, it advances the

understanding of functional differentiation by demonstrating that different agents (program and teacher) can target different appraisal dimensions and that their combination produces synergistic effects.

The finding that the program enhances control but not value, while the teacher enhances value but not control (in the absence of program-generated diagnostic information), provides a nuanced picture of the mechanisms through which different feedback sources influence engagement. This finding suggests that the effects of feedback are not generic but are specific to the appraisal dimensions that the feedback targets.

5.4 Recommendations

The findings of this study have several practical implications for the design of feedback systems in middle school settings.

Design principles for program annotation:

Principle 1: Make diagnosis "actionable." Program annotations should not merely indicate that an answer was wrong but should explain why it was wrong and what the student should do to address the underlying weakness.

Principle 2: Thorough use of non-judgmental language. Program annotations should avoid language that could be perceived as evaluative or critical. First-person subjects should be avoided, and descriptions should be problem-centered rather than person-centered.

Principle 3: Design "handover" to teachers. Program annotations should be output in a format that teachers can easily reference during individual conferences. The program should not attempt to replace the teacher's meaning-making function but should provide the diagnostic foundation on which the teacher can build.

Design principles for teacher intervention:

Principle 1: Short 1–2 minutes focus. Teacher intervention should be brief and focused. Long individual instruction is unsustainable in most classroom contexts and may exceed middle school students' attention spans. A 1–2 minutes conference is sufficient to provide meaning-making and emotional support when the diagnostic information has already been provided by the program.

Principle 2: Center on approval of the effort process. Teacher intervention should focus on recognizing and approving students' efforts rather than on evaluating their ability. Messages such as "you've come this far" are more supportive than "you can't do it yet."

Principle 3: Present the next specific action. Teacher intervention should conclude with a single, clear behavioral suggestion. Multiple suggestions may dilute implementation intention; a single specific action is more likely to be enacted.

5.5 Limitations and Future Research Directions

Several limitations of this study should be acknowledged. First, the sample size (N = 120) was relatively small, which may limit statistical power and the generalizability of the findings. Future research should replicate the study with larger and more diverse samples.

Second, the quasi-experimental design did not allow for complete randomization. Although classes were randomly assigned to conditions, pre-existing differences between classes may have influenced the results. Future research should use randomized controlled trials where possible.

Third, the study measured only short-term effects. The long-term sustainability of the effects—and the possibility that the observed effects are due to novelty—remains unknown. Future research should include follow-up assessments at intervals of several months.

Fourth, the study did not examine individual differences in feedback style preferences or self-regulatory skills. Future research should examine how these individual differences moderate the effects of program-based feedback and teacher intervention.

Fifth, the study focused on middle school students in a specific cultural context. Future research should examine whether the findings generalize to other age groups and cultural settings.

References

[1] Ames, C. (1992). Classrooms: Goals, structures, and student motivation. Journal of Educational Psychology, 84(3), 261–271.
[2] Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman.
[3] Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74.
[4] Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. Review of Educational Research, 65(3), 245–281.
[5] Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. Annual Review of Psychology, 53(1), 109–132.
[6] Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. American Psychologist, 48(2), 90–101.
[7] Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59–109.
[8] Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. American Psychologist, 56(3), 218–226.
[9] Frenzel, A. C., Goetz, T., Lüdtke, O., Pekrun, R., & Sutton, R. E. (2009). Emotional transmission in the classroom: Exploring the relationship between teacher and student enjoyment. Journal of Educational Psychology, 101(3), 705–716.
[10] Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112.
[11] Heift, T., & Schulze, M. (2007). Errors and intelligence in computer-assisted language learning: Parsers and pedagogues. Language Learning & Technology, 11(3), 68–86.
[12] Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. Review of Educational Research, 67(1), 88–140.
[13] Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. Cognitive Science, 36(5), 757–798.
[14] Kulik, J. A., & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. Review of Educational Research, 58(1), 79–97.

[15] Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. Teachers College Record, 115(3), 1–47.
[16] Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. Educational Psychology Review, 18(4), 315–341.
[17] Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. Educational Psychologist, 37(2), 91–105.
[18] Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68–78.
[19] Schaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. Journal of Happiness Studies, 3(1), 71–92.
[20] Shermis, M. D., & Burstein, J. (2013). Handbook of automated essay evaluation: Current applications and new directions. Routledge.
[21] Shute, V. J. (2008). Focus on formative feedback. Review of Educational Research, 78(1), 153–189.
[22] Slavin, R. E. (1995). Cooperative learning: Theory, research, and practice (2nd ed.). Allyn & Bacon.
[23] Smith, R. (2007). An overview of the Tesseract OCR engine. In Proceedings of the International Conference on Document Analysis and Recognition (ICDAR) (pp. 629–633).
[24] Stiggins, R. (2005). From formative assessment to assessment FOR learning: A path to success in standards-based schools. Phi Delta Kappan, 87(4), 324–328.
[25] VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4), 197–221.
[26] Wiliam, D. (2011). Embedded formative assessment. Solution Tree Press.
[27] Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. Educational Psychologist, 52(1), 17–37.
[28] Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, 17(2), 89–100.
[29] Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70.