Academic Education Publishing House
Higher Education and Practice Vol. 3 No. 2, 2026

Design and Effectiveness Evaluation of Classroom Evaluation Literacy Improvement Workshop for College English Teachers: A Practical Approach Based on Activity Theory

Na Zhao*, Ziting Niu
Kunming University, Kunming, Yunnan, China
**Corresponding Author*

Abstract: To address how insufficient assessment literacy constrains teaching quality, this study developed and implemented an activity theory-based intervention program to enhance teachers' evaluation literacy. Using Activity Theory, it examined contradictions within the evaluation system, such as between rigid administrative rules and needs for personalized feedback. A sequential explanatory mixed-methods design was employed. Firstly, a workshop intervention with modules on theory, skills, practice, and reflection was conducted. Pre-test and post-test questionnaires (N=32) quantified changes in evaluation knowledge, skills, and attitudes. Subsequently, interviews (N=15), classroom observations, and analysis of conflict logs qualitatively analyzed the resolution of systemic contradictions. Results showed: 1) The intervention significantly improved teachers' theoretical knowledge ($p<0.01$) and practical skills (case design success increased by 35%); 2) Activity theory analysis identified the core contradiction as between "standardized administrative requirements" and "differential student needs." The workshop mitigated this by guiding teachers to develop flexible rubrics and reconstruct teacher-student collaborative evaluation; 3) Teachers' feedback shifted from unidirectional judgment to bidirectional dialogue, activating students' Zone of Proximal Development (ZPD). This study confirms that systematic intervention guided by Activity Theory can effectively address bottlenecks in teacher evaluation literacy development, offering a theoretically grounded and practical paradigm for professional development and evaluation reform in higher education.

Keywords: Evaluation Literacy; Activity Theory; Workshop; Systemic Contradiction

1. Introduction

1.1 Problem Statement

Against the backdrop of connotative development in higher education and the student-centered teaching paradigm shift, classroom evaluation-as the core feedback mechanism-directly impacts talent cultivation effectiveness [1]. However, current university English teaching in China faces persistent structural challenges in evaluation, hindering teaching quality and students' comprehensive language abilities. These are summarized as three interconnected issues:

Firstly, structural imbalance in evaluation models. Traditional formative assessments have long dominated university English education, with the approach reducing teaching processes to mere exam preparation drills. Despite repeated policy advocacy for formative assessment, practical implementation often sees routine grades and classroom performance metrics treated as superficial formalities. Lacking scientific design, systematic documentation, and effective feedback mechanisms, these evaluations fail to integrate into the teaching cycle or serve as genuine learning facilitators [2]. This imbalance results in disconnection between evaluation systems and instructional practices, preventing timely diagnosis of learning needs and adjustment of teaching strategies.

Secondly, inefficient feedback mechanism. Feedback often remains confined to linguistic error correction, neglecting learners' cognitive processes and strategies. It tends to be generalized, one-way, and delayed, failing to reach students' Zone of Proximal Development (ZPD) [3]. Consequently, learners recognize errors without understanding corrective methods, leaving the feedback's educational potential

untapped.

Thirdly, the systemic deficiency in teachers' evaluation literacy. The root cause of these challenges lies in educators' inadequate evaluation literacy as primary subject of assessment systems. Evaluation literacy transcends basic scoring capabilities; it constitutes a comprehensive skill set encompassing evaluation ethics, theoretical knowledge, tool design, implementation techniques, and data interpretation [4]. Many college English instructors lack systematic evaluation training, demonstrate limited understanding of modern assessment theories, possess insufficient skills in designing diversified evaluation tools, and struggle to translate assessment data into actionable teaching improvement strategies [5]. This creates a vicious cycle where educators "want to evaluate but lack the skills, and when they do, the evaluations prove ineffective.

In summary, reform has shifted from "whether" to "how." The core bottleneck is teachers' insufficient evaluation literacy. Existing development programs often offer fragmented skill instruction rather than viewing evaluation as a complex activity system involving multiple interacting elements and contradictions [6]. Such piecemeal interventions yield limited, unsustainable outcomes.

Therefore, this study addresses a fundamental question: How can we transcend traditional training to develop a systematic intervention that diagnoses and resolves deep-seated contradictions in teacher evaluation practices, thereby substantially and sustainably enhancing college English instructors' evaluation literacy?

1.2 Research Objectives and Value

This study sets dual innovation objectives, with theoretical, practical and policy application values.

Theoretical Objectives and Contributions: It pioneers the application of Activity Theory to the development of college English teachers' evaluation literacy, conceptualizing evaluation as a dynamic contradictory system. By analyzing core elements such as subject, tools, rules and community and their inherent contradictions, it breaks through the single perspective of attributing competency deficiencies to individual skill gaps, reveals structural causes from the perspective of system interaction, provides stronger explanatory power and enriches the theoretical foundation for teacher development.

Practical Objectives and Innovation: Based on the diagnostic results of Activity Theory, it designs and verifies an integrated workshop model of "theory-practice-reflection", focusing on contradiction resolution and system reconstruction. It guides teachers to collaboratively develop innovative tools such as flexible rubrics, negotiate rule restructuring and optimize community collaboration, promoting the evolution of the evaluation practice system and establishing a paradigm for the transformation of teaching practice.

Policy and Application Value: Aligning with China's educational evaluation reform orientation and the requirements of the "Golden Course" initiative, it provides an operable framework with theoretical support for university administrators and faculty development centers. By improving teachers' evaluation literacy, it transforms evaluation into a catalyst for students' learning, cultivates students' higher-order thinking and integrated language abilities, and serves the goal of cultivating high-quality international talents [7].

2. Theoretical Foundations and Framework

2.1 Teacher Evaluation Literacy: Connotation Deconstruction and Disciplinary Expansion of the Three-Dimensional Models

Based on the integrative model proposed by Xu and Brown (2016), this study deconstructs teacher evaluation literacy into three interwoven and dynamically interactive core dimensions: knowledge, skills and ethical awareness, and conducts in-depth disciplinary adaptation in combination with college English teaching.

The knowledge dimension is the foundational layer, requiring teachers to master not only common assessment methods but also their theoretical underpinnings, including the cognitive psychology principles of formative assessment, language competency constructs and ethical standards for fair and transparent practice.

The skills dimension is the practical application layer, covering the ability to design evaluation tools, implement effective feedback strategies and interpret data to diagnose learning needs.

The ethical awareness dimension is the value core, guiding teachers to adhere to fairness and justice, pay attention to the impact of evaluation on different learners, use data responsibly and examine the power relations in the evaluation

system.

For college English teaching, this model needs two key disciplinary adaptations: first, cross-cultural communication assessment capabilities under the background of globalization, with the evaluation focus shifting from linguistic accuracy to cultural negotiation strategies and pragmatic appropriateness; second, technology-integrated assessment capabilities in the digital transformation of education, requiring teachers to use digital tools to improve evaluation accuracy and speciality while alerting to algorithmic biases and data ethics issues [8]. The evaluation literacy defined in this study is a comprehensive professional competency integrating theoretical knowledge, practical skills and value judgment, and the workshop intervention directly targets these disciplinary competency gaps. Therefore, the assessment literacy defined in this study is a comprehensive professional competency, integrating theoretical knowledge, practical skills, and value judgments within the complex context of college English teaching. The workshop intervention is designed to directly address these disciplinary competency gaps.

2.2 Activity Theory: Applicability Argument as a Systematic Framework for Contradiction Analysis

This study adopts Activity Theory as the core analytical framework to diagnose and solve the complex challenges of improving teachers' evaluation literacy. Originating from Vygotsky's cultural-historical psychology and developed by Engeström, Activity Theory is a powerful tool for analyzing human collective practices, conceptualizing them as dynamic systems driven by internal contradictions.

The fundamental proposition of Activity Theory is that any conscious human practice can be regarded as an activity system composed of six interrelated elements. The teacher classroom evaluation system includes: the subject (teachers conducting evaluation), the object (high-quality learning-enhancing assessment), tools (material tools such as rubrics and software, and psychological tools such as evaluation theories), rules (norms from school policies to classroom practices), the community (stakeholders such as students, colleagues and administrators), and division of labor (role and task allocation within the community).

The core advantage of this theory is that it regards contradictions as the driving force for system development and learning. Engeström (2001) divides contradictions into three levels: primary contradictions (conflicts within a single element, such as the tension between teachers' exam-oriented mindset and learning-facilitation philosophy), secondary contradictions (conflicts between different elements of the system, the focus of this study, such as the contradiction between rigid institutional rules and diversified evaluation criteria, and backward evaluation tools and developmental learning objectives), and tertiary contradictions (conflicts between the current activity system and potential advanced models).

The fundamental applicability of Activity Theory to this study lies in its systemic perspective, which places the insufficiency of evaluation literacy in the broader social and cultural system of institutional frameworks, tools and community dynamics, systematically identifies structural contradictions embedded in rules, tools and roles, and provides a precise blueprint for designing intervention programs targeting underlying systemic problems rather than just addressing superficial symptoms.

2.3 Integrative Framework: The Dual-Cycle-Driven Workshop Intervention Model

Building upon the aforementioned two theoretical pillars, this study establishes an integrated theoretical framework termed "Dual-Cycle-Driven" to guide workshop design and elucidate its anticipated mechanisms of action. The core proposition of this framework posits that the effective and sustainable development of teachers' evaluation literacy fundamentally constitutes a culturally-historical evolutionary process within their evaluation activity systems, driven by inherent contradictions.

Cycle I: The workshop serves as a "change laboratory," catalyzing conflict resolution and system reconstruction.

The workshop is designed as a micro "change laboratory". It begins by guiding teachers through self-reflection (e.g., analyzing their own assessment cases, maintaining conflict logs) and group discussions to identify core secondary contradictions within their evaluation systems (such as rule-object contradictions: the simplistic quantitative scoring required by administrative regulations versus the complex needs for

qualitative evaluation of students' cognitive processes). Subsequently, the workshop provides targeted interventions: at the tool level, it facilitates collaborative development of new assessment instruments (e.g., flexible rubrics combining "core indicators + personalized descriptions"); at the rule level, it organizes discussions on creating new practical spaces within existing institutional frameworks (e.g., delegating process evaluation detail formulation authority to course teams); and at the role division level, it designs innovative practical tasks to redefine professional roles (e.g., training students in peer evaluation practices and incorporating evaluation cycles). These interventions aim to directly address identified contradictions and drive adaptive changes in system elements.

Cycle II: Systemic evolution feeding back on the qualitative transformation and solidification of literacy.

The restructuring of activity systems will directly lead to a qualitative leap in teachers' evaluation literacy. When educators utilize new tools to resolve longstanding contradictions, their evaluation knowledge transforms from abstract understanding into contextualized practical expertise. When innovative division of labor models (such as teacher-student co-evaluation) are implemented, ethical considerations in assessment (including fairness and democracy) evolve from external regulations into shared community standards. Moreover, the innovative application of tools and rules inherently refines and enhances teachers' evaluation skills. Most crucially, an evolved, more adaptive activity system—such as teacher communities supporting innovative evaluations and repositories of shareable tools—provides sustainable ecological support for continuous professional development, equipping educators to address future challenges.

The theoretical framework is visually presented as follows:

(Figure description: This diagram depicts a dual-cycle model centered on the "Teacher Evaluation Activity System," represented by Engeström's triangular six-element diagram with lightning symbols indicating "contradictions" between components. The left "Cycle I: Intervention Cycle" arrow points to the system, labeled "Workshop Intervention: Contradiction Diagnosis, Tool Innovation, Rule Negotiation, and Division of Labor Restructuring." The right "Cycle II: Evolutionary Cycle" arrow extends outward from the system toward the peripheral "Three-Dimensional Model of Teacher Evaluation Literacy (Knowledge, Skills, Ethics)," annotated as "System evolution leads to competency qualitative transformation and solidification." The closed loop between cycles demonstrates how interventions drive system evolution, with upgraded systems subsequently laying foundations for higher-level interventions and competency development.)

The innovative aspects of this integrated framework include: Firstly, it explicitly shifts teacher professional development paradigms from "individual cognitive change" to "social practice system reconstruction." Secondly, it establishes a clear logical chain from theoretical diagnosis to practical interventions, ensuring each workshop design element has theoretical grounding and intended functions. Thirdly, it provides analytical coordinates for subsequent mixed-method evaluations—quantifying competency dimension changes while qualitatively tracking the evolution trajectories of activity system elements and contradictions.

3. Workshop Design: Intervention Approaches Based on Activity Theory

3.1 Needs Analysis: Identifying Systemic Contradictions as the Starting Point

To ensure intervention accuracy, this study conducted semi-structured interviews and questionnaires with college English teachers from 30 different universities before workshop design, identifying three typical systemic contradictions in evaluation practice, which became the direct targets of workshop module design: first, the "rule-subject" contradiction, with strong tension between the school's standardized, quantitative formative evaluation rules and teachers' needs to implement personalized, competency-oriented formative evaluation; second, the "tool-subject" contradiction, where traditional evaluation tools such as written tests and simple rating scales are unable to assess complex objectives such as language communication, intercultural awareness and critical thinking; third, the "division of labor-community" contradiction, with evaluation dominated by teachers and students, as core members of the learning

community, excluded, leading to one-way feedback, passive student participation and untapped potential for collaborative learning [9]. These contradictions collectively point to the core problem: teachers' evaluation intentions and behaviors are constrained by the systemic environment. Therefore, the workshop design is not merely knowledge infusion, but a "change laboratory" for teachers to collaboratively explore specific paths to resolve contradictions and reconstruct individual and collective evaluation activity systems.

3.2 Module Design: Targeted Intervention for System Elements

Based on the Activity Theory framework and preliminary needs analysis, this study designed a two-month workshop consisting of four progressive modules. Each module explicitly targets specific elements of the activity system, guiding teachers to collaboratively review and reconstruct their evaluation practices, in line with the core principles of formative interventions such as the Change Laboratory.

Module 1: Updating Evaluation Concepts and Rule Negotiation

Focusing on the interaction between rules and the teacher subject, it is not passive theory transmission but guides teachers to collectively examine and negotiate evaluation rules. Activities include analyzing national and institutional policy documents to distinguish core principles from operational constraints, transforming student-centered ethics into specific classroom rules through case discussions, and drafting localized *Classroom Evaluation Practice Convention* in groups. This drives teachers to transform from passive rule followers to active negotiators and co-constructors, addressing the "rule-subject" contradiction.

Module 2: Innovation in Assessment Tools and Skill Development

Directly targeting the tool element, it empowers teachers to develop mediating tools adapted to complex evaluation objectives. Adopting a practical "design-critique-iteration" model, teachers are divided into disciplinary groups to design preliminary evaluation tools such as oral presentation rubrics around real teaching objectives, which are then optimized through cross-group peer review. This not only cultivates practical tool design skills but also deepens teachers' understanding of how tools as mediators reshape evaluation practice, addressing the "tool-subject" contradiction.

Module 3: Scenario Simulation Practice and Division of Labor Restructuring

Simulating innovative division of labor to observe changes in community interaction, teachers participate in simulated teaching scenarios and take on the roles of students, observers and reflectors. For example, one group acts as students using peer evaluation tools developed in Module 2, while another observes and records the quality of interaction. A structured debriefing is then conducted to analyze changes in teacher-student and peer interaction under shared evaluation authority and teachers' new roles, preparing for the transformation of roles and relationships in the evaluation system and addressing the "division of labor-community" contradiction.

Module 4: Reflective Collaboration and System Evolution

As an integrative module, it guides teachers to collectively reflect on signs of contradiction resolution and plan for the continuous evolution of the system. The core activity is "Evidence-Based Group Diagnosis", where teachers present classroom video clips or student work samples from the implementation of new practices, and conduct in-depth analysis on the contradictions addressed, new tools and rules adopted, changes observed and remaining tensions. This structured collaborative reflection helps teachers synthesize fragmented attempts into a coherent understanding of the evolution of the activity system, establishing a sustainable cycle of improvement and expansive learning.

3.3 Implementation Process and Key Innovation

The workshop adopts a hybrid model of "offline focused discussions + online asynchronous collaboration" spanning two months, including four full-day offline workshops and continuous online collaborative tasks. Two key innovative designs in the implementation reinforce its theoretical orientation:

Conflict Log: Each participant is required to record conflict incidents, coping attempts and challenges in daily evaluation practice, which serves as both a personal reflection tool and raw material for collective diagnosis in Module 4, transforming abstract contradictions into tangible elements [10].

ZPD Collaborative Group: Fixed collaborative

teams are intentionally formed with teachers of diverse teaching experiences and disciplinary backgrounds. This heterogeneous composition creates a ZPD, enabling members to provide conceptual support and practical insights beyond individual literacy through social interaction, facilitating a more effective transformation of systemic perspectives.

In summary, the workshop design presents a clear activity-based intervention path: starting with identifying systemic contradictions in the evaluation system, it implements targeted interventions on core elements such as rules, tools and division of labor, and ultimately guides teachers to foster the collaborative evolution of the evaluation activity system in practice, thereby achieving profound and sustainable improvement of evaluation literacy.

4. Research Methods

4.1 Overall Study Design: Sequential Explanatory Mixed Methods

To comprehensively explore the impact of activity-based workshops on teachers' evaluation literacy and its underlying mechanisms, this study adopted a sequential explanatory mixed-methods design, consisting of two interrelated phases: the first phase uses quantitative methods to measure the overall effects and trends of the workshop intervention on various dimensions of teachers' evaluation literacy; the second phase uses qualitative methods to gain in-depth insights into the processes, contexts and mechanisms behind these changes, explaining the "how" and "why" of the observed transformations [11]. The two phases proceed sequentially, with the second phase's research questions, participant selection and data collection partially informed by the preliminary findings of the first phase, achieving data triangulation and in-depth understanding. This design aligns with the study's objectives: to confirm the overall effectiveness of the intervention (quantitative component) and reveal the micro-dynamics of complex activity system restructuring (qualitative component), ultimately providing a holistic understanding of teachers' professional development as a social and cultural process.

4.2 Study Participants

This study adopted a combination of purposive sampling and convenience sampling for participant recruitment. Through the network of university foreign language departments and faculty development centers, 32 college English teachers from three types of institutions (comprehensive universities, teacher-training colleges, and science and engineering universities) were selected as the research sample, all of whom participated voluntarily and signed informed consent forms. The sample is diverse in terms of teaching experience (2 to 25 years, with an average of 10.4 years), academic titles (lecturers and associate professors) and courses taught (general English, academic English and foundational English majors courses). All 32 teachers participated in the two-month workshop and completed pre-test and post-test questionnaires. For in-depth qualitative research, maximum difference sampling was used to select 15 teachers as core cases, with selection criteria including baseline assessment literacy levels, engagement intensity in workshop activities and institutional environmental variations, ensuring the qualitative sample captures diverse practical contexts and variation patterns to enhance the interpretative validity of the research findings.

4.3 Data Collection Tools and Process

Data collection was conducted throughout the pre-intervention, during-intervention and three months post-intervention phases of the workshop, using multiple tools to obtain multidimensional evidence.

4.3.1 Quantitative data collection: teacher evaluation literacy questionnaire

A self-designed College English Teacher Evaluation Competency Survey Questionnaire was used for pre-test and post-test administration, strictly developed based on Xu & Brown's (2016) integrative model and adapted for the disciplinary context. It consists of four themes: Evaluation Knowledge (understanding of formative assessment principles and language proficiency constructs), Evaluation Skills (ability to design rubrics and provide targeted feedback), Evaluation Ethics Awareness (attitudes toward evaluation fairness and student participation rights), and Evaluation Practice Frequency (frequency of using various assessment tools). The scale uses a 5-point Likert scale, and the questionnaire demonstrated good reliability (Cronbach's $\alpha > 0.85$) and validity (structure validated through exploratory factor analysis) in the pilot test. The pre-test was

conducted one week before the workshop, and the post-test one week after, to directly measure the immediate changes induced by the intervention.

4.3.2 Qualitative data collection: multi-level in-depth data

To capture the dynamic evolution of the evaluation activity system, four types of qualitative data were collected:

Semi-structured in-depth interviews: One-on-one interviews (60-90 minutes each) were conducted with 15 key teachers before and after the workshop. Initial interviews focused on existing evaluation practices, perceived contradictions and challenges, while follow-up interviews explored workshop experiences, new practices implemented, difficulties encountered and reflections on system changes, with interview outlines designed around the six elements of Activity Theory.

Classroom Observation and Teaching Product Analysis: 1-2 classroom observations were conducted for each designated teacher in the semester after the workshop, focusing on documenting evaluation practices. Meanwhile, teaching materials developed under the new evaluation framework (teaching program, assessment tasks, grading rubrics and student assignment comments) were collected and analyzed as direct evidence of evaluation tools.

Conflict Log: All 32 teachers were required to record conflict incidents in evaluation practice at least once a week during and for one month after the workshop, along with the context, personal reactions and reflections, providing real-time and context-specific data on the generation, manifestation and resolution of systemic conflicts.

Workshop process documentation: Full audio recordings of four offline workshops were made, and written outcomes of group discussions and collaborative tool design prototypes were collected, documenting real-time interactions regarding rule negotiation, tool innovation and division of labor restructuring.

4.4 Data Analysis Methods

Data analysis follows the sequential interpretive mixed-methods approach: quantitative data is analyzed first, and the results are used to guide in-depth qualitative analysis.

4.4.1 Quantitative data analysis

Questionnaire data were processed using SPSS 26.0 software. Descriptive statistics were used to present the overall characteristics of the sample and score distribution across various dimensions. Paired-sample t-tests were conducted to compare the total evaluation literacy scores and dimension-specific scores of the 32 teachers between pre-test and post-test, to determine the statistical significance of the immediate effects of the workshop intervention. Correlation analysis was also performed to explore the relationships between changes across different dimensions.

4.4.2 Qualitative data analysis

An iterative analysis method combining thematic analysis with Activity Theory coding was employed for qualitative materials including interview transcripts, observation notes, conflict logs, workshop records and instructional products [12]. The specific steps are: 1) Primary Coding (Open-ended): Repeated text reading to annotate statements related to evaluation practices, conflicts and changes; 2) Secondary Coding (Axial Approach): Categorizing and organizing primary codes using the six core elements of Activity Theory and their corresponding conflict types as the core coding framework; 3) Three-level coding (optional): Integrating all codes to identify core themes and mechanistic patterns explaining the development of evaluation literacy through systemic restructuring; 4) Case Analysis and Cross-Case Comparison: Constructing detailed case narratives for 15 focal teachers to describe the evolutionary trajectory of their individual evaluation activity systems, and conducting cross-case comparisons to identify commonalities and differences in system evolution paths among teachers of diverse backgrounds, enhancing the nuanced understanding of the conditions and boundaries of intervention mechanisms.

4.5 Reliability, Validity, and Ethics of the Study

To ensure research quality, multiple measures were implemented: quantitatively, pre-test was conducted to refine the questionnaire and guarantee its reliability and validity; qualitatively, strategies such as data triangulation, member verification (initial analyses shared with participants for accuracy confirmation) and peer review (regular discussions on coding and interpretation with the research team) were adopted to enhance credibility and reliability. All procedures strictly

adhered to academic ethics standards: institutional review board approval was obtained prior to implementation; participants were fully informed about study objectives, procedures and data usage with written informed consent obtained; personal information and institutional affiliations were hided; data was exclusively used for academic research and properly stored.

5. Study Findings

This study systematically analyzed the impact of Activity Theory-based workshop interventions on college English teachers' evaluation literacy through sequential explanatory mixed methods. Quantitative data revealed significant improvements across all dimensions of literacy, while qualitative data provided in-depth insights into the resolution and restructuring of contradictions within individual and collective evaluation systems. The key findings are presented as follows.

5.1 Quantitative Research Findings: Significant Improvement in All Dimensions of Assessment Literacy

To evaluate the immediate effectiveness of the workshop, pre-test and post-test of the Evaluation Literacy Questionnaire were conducted among 32 participating teachers. Data analysis was performed using paired-sample t-test, with results presented in Table 1.

Table 1. Paired Sample t-Test Results of Teacher Evaluation Literacy Questionnaire pre-test and-post-test (N=32)

Dimension	Pre-test Mean (M)	Pre-test Standard Deviation (SD)	Post-test Mean (M)	Post-test Standard Deviation (SD)	Mean Difference (MD)	t	p (two-tailed)
1. Evaluation Knowledge	3.21	0.58	4.05	0.49	+0.84	8.732	0.000
2. Evaluation Skills	2.89	0.62	3.92	0.54	+1.03	9.415	0.000
3. Ethical Awareness	3.75	0.51	4.18	0.43	+0.43	5.128	0.000
4. Practice Frequency	2.45	0.67	3.38	0.61	+0.93	7.889	0.000
Total Score	3.08	0.48	3.88	0.42	+0.80	10.246	0.000

*Note: The scale uses a 5-point rating system (1= Completely disagree, 5= Completely agree). *

Data indicate that after workshop intervention, teachers demonstrated statistically significant improvements in all four dimensions of evaluation literacy and the total score (all p-values <0.001). The Evaluation Skills dimension showed the largest improvement (MD = +1.03), indicating that the workshop was most effective in enhancing teachers' practical abilities in designing and implementing evaluation tools. The Evaluation Knowledge dimension also showed substantial improvement (MD = +0.84), reflecting deeper understanding of evaluation theories among teachers. The increase in practice frequency (MD = +0.93) indicates that teachers not only understand and are capable of implementing new assessment methods but are also more willing to experiment with them in the classroom. In contrast, while ethical awareness initially had a higher baseline level, it still showed significant improvement (MD = +0.43), suggesting that the workshop facilitated deeper reflection among teachers on ethical issues such as assessment fairness and student participation.

To further explore the correlations among changes across dimensions, we conducted correlation analysis on the differences between post-test and pre-test scores (i.e., change values), with results shown in Table 2.

Table 2. Pearson Correlation Coefficients between Change Values of Evaluation Literacy Dimensions (N=32)

	Knowledge change value	Skill change value	Ethical change value	Practical change value
Knowledge change value	1			
Skill change value	.712	1		
Ethical change value	.421*	.385*	1	
Practical change value	.654	.798	.512**	1

*Note: p < .05, ** p < .01, *** p < .001. *

Table 2 demonstrates a strong positive correlation between knowledge change value

and skill change value (r = 0.712, p <0.001), indicating that theoretical knowledge growth is closely linked to practical skill development. The highest correlation (r = 0.798, p <0.001) is observed between skill change value and practical change value, which strongly supports the core tenet of activity theory: skill development directly drives changes in the frequency and patterns of practical activities. Additionally, ethical change value shows moderate positive correlations with other dimensions, suggesting that value system updates and cognitive-emotional development are mutually reinforcing.

5.2 Qualitative research findings: Conflict Resolution and Restructuring in Evaluation Activity Systems

Qualitative data from interviews, observations and conflict logs revealed the micro-mechanisms by which the workshop intervention drove the evolution of teachers' evaluation activity systems through conflict resolution, yielding three core themes:

From Rule Constraints to Rule Negotiation: Initially, institutional rules were perceived as rigid constraints. The workshop's Rule Negotiation activity reshaped teachers' perspectives, guiding them to collaboratively reinterpret policies and develop compliant implementation methods that expanded the scope of "objective evidence" for evaluation. Follow-up reflections indicated a shift from viewing rules as limitations to seeing them as starting points for negotiation, with a marked decrease in rule-related complaints in conflict logs.

From Universal Tools to Contextualized Tools: A key contradiction was the lack of appropriate tools to assess complex literacy. Through collaborative tool design in Module 2, teachers created contextualized tools such as multi-dimensional rubrics for academic presentations and Collaborative Process Observation Forms for group projects. These innovations reshaped practice, making higher-order literacy measurable and providing a basis for concrete, targeted feedback.

From Teacher Monologue to Community Dialogue: The core conflict was evaluation as a teacher-dominated monologue. The workshop challenged this by restructuring the division of labor, guiding teachers to decentralize evaluation authority and involve students in co-developing scoring criteria and peer review. This shift transformed feedback from one-way teacher annotations to multi-directional, evidence-based dialogues, with teachers' roles changing from referees to coaches and facilitators. Conflict logs reflected this transformation, with early complaints about students ignoring feedback replaced by reflections on guiding constructive peer evaluation.

5.3 Integrated Analysis: The Synergistic Trajectory between System Evolution and Literacy Enhancement

The integration of quantitative and qualitative findings can clearly delineate the development trajectory driven by the workshop (see Figure 1).

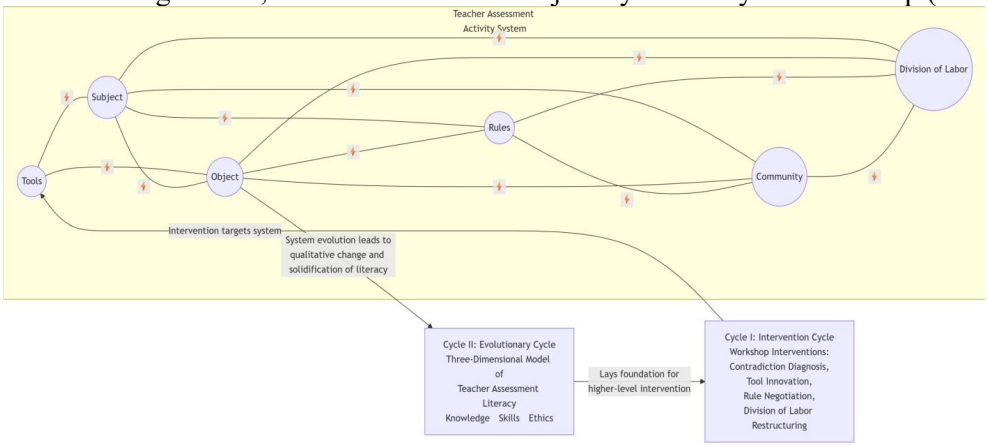

Figure 1. Collaborative Evolution Model of Teacher Evaluation Literacy and Activity System under Workshop Intervention

Figure 1 model demonstrates that the workshop does not directly impart Literacy, but achieves this indirectly through interactive transformations of core elements in the catalytic activity system. Quantitatively, significant skill enhancement corresponds to qualitative innovations in contextualized tools; increased practice frequency reflects routine application

within new rule frameworks and division-of-labor models; while elevated ethical awareness is embedded in a community-based dialogic feedback culture.

A teacher's (T15) summary profoundly reflects this systemic connection: "Previously, I thought evaluation literacy meant creating test questions and grading. Now I see it as a systematic project. I've learned how to collaborate with colleagues and students to build our course evaluation framework. My capabilities have grown within this new system, making them more solid and enabling continuous improvement."

In summary, the study's findings strongly demonstrate that activity theory-guided workshops focusing on resolving systemic contradictions can effectively drive comprehensive enhancement of college English teachers' evaluation literacy. The mechanism lies in guiding teachers to collectively negotiate evaluation rules, innovate assessment tools, and restructure evaluation responsibilities, thereby promoting adaptive evolution of their evaluation systems. Through this evolutionary process, teachers' knowledge, skills, ethical awareness, and practical behaviors develop. This provides an effective pathway for understanding and supporting teacher professional development through systematic restructuring.

6. Discussion

6.1 Theoretical Dialogue: Expanding the Application Scope of Activity Theory in Teacher Professional Development

The primary theoretical contribution of this study is its successful transformation of Activity Theory from a descriptive and analytical framework into a robust tool for designing and evaluating professional development interventions. By applying this framework to the specific domain of college English teachers' evaluation literacy, the study significantly broadens its application boundaries and enhances its interpretive depth.

Traditional teacher development research often views competency enhancement as linear changes in individual cognitive structures, focusing on measuring knowledge or belief shifts, but struggles to explain why many training programs are difficult to implement and sustain. This study introduces Activity Theory to address this challenge, and the findings strongly support Engeström's core argument that individuals' cognition and actions are deeply embedded in and constrained by their collective activity systems. The systemic contradictions identified in this study cannot be overcome through individual effort alone, and the workshop's effectiveness stems from its creation of a "Change Laboratory" that guides teachers to collectively examine, negotiate and resolve these contradictions through collaboration.

Furthermore, this study refines the micro-process of "contradiction-driven development" within Activity Theory. It finds that contradictions do not automatically transform into developmental momentum, and the workshop's structured interventions serve as critical catalysts, converting teachers' vague frustrations (primary contradictions) into concrete systemic issues (secondary contradictions) within the community. This study not only validates Activity Theory's applicability but also demonstrates how systemic contradictions can be consciously converted into core resources for professional learning, enriching the operational implications of the "contradiction-driven" theory.

6.2 Mechanistic Explanation: The Synergistic Evolution Chain of "Contradiction-Tool-Practice"

The most pivotal finding of this study is the identification of a clear synergistic evolution chain of "contradiction-tool-practice" underlying the enhancement of teachers' evaluation literacy, providing concrete empirical evidence for understanding the sociocultural essence of professional development.

The strong correlation between changes in evaluation skills and practice frequency directly reflects the quantitative characteristics of this evolutionary chain. Qualitative analysis provides vivid insights: when teachers face the dilemma of assessing complex literacy without appropriate tools (tool-object contradiction), the workshop guides them to collaboratively develop new tools, which immediately transform teaching practices. More importantly, the introduction of new tools often exposes deeper contradictions, prompting teachers to design new strategies or adjust rules, initiating another evolutionary cycle. This creates a self-reinforcing spiral: contradictions drive tool innovation, new tools generate new practices, and these practices reveal new contradictions or solidify established rules, ultimately stabilizing

and deepening newly acquired skills and knowledge.

In this chain, tools serve as indispensable intermediary agents. The contextualized tools independently developed by teachers in this study are not generic technical kits but practical wisdom products generated by teacher communities to address unique challenges. This endogenous nature ensures seamless alignment between tools and educators' intentions and local contexts, significantly enhancing their adoption and sustained use, and making tools bridges connecting individual cognitive development with systemic transformation.

6.3 Model Construction: Literacy and System Synergistic Evolution Model under Dual-Cycle Dynamics

Based on the preceding discussions, this study empirically validated and refined the initial theoretical framework proposed in Part 2, ultimately developing a more explanatory Competency and System Synergistic Evolution Model (see Figure 2). The model transcends the simplistic linear paradigm of training influencing individuals, who then modify practices, instead depicting a dynamic process driven by dual-cycle mechanisms.

The core of this model consists of two nested and mutually reinforcing cycles. Cycle A (System Reconstruction Cycle) represents the immediate operational level of workshop interventions, yielding an evolved local activity system with temporarily resolved contradictions. Cycle B (Competency Internalization Cycle) describes how teachers' literacy are substantially developed and solidified through practical application within this new framework. This aligns closely with Vygotsky's social and cultural learning theory: higher psychological functions (here referring to evaluative literacy) first emerge at the social interaction level (collective tool design and rule negotiation), before being internalized by individuals.

The quantitative and qualitative data from this study provide robust support for the model. The workshop (external intervention) successfully initiated Cycle A, manifested through rule negotiation, tool innovation, and division of labor restructuring in qualitative themes. Subsequently, teachers' enhanced evaluation practices within the new system (quantitative data showing increased "practice frequency") marked the beginning of Cycle B. The knowledge and skill improvements gained through these practices (significant growth across all dimensions in quantitative data) represented the operational outcomes of Cycle B. Teacher T15's observation "My capabilities have grown within this new system" serves as the most straightforward empirical validation of the proposition literacy internalized within the system. Feedback arrows in the model indicate that elevated internalized literacy enable teachers to become more acute conflict identifiers, thereby preparing conditions for deeper systemic evolution in subsequent cycles (Cycle A) and revealing sustainable pathways for professional development.

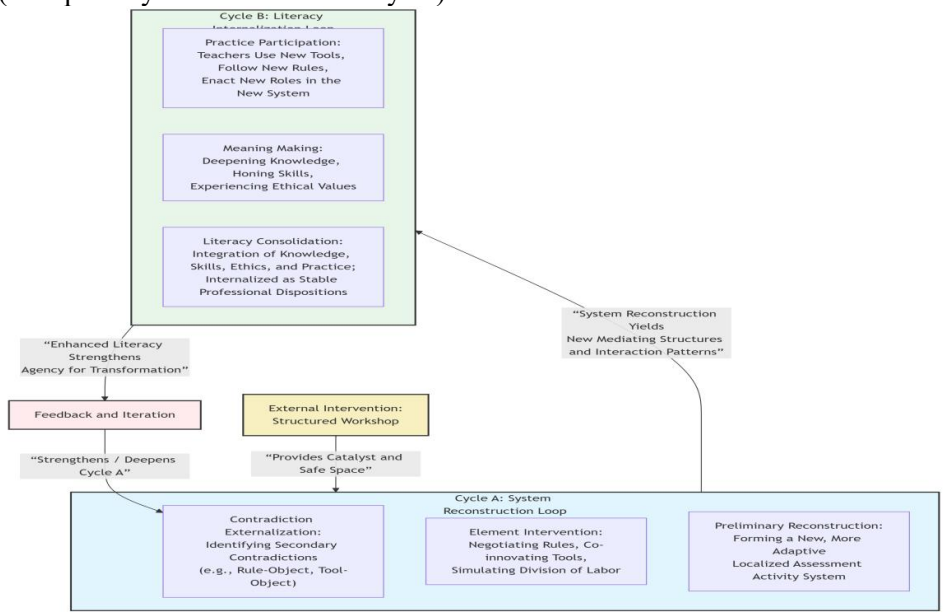

Figure 2. Synergistic Evolution Model of Teacher Evaluation Literacy and Activity System

6.4 Practical Implications: A Teacher Development Paradigm toward Support System Restructuring

The findings of this study provide clear practical implications for college English teacher development, curriculum evaluation reform and broader teaching reforms in higher education institutions:

Teacher development programs must transition from the knowledge renewal paradigm to a system reconstruction support paradigm. Future teacher training should adopt the workshop design model, transforming into a collaborative workplace for educators to resolve real-world systemic challenges, with objectives focusing on generating concrete, locally tailored solutions. Faculty development centers should serve as facilitators of "change laboratories" and provide methodological guidance.

The teaching management system should provide flexibility for teachers' rule negotiation. Schools and academic departments should review existing evaluation management systems, reduce rigid "one-size-fits-all" regulations while ensuring quality benchmarks and fairness, and empower teaching teams to develop customized formative assessment frameworks tailored to curriculum objectives and disciplinary characteristics.

Evaluation reform should prioritize infrastructure development for tool development. Universities or academic departments are advised to establish a Teaching Evaluation Tool Resource Library, collecting and promoting exemplary evaluation tools developed by faculty members. This reduces individual innovation costs and facilitates the sharing of practical expertise, fostering a professional community culture supporting evaluation reform.

6.5 Limitations and Future Directions of the Study

While this study offers valuable insights, it has several limitations that point to fruitful avenues for future research:

Sample size and speciality: The core sample of 32 volunteer teachers is relatively small and regionally concentrated, limiting the application of the findings. Future studies should expand sample sizes and conduct cross-regional or cross-cultural comparative analyses to test the model's applicability in different institutional contexts.

Long-term sustainability and application: This study measured immediate and short-term outcomes, and the consolidation of new practices and the stability of systemic restructuring require longer validation periods. Longitudinal studies with follow-ups one or more years post-intervention are essential to assess the durability of effects and the application of transformations to different courses.

Indirect impact on student learning outcomes: The study focused on teacher-level changes, with only indirect evidence of student engagement captured. Future research should incorporate long-term, systematic measurement of student learning outcomes to establish a complete evidence chain of teacher development → improved assessment practices → enhanced student learning.

Research methodology: The study relied partly on self-reported data such as conflict logs, which may be subject to bias. Future research could employ more diverse methods such as recording teachers' verbalized decision-making during assessment or conducting student focus groups to provide richer, multi-perspective validation of systemic changes.

In conclusion, this study demonstrates the efficacy and sociocultural mechanisms of Activity Theory-based interventions for enhancing teacher evaluation literacy, deepening theoretical understanding and providing a practical pathway for reform. Acknowledging these limitations will help advance research in this field, collectively promoting college English teaching evaluation in higher education toward a more scientific, equitable and effective new phase.

Acknowledgments

This work was supported by the Foreign Languages and Studies Press under the College English Research Project (Grant No. 2020111513).

References

[1] Black, P., & Wiliam, D. (1998). Inside the Black Box: Raising Standards through Classroom Assessment. Phi Delta Kappan, 80(2), 139-148.

[2] Guo X Q. (2026). Research on the Construction of a Student-Centered Classroom Teaching Quality Evaluation System for College English in Vocational

Undergraduate Universities. Knowledge Window (Teacher Edition), (01): 81-83.
[3] Vygotsky, L. S., & Cole, M. (1978). Mind in Society: The Development of Higher Psychological Processes. Cambridge: Harvard University Press, 15-18.
[4] Xu, Y., & Brown, G. T. L. (2016). Teacher Assessment Literacy in Practice: A Reconceptualization. Teaching and Teacher Education, 58, 149-162.
[5] Popham, W. J. (2009). Assessment Literacy for Teachers: Faddish or fundamental? Theory into Practice, 48(1), 4-11.
[6] Engeström, Y. (2001). Expansive Learning at Work: Toward an Activity Theoretical Reconceptualization. Journal of Education and Work, 14(1), 133-156.
[7] Tong D D. (2025). Research on the Construction of an Intelligent Teaching Evaluation System for College English. Journal of Jilin Agricultural Science and Technology College, 34(05): 95-99.
[8] Meng W T, & Pang X W. (2025). Research on the Construction and Application of an Intelligent Evaluation System for College English. Modern English, (19): 62-64.
[9] Pan W. (2026). Research on the Reform Path of College English Teaching Evaluation Empowered by Digital Intelligence. Overseas English, No.557(01): 112-114.
[10] Zhang Y X. (2025). Research on the Reconstruction of College English Blended Teaching Evaluation System in the "Intelligence Plus" Era. Modern English, (14): 26-28.
[11] Chen Y. (2017). An Investigation and Study on College English Teachers' Teaching Evaluation Literacy. Journal of Lanzhou Institute of Education, 33(10): 107-109.
[12] Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. Qualitative Research in Psychology, 3(2), 77-101.