

Design, Implementation, and Evaluation of a Needs-Based Language Assessment Literacy Training Program for Rural English Teachers in China

Qiao Huang, Khairul Azhar Bin Jamaludin

Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

Email: P145876@siswa.ukm.edu.my

Corresponding Author Email: khairuljamaludin@ukm.edu.my

DOI Link: <http://dx.doi.org/10.6007/IJARPED/v14-i3/26286>

Published Online: 03 September 2025

Abstract

The shift from assessment of learning (AoL) to assessment for learning (AfL) and assessment as learning (AaL) has heightened the demand for teacher language assessment literacy (LAL). Yet rural educators often have limited access to professional development in assessment. This study evaluated the LAL of rural English teachers in China (specifically in Shaoyang, Hunan Province), identified their urgent training needs, and implemented a contextually tailored LAL training program. A mixed-methods design was adopted, gathering data from 113 teachers through surveys, interviews, and classroom observations. Quantitative results (descriptive statistics, t-tests, and ANOVA) revealed a baseline mean LAL score of 16.42 out of 35, with highest knowledge in test methods and lowest in ethics, and significant differences across LAL dimensions ($\chi^2(6) = 52.37, p < .001, \eta^2 = .18$). Key needs included better skills in communicating results, using scoring rubrics, and conducting formative assessment. A six-week needs-based training intervention—combining face-to-face workshops, online modules, and one-on-one mentoring—led to significant improvements across all seven LAL dimensions. For example, the mean score for providing feedback increased from 2.31 to 3.15 on a 5-point scale ($t(112) = 11.26, p < .001, d = 1.06$). Qualitative feedback indicated that teachers valued the program's practical activities and local relevance, though engagement with online components was modest. These findings demonstrate that a needs-based, context-sensitive LAL training program can substantially enhance rural teachers' assessment literacy. The study offers a replicable model for teacher professional development in resource-constrained settings, with implications for policy and practice in narrowing urban–rural gaps in educational quality.

Keywords: Language Assessment Literacy, Rural Teachers, Teacher Training, Formative Assessment

Introduction

Language assessment literacy (LAL) is increasingly recognized as a cornerstone of effective teaching, student learning, and educational equity. Teachers' ability to design, administer, and interpret assessments not only enhances instructional quality but also ensures that students receive meaningful feedback and opportunities for growth. Strengthening teachers' LAL is therefore essential for narrowing achievement gaps and promoting fairness, particularly in rural and under-resourced contexts where professional development opportunities are limited and reliance on exam-oriented methods remains widespread. By equipping rural teachers with practical assessment skills, formative feedback strategies, and awareness of ethical practices, LAL development can empower them to provide more equitable learning opportunities and support inclusive, high-quality education.

Against this backdrop, language assessment has undergone a significant paradigm shift over the past two decades, evolving from an exclusive focus on assessment of learning (AoL) toward assessment for learning (AfL) and assessment as learning (AaL) (Iyanda, 2025; Schellekens et al., 2021; Yang & Xin, 2022). This transformation has expanded the role of assessment from a tool for measuring learning outcomes to an integral process that informs instruction, supports learner autonomy, and fosters ongoing improvement (Delavarpour & Safarnejad, 2024; Hidayat et al., 2023; Hidayati, 2024). Central to this transformation is the concept of LAL, which encompasses teachers' knowledge of assessment principles, their ability to design and implement effective assessment practices, and their awareness of the social, ethical, and contextual implications of assessment (Inbar-Lourie, 2016). International research consistently highlights LAL as a critical component of teacher professionalism, linking it directly to student achievement, equitable evaluation, and curriculum alignment (Giraldo, 2021; Tsagari & Armostis, 2025). However, many teachers still equate assessment primarily with exams and grades, reflecting a limited conception of assessment's role in learning (Aliakbari et al., 2023; Baidoo-Anu et al., 2023).

Despite growing recognition of LAL's importance, disparities in teacher training opportunities persist, particularly in rural and under-resourced regions in China (Jin, 2018; Liao et al., 2025). Rural English teachers often work in contexts where access to updated assessment theory, professional development programs, and collaborative learning opportunities is severely limited. This leads to reliance on traditional, exam-oriented assessment methods, with insufficient attention to formative feedback, diagnostic evaluation, and student self-assessment (Xu & Qiu, 2022; Zhu, 2024). Moreover, the rapid rollout of new curriculum standards in China has introduced more complex assessment requirements, yet many rural teachers lack the necessary training and confidence to meet these demands effectively (Peng & Tang, 2022; Yunfeng et al., 2023). These challenges are especially pronounced in less-developed areas such as Shaoyang in Hunan Province, where limited educational resources and scarce formal training exacerbate difficulties in translating assessment policy into classroom practice.

Existing studies have predominantly examined LAL in urban or well-resourced school contexts, leaving a significant gap in understanding the unique needs of rural educators. While some research has focused on pre-service teacher preparation, far fewer studies have investigated in-service training explicitly tailored to rural realities, such as large class sizes, constrained instructional time, and minimal access to assessment materials. Addressing this gap requires

not only measuring current levels of LAL but also designing interventions that are needs-based, context-sensitive, and grounded in practical classroom application.

In response to these challenges, the present study focuses on rural English teachers in Shaoyang, aiming to evaluate their LAL, identify urgent training needs, and implement a targeted professional development program to address those needs. Adopting a mixed-methods approach, the study provides empirical evidence of baseline LAL levels and assesses the impact of a needs-based training intervention on teachers' knowledge, skills, and principles. Specifically, the study was guided by the following objectives and research questions:

Diagnose current LAL levels and needs: What is the baseline language assessment literacy of rural English teachers, and what are their most prominent weaknesses and urgent needs in assessment knowledge and practice?

Inform training content and delivery: What training content and delivery formats do these teachers prefer, and how can a professional development program be designed to align with their authentic classroom challenges and workload constraints?

Evaluate training effectiveness: How effective is a needs-based, context-sensitive LAL training program in improving teachers' assessment literacy? In what areas does the program produce the most improvement, and which areas still require further support?

By addressing these questions, this work contributes to extending LAL research to rural in-service teachers in central China and offers a replicable model for enhancing assessment competence in resource-constrained environments. The findings carry practical implications for policymakers, curriculum designers, and teacher educators seeking to narrow the urban–rural divide in professional development opportunities. Ultimately, improving rural teachers' assessment competence is expected to promote more equitable and effective classroom assessment practices, benefiting student learning in line with the goals of AfL and AaL.

Literature Review

Conceptualizing Language Assessment Literacy

Language assessment literacy (LAL) refers to the knowledge, skills, and principles that enable teachers to design, implement, interpret, and use assessments effectively to support student learning (Tajeddin et al., 2022). LAL extends beyond technical competence in test construction to include an understanding of the social, ethical, and policy dimensions of assessment (Giraldo, 2022, 2025). Drawing on Vogt et al. (2024), LAL can be viewed as encompassing three interrelated domains: (1) knowledge of testing concepts and methods, (2) the skills needed to apply these concepts in varied educational contexts, and (3) an awareness of the consequences and implications of assessment decisions. In the language teaching profession, LAL is increasingly recognized as a core competency underpinning equitable, valid, and reliable evaluation practices (Gan & Lam, 2022; Kianinezhad, 2023; Kiran, 2023).

In recent years, the expansion of formative assessment practices has heightened the demand for teachers to adopt a more holistic view of LAL. This includes the ability to integrate assessment seamlessly into instruction, provide actionable feedback, and foster students' self-regulated learning (Kamran, 2024; Khursheed & Alwi, 2023). However, despite widespread acknowledgment of these competencies in international policy frameworks (e.g.

the Common European Framework of Reference for Languages), many teachers, particularly in rural contexts, struggle to apply them consistently due to limited professional development opportunities. Many rural teachers continue to rely on narrow assessment approaches focused on quizzes and examinations, lacking exposure to alternative strategies such as peer assessment or portfolio use.

International and Domestic Research on LAL

International research on LAL has documented diverse approaches to its development, ranging from pre-service teacher education modules to ongoing in-service training programs. Studies in Europe, North America, and Australia have emphasized collaborative learning communities, mentoring, and reflective practice as effective mechanisms for building assessment capacity (Jäppinen et al., 2016; Nolan & Molla, 2018; Zamiri & Esmaeili, 2024). In contrast, research in developing contexts often underscores structural barriers, such as inadequate resources, high teacher–student ratios, and minimal institutional support, which hinder the implementation of innovative assessment practices (Adeniyi et al., 2024).

In China, research on teacher assessment literacy has grown in parallel with curriculum reforms promoting competency-based education and formative evaluation (Deng et al., 2024). Urban teachers often have access to workshops, conferences, and professional networks, but rural teachers remain disadvantaged in exposure to updated assessment theories and best practices (Li et al., 2020; Wang et al., 2017). Yan (2015) and Zhao and Chen (2022) have revealed substantial gaps between the intended assessment policies outlined in national curriculum standards and the actual practices in rural classrooms. These studies point to a persistent reliance on traditional, summative assessments with insufficient emphasis on diagnostic and formative techniques. For example, teachers may focus on teaching to the test and assigning grades, while providing little constructive feedback to guide student improvement. Such mismatches highlight the need for professional learning that can bridge policy and practice in assessment, especially in under-resourced school settings.

Needs-Based Training for Rural Teachers

Needs-based professional development recognizes that effective training must be tailored to the specific challenges, contexts, and priorities of its participants (Edu, 2025). In the case of rural English teachers, these challenges often include large class sizes, limited access to teaching and assessment materials, and a lack of peer support. Research on needs-based LAL training suggests that programs should combine explicit instruction in assessment principles with opportunities for hands-on practice, peer feedback, and reflection (Bommanaboina & Bommanaboina, 2025). Such training is most impactful when it aligns closely with teachers' everyday instructional realities and is sustained over time rather than delivered as isolated one-off workshops.

In China's rural settings, needs-based training must also address broader systemic issues. These include integrating assessment into lesson planning, using formative assessment tools to support learning, and considering the ethical dimensions of testing (such as fairness and transparency in scoring and feedback). Few studies have comprehensively examined how a targeted LAL training program, informed by a detailed needs analysis, can tackle these multifaceted challenges. This gap underpins the significance of the present study, which aims to design, implement, and evaluate such a program within a rural Chinese context. Grounding

the training in local needs and conditions is expected to improve its relevance and effectiveness. Moreover, examining the outcomes of this intervention can provide evidence to guide future teacher training initiatives in similar low-resource environments.

Methodology

Research Design

This study employed a mixed-methods research design to investigate rural English teachers' LAL, identify their specific training needs, and evaluate the impact of a targeted professional development program. A pre- and post-intervention framework was implemented, combining quantitative survey data with qualitative interviews and classroom observations. This approach enabled both statistical measurement of changes in teachers' assessment literacy and an in-depth exploration of their perceptions and practices. By sequentially integrating needs analysis, training implementation, and evaluation, we ensured that the professional development intervention was data-driven and contextually grounded.

Research Participants

The participants were 113 in-service English teachers from rural primary and junior secondary schools in Shaoyang, Hunan Province, China (see Table 3.1). They were drawn from multiple township schools across the region, ensuring diverse representation in terms of school contexts and teacher backgrounds. Teaching experience among the participants averaged 12.4 years (SD = 6.8), with a range from 2 to 28 years. Their educational backgrounds and professional roles varied (including subject teachers and a few with additional responsibilities like grade head or teaching researcher), reflecting the typical staffing profile of rural schools. Most of the participants were female (approximately 74%, see Table 3.1), and about half taught at the primary level while half taught at the junior secondary level. Notably, most participants (81.4%) reported no prior formal training in language assessment – underscoring the need for professional development in this area. Those who had some previous exposure to assessment training (18.6%) indicated that it was limited in scope (often a brief workshop or lecture). Overall, the sample represented the population of rural English teachers in the region, who often have substantial classroom experience but lack specialized training in assessment.

Table 3.1

Demographic Profile of Participants (N = 113)

Variable	Category	n	%
Gender	Female	84	74.3%
	Male	29	25.7%
Teaching Level	Primary School	58	51.3%
	Secondary School	55	48.7%
Teaching Experience	1–5 years	19	16.8%
	6–15 years	52	46.0%
	16+ years	42	37.2%
Prior LAL Training	Yes	21	18.6%
	No	92	81.4%

Research Instruments

The primary instrument for quantitative data collection was a Language Assessment Literacy Questionnaire (Cui, 2023) adapted from established frameworks. This questionnaire covered seven key dimensions of LAL: (1) Test concepts and methods (e.g. knowledge of various test formats, item types, validity/reliability concepts), (2) Scoring and rating (skills in developing rubrics and scoring consistently), (3) Interpreting and using results (ability to analyse assessment data and inform instruction), (4) Formative assessment (use of assessment to provide feedback and guide learning), (5) Feedback and result communication (communicating assessment outcomes to students and stakeholders effectively), (6) Assessment ethics (awareness of fairness, bias, confidentiality, etc.), and (7) Policy and curriculum alignment (aligning assessments with curriculum standards and policies). The questionnaire items were statements rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree) indicating teachers' self-reported knowledge or practice in each area.

Content validity of the questionnaire was established through expert review by two assessment specialists who examined the relevance and clarity of each item. The instrument demonstrated high internal consistency: Cronbach's α for the overall scale was 0.91, and the α values for the seven subscales ranged from 0.82 to 0.88 (see Table 3.2). These reliability indices indicate that the questionnaire provided a coherent measure of various LAL components. In addition to the Likert-scale items, the survey included a short section for teacher background information (age, years of experience, grade level, and prior training) as summarized in Table 3.1.

Table 3.2

Reliability of the Language Assessment Literacy Questionnaire

Dimension	No. of Items	Cronbach's α
Test concepts and methods	5	0.85
Scoring and rating	4	0.84
Interpreting and using results	5	0.88
Formative assessment	5	0.86
Feedback and result communication	4	0.82
Assessment ethics	4	0.83
Policy and curriculum alignment	3	0.84
Overall scale	30	0.91

To gather qualitative insights, we developed a semi-structured interview protocol and an observation checklist. The interview protocol for the needs analysis phase included questions about teachers' current assessment practices (e.g., "How do you typically assess your students' learning in English?"), challenges faced ("What difficulties do you encounter in designing or grading assessments?"), and perceived needs for training ("What assessment skills or knowledge would you most like to improve?"). The post-training interview protocol focused on experiences with the training program (e.g., "Which parts of the training did you find most useful?"), changes in practice ("Have you tried any new assessment techniques in your classes since the training?"), and suggestions for further support. A total of 15 teachers were interviewed in the needs analysis stage (see Procedure below), and 6 teachers were interviewed after the training. Each interview lasted approximately 30–45 minutes and was conducted in Chinese (the participants' native language) for openness and accuracy, then transcribed and translated for analysis.

The classroom observation checklist was used to document specific assessment-related behaviors during English lessons. It included items such as “Uses a variety of question types (open-ended, objective, etc.) to gauge understanding,” “Provides feedback or discusses results with students,” and “Engages students in self- or peer-assessment activities.” Three teachers’ classes were observed (before and after the training, as examples) with each observation lasting one class period (~40 minutes). Field notes were taken to capture instances of assessment for learning practices versus purely summative practices. These qualitative instruments were intended to complement the survey results, offering contextual examples of how teachers assess and whether any shifts occurred after the intervention.

Research Procedure

The study was carried out in three stages: (1) Needs Analysis, (2) Training Intervention, and (3) Evaluation. In Stage 1 (Needs Analysis), the LAL questionnaire was administered to all 113 participating teachers to establish baseline assessment literacy levels and identify common areas of strength and weakness. The survey was conducted in person during a teacher professional development meeting in Shaoyang (with necessary permissions from school administrators), and an excellent response rate (91%) was achieved, yielding 113 valid responses. Quantitative needs analysis was supplemented by follow-up interviews with 15 volunteer teachers (selected to represent different schools and experience levels) to gain qualitative insights into specific assessment challenges and professional learning desires. Additionally, three teachers (who were among those interviewed) were observed in their classrooms and asked to share samples of their assessment materials (e.g., recent unit tests or quizzes they had designed). These in-depth cases illustrated typical assessment practices in the local rural context and helped validate the questionnaire findings.

In Stage 2 (Training Intervention), a six-week needs-based LAL training program was designed and delivered to address the most urgent needs identified in Stage 1. The training program was organized through the local teacher training center in Shaoyang in collaboration with the researchers. To maximize access and relevance, the program combined multiple components: an initial offline interactive workshop, subsequent online learning modules, and ongoing one-to-one guidance (mentoring). In the first week, an in-person workshop (one day, 6 hours) was conducted at a centrally located school, where trainers introduced key concepts (e.g., principles of formative assessment, designing rubrics) and engaged teachers in hands-on activities (such as critiquing test questions and practicing giving feedback) using examples drawn from the local curriculum. Over the next four weeks, teachers participated in online modules via a messaging app and web platform. These modules included short video lectures, reading materials, and practical assignments. For instance, each teacher had to draft a sample scoring rubric for an upcoming speaking task and share it in an online forum for peer review. Progress in the online portion was monitored, though, as expected in this context, not all teachers completed every module. Throughout the intervention, the researchers and local teacher trainers provided individualized guidance, either in person or via phone/WeChat, answering questions and coaching teachers on applying new strategies in their own classrooms. This personalized support was intended to help teachers bridge the gap between theory and practice. In total, 113 teachers were invited to the training; attendance at the initial workshop was high (approximately 100 teachers), while engagement in the voluntary online activities varied (on average about 60% actively participated). All participants, however,

received the same set of materials and were encouraged by their school leaders to implement what they learned.

In Stage 3 (Evaluation), the effects of the training program were assessed. Immediately after the six-week intervention, the LAL questionnaire was re-administered to the participants as a post-test (with 113 responses matched to their pre-test). This provided quantitative data on any changes in teachers' self-reported assessment literacy across the seven dimensions. To triangulate these results, post-training interviews were conducted with 6 teachers (randomly selected from those who completed the training, ensuring a mix of primary vs. secondary and varying years of experience). In these interviews, teachers reflected on what they had learned and how (if at all) their assessment practices were changing. Furthermore, the same 3 case-study teachers from Stage 1 were observed in their classrooms again about one month after the training to look for evidence of new assessment practices (such as using rubrics or giving more feedback). These follow-up observations and interviews provided qualitative evidence of the training's impact and captured any challenges teachers faced in implementation. All qualitative data were recorded with consent and later analysed alongside the survey results. Overall, this three-stage procedure ensured that the training program was grounded in actual needs and that its impact could be rigorously evaluated. Ethical approval for the study was obtained from the university's research ethics committee. Participants gave informed consent, and all data were kept confidential. During the study, teachers and school administrators were periodically briefed on preliminary findings (especially the needs analysis results, which informed the training content), fostering a collaborative and transparent research process.

Data Analysis

Quantitative data were analyzed using IBM SPSS Statistics (Version 27). Descriptive statistics (means, standard deviations, minima/maxima) were computed for LAL questionnaire responses to summarize teachers' baseline levels on each dimension. A paired-samples t-test was used to compare pre- and post-training scores for each LAL dimension and the overall LAL score, to determine whether any gains were statistically significant. Effect sizes for the pre-post differences were calculated using Cohen's *d* (for paired comparisons) to assess the practical significance of any improvements. Additionally, a one-way ANOVA was conducted to examine whether teachers' gains from the training differed by their years of teaching experience (comparing novice, mid-career, and veteran teachers, as defined in Table 1). The ANOVA tested for any interaction between experience level and improvement in LAL, using an alpha level of .05. Partial eta-squared (η^2) was reported as a measure of variance explained by group differences.

Qualitative data from interviews and observations were analyzed through thematic analysis. Interview transcripts and observation notes were transcribed and imported into NVivo 12 for coding. The researchers first read through all qualitative data to become familiar with the content. Then, an initial coding scheme was developed combining deductive categories (based on the research questions and LAL dimensions, e.g. "challenge in giving feedback," "pre-training practice: primarily tests," "post-training change: uses rubric") and inductive themes that emerged from the data (e.g. "teachers worry about time for new methods," "preference for face-to-face learning"). Two researchers independently coded a subset of the interview transcripts to refine the code definitions and ensure inter-coder agreement. After

achieving a high consistency in coding application, one researcher coded the remaining data with periodic peer checks. Key themes were then identified, such as teachers' perceived weaknesses (e.g. difficulty in designing listening test items), desired training features (e.g. "more examples, less theory"), and reported changes (e.g. starting to give written feedback, involving students in self-assessment). Representative quotations from teachers were extracted to illustrate these themes. Triangulation of quantitative and qualitative findings was conducted during interpretation: for instance, if the survey indicated low confidence in "feedback and result communication," the interviews were examined to see how teachers described their feedback practices and whether the training helped. This mixed-methods analytic strategy provided a rich, contextualized understanding of the data, strengthening the validity of the conclusions.

Findings

Baseline LAL Levels

The descriptive analysis of the baseline (pre-intervention) survey data revealed that the rural English teachers in our sample had moderate overall language assessment literacy, with an average total LAL score of 16.42 out of a possible 35 (which corresponds to a mean of 2.35 on the 5-point Likert scale; $SD = 4.13$). This level indicates a partial familiarity with assessment concepts and practices but also highlights considerable room for improvement. In practical terms, many teachers agreed with some basic statements about assessment (e.g. understanding different test question types), yet they tended to disagree or be unsure about more complex aspects (e.g. how to use assessment results to inform teaching).

There was substantial variation across the seven LAL dimensions assessed. As shown in Table 4.1, teachers performed best in "Test concepts and methods," with an average dimension score of $M = 2.78$ ($SD = 0.65$) on the 5-point agreement scale. This suggests a reasonable understanding of basic testing formats and item types (perhaps reflecting that most teachers are comfortable preparing traditional quizzes and exams). Conversely, the lowest performance was in "Assessment ethics," with $M = 2.12$ ($SD = 0.71$), indicating limited awareness of ethical considerations such as fairness, confidentiality of results, and avoiding biases in testing. Other dimensions showed intermediate scores: for example, "Scoring and rating" ($M = 2.54$), "Interpreting and using results" ($M = 2.48$), and "Policy and curriculum alignment" ($M = 2.39$) were all in the mid-2 range, suggesting that while teachers had some relevant knowledge, their ability to apply it in practice was inconsistent. The dimension "Formative assessment" ($M = 2.36$, $SD = 0.74$) and "Feedback and result communication" ($M = 2.31$, $SD = 0.69$) were also on the lower end, pointing to weaknesses in using assessment to support learning and in providing feedback to students or parents.

A chi-square test for goodness-of-fit confirmed significant differences in mean scores across the seven LAL dimensions, $\chi^2(6) = 52.37$, $p < .001$, with a large effect size (Cramer's V approximated by $\eta^2 = 0.18$). This indicates that certain areas of LAL are markedly more developed than others among these teachers. Knowledge of basic test formats outstripped skills in formative practices and ethical understanding. These baseline findings clearly validated the need for targeted professional development focusing on the dimensions that scored lowest – notably assessment ethics, feedback/result communication, and formative assessment. Improving these areas became a priority for the training intervention. It was also evident that even the higher-scoring areas could be strengthened; for instance, although

“Test methods” was highest, a mean of 2.78/5 still reflects only a modest level of confidence. Overall, the needs analysis demonstrated an uneven LAL profile, with critical gaps that the subsequent training was designed to address.

Table 4.1

Descriptive Statistics of Baseline LAL Scores by Dimension (N = 113)

LAL Dimension	M	SD	Min	Max
Test concepts and methods	2.78	0.65	1.40	4.00
Scoring and rating	2.54	0.70	1.20	4.20
Interpreting and using results	2.48	0.68	1.00	4.10
Formative assessment	2.36	0.74	1.00	4.00
Feedback and result communication	2.31	0.69	1.00	4.10
Assessment ethics	2.12	0.71	1.00	3.80
Policy and curriculum alignment	2.39	0.72	1.00	3.90
Overall LAL score (sum)	16.42	4.13	8.00	27.20

Pre- and Post-Training Changes

Following the implementation of the six-week needs-based LAL training program, participants demonstrated statistically significant gains across all seven LAL dimensions. Figure 1 (see Table 4.2 for exact values) illustrates the pre- vs. post-training mean scores by dimension. The improvements were evident in the raw scores and were confirmed by paired t-tests for each dimension (using $\alpha = 0.05$ with Bonferroni adjustment for multiple comparisons).

The largest improvement occurred in “Feedback and result communication.” Teachers’ mean rating on this dimension increased from 2.31 (pre) to 3.15 (post), a gain of $\Delta M = 0.84$. This change was highly significant, $t(112) = 11.26$, $p < .001$, with a Cohen’s $d = 1.06$, indicating a large effect. This suggests that the training’s emphasis on how to deliver constructive feedback (including role-playing feedback sessions and creating simple progress charts for students) had a strong impact on teachers’ skills and confidence in this area. Many teachers moved from “disagree” toward “agree” on statements such as “I know how to give students specific suggestions for improvement based on assessment results,” reflecting newly acquired strategies for result communication.

Similarly, the dimension of “Formative assessment” showed a substantial improvement (pre $M = 2.36$ to post $M = 3.15$, $\Delta M = 0.79$, $t(112) = 10.14$, $p < .001$, $d = 0.95$). This indicates enhanced capacity to design and use assessments for learning, such as incorporating informal checks during lessons or using assessment information to adjust teaching. Teachers’ qualitative comments also supported this gain; for example, one teacher noted in an interview, “I started giving short quizzes with no grades, just to see who needs help, and I think it’s working.” The training’s focus on formative techniques (like exit tickets and peer assessment exercises) seems to have been translated into practice.

Interestingly, even areas that were relative strengths at baseline saw significant gains. “Test concepts and methods,” which had the highest pre-score, still improved from 2.78 to 3.39 ($M = 0.61$, $t(112) = 9.28$, $p < .001$, $d = 0.87$). This suggests that the training benefited teachers across the board, reinforcing foundational knowledge and introducing new test design skills. For instance, teachers reported learning how to craft listening and speaking tasks, which are

areas not traditionally emphasized, thus broadening their assessment toolkit despite already feeling comfortable with basic testing.

All other dimensions (Scoring, Interpreting Results, Assessment Ethics, Policy Alignment) also showed significant pre–post improvements (all $p < .001$, see Table 4). Of note, “Assessment ethics” increased from 2.12 to 2.89 ($\Delta M = 0.77$, $t(112) = 10.08$, $p < .001$, $d = 0.95$). While the absolute post-training level for ethics is still moderate, the jump indicates that teachers gained awareness of important principles (such as transparency in grading criteria and avoiding favoritism). They likely benefited from training discussions on ethical dilemmas and the introduction of simple guidelines (e.g., not using test scores publicly to rank students, a practice some admitted doing before).

Overall, the consistently large effect sizes for the gains (Cohen’s d ranging roughly from 0.87 up to 1.06 for individual dimensions, and $d = 1.16$ for the overall score) underscore the practical significance of the program. The overall LAL score (sum of all items) rose from an average of 16.42 to 21.15 out of 35 (a 28.8% increase), which was highly significant ($t(112) = 12.37$, $p < .001$). These results demonstrate that a needs-based, context-sensitive training intervention can effectively enhance teachers’ assessment literacy in a relatively short period. The gains were most pronounced in the domains that were initially weakest (feedback, formative use of assessment, ethics), indicating that the training successfully targeted the right areas. At the same time, improvements in traditionally stronger domains (test design, scoring) suggest a comprehensive uplift in teachers’ competencies.

Table 4.2

Pre- and Post-Training LAL Scores (N = 113)

Dimension	Pre M	Pre SD	Post M	Post SD	M	t (112)	p	Cohen’s d
Test concepts and methods	2.78	0.65	3.39	0.61	0.61	9.28	< .001	0.87
Scoring and rating	2.54	0.70	3.28	0.68	0.74	10.45	< .001	0.98
Interpreting and using results	2.48	0.68	3.17	0.66	0.69	9.82	< .001	0.92
Formative assessment	2.36	0.74	3.15	0.71	0.79	10.14	< .001	0.95
Feedback & result communication	2.31	0.69	3.15	0.65	0.84	11.26	< .001	1.06
Assessment ethics	2.12	0.71	2.89	0.68	0.77	10.08	< .001	0.95
Policy & curriculum alignment	2.39	0.72	3.12	0.69	0.73	9.94	< .001	0.93
Overall LAL score	16.42	4.13	21.15	3.98	4.73	12.37	< .001	1.16

Differences by Teaching Experience

An analysis was conducted to examine whether the training’s effectiveness varied according to teachers’ years of teaching experience. Participants were grouped into three categories by experience level: novice (1–5 years, $n = 19$), mid-career (6–15 years, $n = 52$), and veteran (16+ years, $n = 42$) (as previously described in Table 3.1). For each teacher, an “overall gain score” was calculated (their post-training total LAL score minus pre-training total score). A one-way ANOVA was then performed with experience group as the independent variable and LAL gain as the dependent variable.

The ANOVA indicated no statistically significant differences in overall LAL improvement among the three experience groups, $F(2, 110) = 1.42$, $p = .247$, $\eta^2 = .025$. In other words, novice teachers, those with moderate experience, and very experienced teachers all benefited similarly from the program. Table 4.3 presents a summary of the ANOVA results. While there were minor numerical differences (the veteran teachers showed a slightly higher mean gain than novices, for instance), these were not large enough to be meaningful or to reach significance. The effect size η^2 of 0.025 suggests that only 2.5% of the variance in gain scores could be attributed to teaching experience, which is negligible.

Table 4.3

ANOVA Results for LAL Gain Scores by Teaching Experience

Source	SS	df	MS	F	p	η^2
Between Groups (experience)	2.84	2	1.42	1.42	.247	.025
Within Groups	110.00	110	1.00			
Total	112.84	112				

This finding suggests that the LAL training program was equally beneficial for teachers at all career stages. Novice teachers (who might have been expected to gain more, being “blank slates”) improved substantially, but so did their more experienced counterparts. Conversely, veteran teachers (who might be presumed to resist new methods or already know a lot) showed as much improvement as less experienced teachers. This result reinforces the notion that effective professional development can yield positive outcomes regardless of prior teaching experience. All teachers, whether in their first few years or decades into their career, appeared to learn and apply new assessment concepts when given a structured and relevant training opportunity. It challenges any assumption that only younger teachers are open to change, or that only older teachers struggle with assessment—rather, LAL gaps span all groups, and all groups responded well to training. This has important implications: it suggests that capacity-building in assessment literacy should not be limited to novices or assumed unnecessary for veterans; a needs-based approach can uplift practitioners across the board in a rural teaching community.

(No significant differences by other background factors were detected either. For instance, an independent t-test comparing those with vs. without prior LAL training found no significant gap in baseline scores or in gains, reinforcing that prior one-off workshops had minimal lasting impact.)

Discussion

This study provides compelling evidence that a needs-based, context-sensitive LAL training program can significantly enhance rural English teachers’ assessment literacy. The substantial gains observed across all seven LAL dimensions indicate that targeted professional development can fill both conceptual and practical gaps even in resource-constrained settings. Notably, our results align with prior research showing that rural teachers tend to have fragmented LAL, stronger in basic testing techniques but weaker in formative assessment, feedback, and ethics (Berisha et al., 2024; Giraldo, 2021; Pastore, 2023). In our sample, the lowest baseline scores were indeed in assessment ethics and feedback, mirroring international findings that these domains are often underdeveloped (Coombe et al., 2020; Kianinezhad, 2023). After the training, those areas showed the greatest improvement (effect sizes > 0.95), demonstrating the efficacy of explicitly addressing neglected aspects of LAL.

rather than focusing solely on test design. This outcome resonates with Vogt et al. (2024) argument that teachers need to reconceptualize assessment as a two-way dialogue to support learning, and with Giraldo (2021) observation that practice-oriented workshops can help teachers integrate formative assessment into daily instruction. In short, when training is hands-on and relevant to teachers' classroom context, even long-standing weaknesses in assessment literacy can be rapidly improved.

Another important finding is the program's uniform effectiveness across teachers of varying experience levels. Both novice and veteran teachers benefited comparably from the training, challenging the assumption that only new teachers need assessment training. This result supports the view that professional growth stems from meaningful learning opportunities rather than years of service. In line with Lave and Wenger (2001) situated learning theory, even experienced teachers expanded their assessment techniques when engaged in a new community of practice. Some veteran teachers in our study acknowledged blind spots (for example, not having considered test fairness before) that were addressed through the training, reinforcing the idea that LAL is multidimensional and never fully "complete" even for seasoned educators. From a theoretical standpoint, our findings confirm that LAL comprises cognitive, practical, and ethical components (Giraldo, 2018; Giraldo & Quintero, 2019) which must be developed holistically. The uneven pre-training profile of participants, high knowledge of test formats but low confidence in ethical and formative practices, underscores the need to treat LAL as an integrated competency rather than a single skill. Encouragingly, the significant improvements across knowledge, skills, and principles after the intervention suggest that a comprehensive training approach can build capacity in all dimensions, aligning with models that emphasize knowledge, skills, and values as interrelated facets of teacher assessment literacy (Chan & Luk, 2022; Pastore, 2023).

The rural context of this study offers additional insight into the role of environment in developing LAL. Consistent with Yan (2015) perspective, teachers' assessment practices were influenced by contextual constraints such as large class sizes and an exam-driven school culture. Before training, many participants conceptually understood the value of formative assessment but felt unable to implement it due to these pressures. By situating the training in their local reality, using examples from their classrooms and acknowledging challenges like limited class time, the program helped bridge the gap between knowing and doing. Teachers reported adopting new practices (e.g. giving more formative feedback) and even shifting their beliefs about assessment. Several noted that they no longer equate assessment solely with testing, but see it as a broader, continuous process to support learning. This change in mindset is significant: it shows that improving LAL in under-resourced settings involves not just building skills but also reshaping teachers' perceptions of assessment's purpose. In sum, our study expands the scope of LAL research to a less-developed rural context, highlighting that context is integral to teachers' professional learning. Lack of exposure to professional communities and resources in places like Shaoyang can stagnate certain dimensions of LAL, so interventions must address not only individual knowledge but also the support systems and communities in which teachers operate.

From a practical perspective, the findings carry several implications for teacher development programs. First, they underscore the value of conducting a thorough needs analysis before designing training. Rather than offering generic, one-size-fits-all workshops, professional

development should target the specific gaps identified among the teachers. In our case, diagnosing teachers' LAL revealed critical deficits in providing feedback and understanding assessment ethics, which guided the content of the training. Addressing these genuine needs made the program highly relevant and effective; for instance, without the needs assessment, we might have overemphasized test construction, an area where teachers were already relatively comfortable and failed to allocate enough time to feedback strategies. Systematically aligning training content with teachers' actual needs ensures that professional development efforts yield maximum impact. Second, the success of the blended training approach suggests that combining delivery modes can enhance learning outcomes. Our program integrated face-to-face workshops, online modules, and one-on-one mentoring, allowing teachers to learn new concepts, practice them with guidance, and receive feedback on implementation. Teachers particularly valued the interactive in-person sessions for their hands-on practice and immediate feedback, reflecting best practices that emphasize active learning opportunities (Chandran et al., 2021). At the same time, we found that online components alone had limited engagement, implying that simply providing digital resources is insufficient in this context. Effective professional development for rural teachers may require scaffolding online learning with clear expectations, incentives, and technical support to ensure teachers remain engaged. Additionally, incorporating assignments that prompt immediate classroom application, such as trying a new formative assessment technique and reflecting on it, helped teachers transfer training to their daily practice, reinforcing the training's practical impact.

Equally important are the strategies to motivate and support teachers throughout the professional development process. Our observations indicate that many rural teachers were initially unaware of specific shortcomings in their assessment practices. The act of self-evaluation, through questionnaires and interviews, served as a wake-up call, helping teachers recognize gaps (e.g. "I rarely give feedback beyond scores") and thus priming them to learn. Professional development programs should consider building in a self-assessment or reflection phase at the start, so that teachers develop a sense of need and ownership over their learning. Furthermore, training content must be authentically contextualized to teachers' local realities. Participants responded enthusiastically to examples and case studies drawn from rural classrooms like their own, whereas they had found previous, more theoretical workshops irrelevant to their situation. This suggests that teacher training in rural areas should be designed and delivered with a deep understanding of the local context, addressing concrete issues such as multi-level classes or exam pressures, rather than importing generic solutions. Another practical insight is the importance of incentives and recognition. Given heavy workloads and limited extrinsic motivation, rural teachers are more likely to invest effort in training if it is tied to tangible benefits. In our program, modest steps such as scheduling sessions during convenient times, providing a stipend, and highlighting how training achievements could count toward professional portfolios helped encourage participation. Education authorities might further motivate teachers by offering formal recognition, certificates, credits, or points for promotion, for those who engage in extensive professional learning. Such incentives can signal that developing skills like assessment literacy is an expected and rewarded part of a teacher's role. Finally, the need for ongoing support after the initial training is paramount. Without follow-up, teachers often revert to old habits once the initial enthusiasm fades. Our inclusion of follow-up observations and mentoring, although brief, was a step in the right direction, and teachers indicated a desire for more

continuous support. Establishing communities of practice or regular peer support meetings, even virtual ones given to remote locations, can help sustain the momentum of change. For example, forming a county-wide teacher discussion group on assessment or scheduling periodic coaching sessions would provide teachers with forums to share experiences, troubleshooting challenges, and keep each other accountable. In summary, effective teacher training, especially in rural settings, should be needs-driven, locally relevant, incentive-aligned, and sustained. Adopting these strategies can amplify the impact of professional development, leading not only to improved LAL but potentially to enhancement in other areas of teaching competence as well.

However, several limitations of this study should be acknowledged when interpreting the results. First, the research was conducted in a single region, Shaoyang, Hunan Province, which may limit the generalizability of the findings to other contexts. Rural schools in different provinces or countries vary in resources, policies, and challenges, thus similar interventions elsewhere might yield different outcomes. Future studies should examine needs-based LAL training in multiple regions to see if our results hold broadly and to identify any context-specific factors that influence effectiveness. Second, our evaluation captured only the short-term impact of the training. We measured teachers' knowledge and practices immediately after the program, but it remains uncertain whether these gains will be sustained over time. It is possible that without continued support, some teachers might regress to previous habits after several months. Longitudinal research, for example follow-up after 6–12 months, is needed to determine the durability of the improvements and whether enhanced teacher LAL ultimately translates into better student outcomes. Third, there is a potential for self-report bias and the Hawthorne effect. Teachers knew they were part of a study and training initiative, which may have motivated them to temporarily change behavior or give more positive self-assessments. We tried to mitigate this by including classroom observations and assuring participants that this was not an evaluation of their performance. While observations did confirm some changes, such as a teacher using rubric for the first time, we cannot be certain that all reported changes in practice were fully realized or will persist. Incorporating student feedback or more objective performance tasks in future research could provide additional validation of teacher-reported gains. Additionally, our qualitative data was relatively limited: only a subset of teachers were interviewed and only a few classes observed, which might not capture the full variability of teachers' experiences. A more extensive qualitative follow-up, such as repeated interviews or teacher journals, would deepen understanding of how teachers integrate new assessment techniques over time. Finally, because the intervention combined workshops, online learning, and mentoring, we are unable to isolate which component was most impactful. Given the low uptake of the online modules, it seems likely that the face-to-face and mentoring elements were critical, but future studies could experiment with different training models, for instance comparing blended versus online-only approaches, to optimize cost-effectiveness. Despite these limitations, the study offers valuable proof of concept: even in a challenging rural context, a carefully designed, needs-driven training program can achieve meaningful improvements in teachers' assessment literacy. These insights should inform ongoing efforts to support rural educators, while recognizing that further research is needed to fully understand long-term impacts and to refine the model for broader application.

Conclusion and Recommendations

This study evaluated the language assessment literacy of rural English teachers in Shaoyang, China, identified their most pressing needs, and implemented a tailored professional development program to address those needs. The findings reveal that these teachers had uneven LAL profiles, they were relatively confident in traditional test design and methods, but had notable deficits in formative assessment, providing feedback, and assessment ethics. After the needs-based training intervention, teachers' knowledge and skills improved significantly across all seven LAL dimensions, with the largest gains in the areas that were initially weakest, feedback, formative use of assessment, and ethics. This demonstrates that a context-specific, targeted training program can substantially build teachers' assessment competencies. Importantly, the benefits were observed for both less experienced and veteran teachers, indicating that even experienced teachers can adopt new assessment practices when given appropriate support. In other words, professional learning in assessment should not be limited to novices, with the right training, teachers at any career stage can strengthen their LAL.

These results also illustrate that contextually adapted training can be highly effective in resource-limited settings. By grounding the program in local needs and conditions, we enabled teachers to directly apply new strategies in their classrooms, helping to bridge the gap between urban and rural educational quality. Participants not only gained new theoretical insights but also began changing their day-to-day assessment practices, as evidenced by their feedback and our observations. They especially valued the practical focus of the training and the consideration of their workload and challenges, which bolstered their motivation to implement what they learned. This underscores that effective teacher development in rural areas must prioritize relevance and support, teachers are more likely to embrace new methods if the training aligns with their real-world classroom context and if they receive encouragement, such as recognition or peer support, to sustain these changes. In short, improving assessment literacy in such settings requires attention to both content and conditions, delivering useful strategies and fostering an environment that motivates continuous improvement.

Based on the study's conclusions, several recommendations can be made to strengthen teacher training in assessment. Education authorities and professional development providers should institutionalize a needs analysis as a precursor to any training program, so that the content can target genuine gaps in teachers' knowledge and practice, for example focusing on how to give effective feedback and ensure fairness in assessment, not only on test construction. Training curricula and teacher standards should explicitly include these often-neglected competencies, for instance emphasizing ethical assessment practices and feedback skills as core teaching competencies, to signal their importance. Teacher training programs should adopt blended, practice-oriented learning approaches, for example combining interactive workshops, coached practice opportunities, and online resources to reinforce learning, while ensuring the training activities closely relate to teachers' everyday work. It is also critical to ensure contextual relevance in professional development for rural teachers. Using local examples, addressing region-specific challenges, and delivering training on-site or by trainers familiar with the local context can make the experience more relatable and impactful. Furthermore, policymakers and school leaders should provide incentives and ongoing support to encourage teachers to apply new assessment techniques. This might

include offering certificates or credits for completed training, recognizing improvements in assessment practice in teacher evaluations, and establishing follow-up support systems such as mentoring or teacher learning communities. By implementing these strategies, stakeholders can improve the effectiveness of teacher professional development and help narrow the urban–rural gap in assessment literacy. As teachers become more assessment-literate, their assessments are likely to be fairer, more aligned with learning goals, and more useful for guiding student learning, ultimately contributing to better educational outcomes. Finally, although this intervention proved successful in the short term, further research should examine its long-term impact and scalability. Longitudinal studies tracking these teachers would determine whether the improvements in LAL are sustained over time and whether they translate into enhanced student learning, for example through improved student performance or more positive student feedback on assessment practices. Comparative studies across different regions or educational contexts would also be valuable to see if similar needs-based training programs yield comparable results and to identify any necessary adaptations for different settings. Overall, this study demonstrates the potential of a needs-driven, context-sensitive training model to substantially raise teachers' assessment literacy in a rural context. With continued efforts to refine and support such initiatives, we move closer to an education system in which all teachers, rural and urban alike, are equipped to use assessment as a powerful tool to enhance student learning.

References

- Adeniyi, I. S., Al Hamad, N. M., Adewusi, O. E., Unachukwu, C., Osawaru, B., Onyebuchi, C., & David, I. (2024). Educational reforms and their impact on student performance: A review in African Countries. *World Journal of Advanced Research and Reviews*, 21(2), 750-762.
- Aliakbari, M., Yasini, A., & Sadeghi, S. (2023). Iranian EFL Teachers' Classroom Assessment Practices: Discrepancy between Theory and Practice. *International Journal of Language Testing*, 13(2), 149-169.
- Baidoo-Anu, D., Rasooli, A., DeLuca, C., & Cheng, L. (2023). Conceptions of classroom assessment and approaches to grading: Teachers' and students' perspectives. *Education Inquiry*, 1-29.
- Berisha, F., Vula, E., Gisewhite, R., & McDuffie, H. (2024). The effectiveness and challenges implementing a formative assessment professional development program. *Teacher Development*, 28(1), 19-43.
- Bommanaboina, L. D., & Bommanaboina, R. D. (2025). Enhancing ESL Preservice Teachers' Formative Assessment Skills: Developing Reflective Analytic Rubrics for Language Assessment Tasks. *International Journal of Interdisciplinary Social & Community Studies*, 20(1).
- Chan, C. K., & Luk, L. Y. (2022). A four-dimensional framework for teacher assessment literacy in holistic competencies. *Assessment & Evaluation in Higher Education*, 47(5), 755-769.
- Chandran, K. B., Haynie, K. C., Tawbush, R., & Wyss, J. M. (2021). Effectively adapting and implementing in-person teacher professional development to a virtual format. *Journal of STEM outreach*, 4(3), 10.15695/jstem/v15694i15693. 15612.
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia*, 10(1), 3.
- Cui, T. (2023). *Design, practice, and evaluation of language assessment literacy training for township English teachers* [Master, <https://link.cnki.net/doi/10.27032/d.cnki.ggdwu.2023.000045>]

- Delavarpour, P., & Safarnejad, E. (2024). Identifying Indicators of Teacher Proficiency in Assessment for Learning (AfL) Frameworks. *Assessment and Practice in Educational Sciences*, 2(3), 1-10.
- Deng, L., Wu, Y., Chen, L., & Peng, Z. (2024). 'Pursuing competencies' or 'pursuing scores'? High school teachers' perceptions and practices of competency-based education reform in China. *Teaching and teacher education*, 141, 104510.
- Edu, N. (2025). Enhancing Teachers' Professional Development for more Productive Society. *International Journal of Innovative Development and Policy Studies*, 13(1), 197-203.
- Gan, L., & Lam, R. (2022). A review on language assessment literacy: Trends, foci and contributions. *Language Assessment Quarterly*, 19(5), 503-525.
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179-195.
- Giraldo, F. (2021). Language assessment literacy and teachers' professional development: A review of the literature. *Profile: Issues in Teachers' Professional Development*, 23(2), 265-279.
- Giraldo, F. (2022). *Language assessment literacy and the professional development of pre-service foreign language teachers*. Editorial Universidad de Caldas.
- Giraldo, F. (2025). *Fostering Teachers' Language Assessment Literacy: Theory & Practice*. UNIVERSIDAD DE CALDAS.
- Giraldo, F., & Quintero, D. M. (2019). Language assessment literacy and the professional development of pre-service language teachers. *Colombian Applied Linguistics Journal*, 21(2), 243-259.
- Hidayat, R., Sujadi, I., & Usodo, B. (2023). Description of assessment: Assessment for learning and assessment as learning on teacher learning assessment. *Journal of Education Research and Evaluation*, 7(4), 653-661.
- Hidayati, D. (2024). Integrating Student Self-Assessment as a Key Instrument for Achieving Need-Supportive Learning Outcomes. *English Language and Literature in Education Journal*, 2(2), 24-41.
- Inbar-Lourie, O. (2016). Language assessment literacy. In *Language testing and assessment* (pp. 1-14). Springer.
- Iyanda, J. (2025). Transforming Assessment Practices in Higher. *Quality in African Higher Education: Development Perspectives from Selected Regions*, 20, 301.
- Jäppinen, A.-K., Leclerc, M., & Tubin, D. (2016). Collaborativeness as the core of professional learning communities beyond culture and context: Evidence from Canada, Finland, and Israel. *School effectiveness and school improvement*, 27(3), 315-332.
- Jin, Y. (2018). Development of foreign language teachers' assessment literacy: Theoretical framework and path exploration. *Frontiers of Foreign Language Education Research*, 4, 65-72.
- Kamran, F. (2024). *Relevance of Formative Assessment and Feedback Practices of Language and Science Teachers for Students' Motivation and Self-regulation at Public Higher Education Institutions in Pakistan*. Friedrich-Alexander-Universitaet Erlangen-Nuernberg (Germany).
- Khursheed, S., & Alwi, S. K. K. (2023). Evaluation of formative assessment strategies in enhancing student achievements and learning motivation in higher secondary schools. *Pakistan Journal of Educational Research*, 6(2).
- Kianinezhad, N. (2023). Language assessment literacy: The role of self-reflection. *Journal of English Language and Culture*, 14(1).

- Kiran, A. (2023). English Language Assessment: Innovations, Validity, and Reliability. *Journal of International English Research Studies (JIERS)*, ISSN: 3048-5231, 1(2), 1-8.
- Lave, J., & Wenger, E. (2001). Legitimate peripheral participation in communities of practice. In *Supporting lifelong learning* (pp. 121-136). Routledge.
- Li, J., Shi, Z., & Xue, E. (2020). The problems, needs and strategies of rural teacher development at deep poverty areas in China: Rural schooling stakeholder perspectives. *International Journal of Educational Research*, 99, 101496.
- Liao, S., Chen, X., Jiao, C., Huang, F., Zhou, S., & Ye, J. (2025). Current situation and perspective strategies for the development of rural primary school English education. *Advances in Social Sciences*, 14, 208.
- Nolan, A., & Molla, T. (2018). Teacher professional learning as a social practice: An Australian case. *International studies in sociology of education*, 27(4), 352-374.
- Pastore, S. (2023). Teacher assessment literacy: A systematic review. *Frontiers in Education*, Peng, H., & Tang, Z. (2022). Research hotspots, frontier directions, and future prospects of teacher training in China. *Journal of Teacher Education*, 9(6), 155-164.
- Schellekens, L. H., Bok, H. G., De Jong, L. H., Van der Schaaf, M. F., Kremer, W. D., & Van der Vleuten, C. P. (2021). A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). *Studies in Educational Evaluation*, 71, 101094.
- Tajeddin, Z., Khatib, M., & Mahdavi, M. (2022). Critical language assessment literacy of EFL teachers: Scale construction and validation. *Language Testing*, 39(4), 649-678.
- Tsagari, D., & Armostis, S. (2025). Contextualizing Language Assessment Literacy: A Comparative Study of Teacher Beliefs, Practices, and Training Needs in Norway and Cyprus. *Education Sciences*, 15(7), 927.
- Vogt, K., Bøhn, H., & Tsagari, D. (2024). Language assessment literacy. *Language Teaching*, 57(3), 325-340.
- Wang, D., Wang, J., Li, H., & Li, L. (2017). School context and instructional capacity: A comparative study of professional learning communities in rural and urban schools in China. *International journal of educational development*, 52, 1-9.
- Xu, Y., & Qiu, X. (2022). Online assessment practices and assessment literacy of college English teachers: A qualitative study. . *Frontiers of Foreign Language Education Research*.
- Yan, C. (2015). 'We can't change much unless the exams change': Teachers' dilemmas in the curriculum reform in China. *Improving schools*, 18(1), 5-19.
- Yang, L. P., & Xin, T. (2022). Changing educational assessments in the post-COVID-19 era: from assessment of learning (AoL) to assessment as learning (AaL). *Educational Measurement: Issues and Practice*, 41(1), 54-60.
- Yunfeng, H., Zhojing, W., Jingjing, M., Tiejie, T., Hong, W., Peizhong, L., Ping, L., & Yingchun, X. (2023). Multi-dimensional Reflection on the High-quality Development of Teacher Education in the New Era. *Journal of Jishou University (Social Sciences Edition)*, 44(6), 17.
- Zamiri, M., & Esmaeili, A. (2024). Strategies, methods, and supports for developing skills within learning communities: A systematic review of the literature. *Administrative Sciences*, 14(9), 231.
- Zhao, D., & Chen, Y. (2022). Research on the optimal allocation of teacher resources in small rural schools: Experience and implications from the United States. *Promoting High-Quality Development of Basic Education*, 197.
- Zhu, Y. (2024). Manifestations, causes, and countermeasures of mental health problems among rural left-behind children. *Advances in Psychology*, 14, 570.