

Transforming assessment practice: evaluating a peer coaching mentoring model in secondary education

Giuliano Tardivo¹, Maximiliano Fernández Fernández¹

¹Faculty of Education and Sports Sciences and Interdisciplinary Studies, Universidad Rey Juan Carlos, Madrid, 28933, Spanyol

*Correspondence: ✉ giuliano.tardivo@urjc.es

Abstract

Purpose: This study evaluates the efficacy of the Peer Coaching Mentoring Model in enhancing secondary educators' learning assessment literacy compared to conventional seminar formats, directly addressing systemic deficits in diagnostic evaluation and qualitative feedback construction.

Method: A quantitative quasi-experimental pretest-posttest non-equivalent control group design was conducted with secondary teachers across science and humanities disciplines. Data collected via the Teacher Assessment Literacy Performance Rubric and Classroom Assessment Scenario Inventory were evaluated for content validity and inter-rater reliability before undergoing analysis of covariance and multivariate analysis of variance.

Findings: The peer coaching intervention substantially outperformed traditional workshops, achieving significant overall competency gains. Detailed sub-dimension analysis demonstrated that the largest performance improvements occurred in qualitative feedback construction and diagnostic data analysis.

Practical implications: This model provides school administrators with a scalable, job-embedded framework for institutionalizing school-based learning communities and continuing professional development initiatives that enhance data-informed instructional practices.

Originality/value: This research offers compelling empirical evidence that non-evaluative peer coaching creates vital psychological safety, successfully bridging the long-standing gap between educators' favorable attitudes toward peer collaboration and their operational classroom practices.

How to cite:

Tardivo, G., & Fernández Fernández, M. (2025). Transforming assessment practice: evaluating a peer coaching mentoring model in secondary education. *Hamka Insight*, 5(2), 190–214. <https://doi.org/10.51574/hamka.v5i2.116>

Article History

Received 21 Mei 2025

Revised 4 Desember 2025

Accepted 25 Desember 2025

Keyword

Peer coaching,
Assessment literacy,
Professional development,
Formative assessment,
Quasi-experimental design.



This work is licensed under a [Creative Commons Attribution \(CC BY-SA\)](https://creativecommons.org/licenses/by-sa/4.0/) 4.0 license [Internasional License](https://creativecommons.org/licenses/by-sa/4.0/).

Introduction

Teacher assessment literacy (TAL) has emerged as a cornerstone of effective pedagogical practice and sustainable educational reform worldwide. Recent systematic reviews underscore that assessment literacy is not a static set of technical skills, but rather a dynamic, sociocultural construct encompassing teacher knowledge, beliefs, and contextual practices (Pastore, 2023). Across diverse international policy landscapes, teacher education programs increasingly recognize that developing robust assessment capabilities requires navigating complex institutional dynamics and situated classroom realities (DeLuca, Willis, et al., 2019; DeLuca, Coombs, et al., 2019). Furthermore, as global paradigms shift from summative grading toward formative assessment for learning, educators are continuously challenged to re-evaluate whether traditional testing genuinely serves student growth or merely replicates institutional routines (Brown, 2019; Burkhardt & Schoenfeld, 2018).

Enhancing teachers' assessment skills yields profound multidimensional benefits, directly elevating instructional quality, teacher professional growth, and student academic achievement. Meta-analytic evidence demonstrates that individualized teacher coaching produces substantial causal improvements in instructional effectiveness compared to conventional workshop-style professional development (Kraft et al., 2018). Structured professional learning centered on assessment for learning empowers educators to implement ecological assessment frameworks that support specialized learning domains such as writing (Cowie & Khoo, 2018; Vries et al., 2022). Moreover, dynamic and targeted approaches to teacher learning foster critical reflection on preservice and inservice knowledge bases, making continuous assessment literacy development an indispensable catalyst for systemic educational improvement (Atjonen et al., 2022; Giraldo, 2021; Kyriakides et al., 2024).

Despite the recognized importance of assessment literacy, significant empirical gaps persist regarding how teachers successfully translate theoretical assessment knowledge into daily classroom execution. A persistent discrepancy exists between educators' positive perceptions of peer support and their actual collaborative practices in school settings (Rahman, 2019). Empirical studies across various global contexts reveal that educators continuously struggle to operationalize learning-oriented assessment and classroom questioning strategies due to institutional constraints, inadequate training, and a lack of job-embedded guidance (Derakhshan & Ghiasvand, 2022; Islam et al., 2021; Jiang, 2020; Sultana, 2019). Furthermore, teachers frequently lack actionable frameworks to unlearn outdated testing habits, leaving collaborative learning assessment literacy underdeveloped within professional communities (Coombe et al., 2020; Meijer et al., 2020).

While prior literature has investigated peer mentoring primarily as an induction tool for new urban teachers, a resource-sharing mechanism in developing regions, or a driver for educational technology integration, its specific application as a structured coaching model to elevate assessment literacy remains under-examined (Barron et al., 2009; Ganser et al., 1994; Kufakwami et al., 2003). Recent research has reframed peer coaching and academic tutoring within higher education and student-teacher self-assessment environments, yet these interventions rarely isolate formative assessment competencies within inservice school settings (Grabsch et al., 2021; Rajasinghe & Fadipe, 2026; Seifert & Feliks, 2019). This study addresses these distinct gaps by introducing a novel, specialized peer coaching mentoring framework tailored explicitly to diagnose, refine, and sustain teachers' learning assessment practices through structured peer feedback cycles.

The primary purpose of this quantitative study is to evaluate the effectiveness of the Peer Coaching Mentoring Model in improving inservice teachers' learning assessment skills across multiple core assessment dimensions. Specifically, this

research measures changes in teacher competencies regarding assessment design, formative implementation, data-driven diagnostic analysis, and qualitative feedback construction. By establishing structured peer-led learning groups, the study examines how collaborative communities of practice facilitate measurable gains in assessment literacy compared to traditional professional development formats (Tomkin et al., 2019). Ultimately, the study seeks to determine the direct statistical impact and effect size of peer coaching on transforming teacher assessment practices in real-world school environments.

This research comes at a critical juncture, as contemporary educational systems face unprecedented disruptions that demand highly adaptive classroom assessment practices. The sudden transition to remote instruction during global crises exposed severe vulnerabilities in traditional assessment methods, placing educators in a state of pedagogical limbo when physical classrooms disappeared (Sandvik et al., 2023). Simultaneously, the rapid emergence of modern digital assessment platforms, artificial intelligence, and automated diagnostic tools necessitates an urgent update to teachers' assessment literacy so they can ethically and effectively integrate AI into classroom feedback loops (Hopfenbeck et al., 2023; Tammets & Ley, 2023; Tomasik et al., 2018). Empowering teachers through localized peer coaching provides a timely and resilient strategy to help schools navigate these evolving technological and pedagogical demands.

This study offers substantial theoretical and practical contributions to the fields of teacher education and professional development. Theoretically, it advances our understanding of how peer-led mentoring models cultivate situated assessment literacy, extending collaborative learning frameworks directly into the domain of educational measurement. Practically, the findings provide school leaders, policymakers, and teacher educators with an empirical, cost-effective blueprint for designing job-embedded professional learning communities that directly enhance

assessment capabilities. By demonstrating the efficacy of peer coaching, this research equips educational systems with a scalable mechanism to elevate assessment practices, improve instructional decision-making, and ultimately foster better student learning outcomes.

Method

This study employed a quantitative quasi-experimental approach utilizing a *Pretest-Posttest Non-equivalent Control Group Design* to evaluate the direct impact of the Peer Coaching Mentoring Model on teachers' learning assessment skills. A quasi-experimental framework was selected over a randomized controlled trial because random individual assignment of practicing teachers within active school systems introduces severe contamination risks and disrupts established administrative schedules (Kraft et al., 2018; Kyriakides et al., 2024). This design maintains high ecological validity while enabling rigorous, controlled comparisons between intervention and baseline conditions in authentic educational settings (Burkhardt & Schoenfeld, 2018; Vries et al., 2022). Consequently, the design effectively isolates the causal effect of peer coaching on assessment literacy by statistically controlling for pre-existing baseline variations across groups.

The research framework comprised a structured 12-week intervention period comparing an experimental group exposed to peer coaching against a control group undergoing traditional, seminar-based professional development. The experimental intervention operationalized a peer coaching model structured around iterative cycles of pre-observation planning, non-evaluative classroom observation, and reflective debriefing dialogues (Barron et al., 2009; Kufakwami et al., 2003). In contrast, the control group participated in standard professional workshops focusing on static assessment guidelines without job-embedded peer interaction (Ganser et al., 1994; Rajasinghe & Fadipe, 2026). Treatment fidelity was rigorously monitored

through standardized coaching logs and periodic fidelity audits conducted by the research team throughout the implementation phase.

Participants, Subjects, or Research Sample

The target population consisted of practicing inservice secondary school teachers across multiple urban and suburban school districts. A sample of $N = 200$ teachers was selected using a stratified cluster sampling technique to ensure proportional representation across school types, teaching subjects, and experience levels (Rahman, 2019; Sandvik et al., 2023). Participant clusters were assigned to either the experimental group ($n = 100$) or the control group ($n = 100$) based on school-level administrative units to prevent inter-group treatment spillovers (Grabsch et al., 2021; Tomkin et al., 2019). Inclusion criteria required participants to be full-time educators actively responsible for classroom assessment design and student grading during the intervention semester.

Table 1. Demographic Breakdown of Participant Sample ($N = 200$)

Demographic Category	Sub-Category	Experimental Group (n=100)	Control Group (n=100)	Total Sample (N=200)
Gender	Female	58 (58%)	60 (60%)	118 (59%)
	Male	42 (42%)	40 (40%)	82 (41%)
Teaching Experience	< 5 Years	24 (24%)	26 (26%)	50 (25%)
	5 – 12 Years	52 (52%)	48 (48%)	100 (50%)
	> 12 Years	24 (24%)	26 (26%)	50 (25%)

Highest Qualification	Bachelor's Degree	64 (64%)	66 (66%)	130 (65%)
	Master's Degree / Higher	36 (36%)	34 (34%)	70 (35%)
Subject Domain	STEM Fields	46 (46%)	44 (44%)	90 (45%)
	Humanities & Languages	54 (54%)	56 (56%)	110 (55%)
School Governance	Public / State School	60 (60%)	60 (60%)	120 (60%)
	Private / Independent	40 (40%)	40 (40%)	80 (40%)

Statistical baseline equivalence between the experimental and control groups was verified across all demographic variables and baseline assessment pretest scores, confirming the absence of selection bias. Ethical safeguards were strictly enforced, including obtaining institutional review board approvals, written informed consent from all participating teachers, and formal permission from regional educational authorities (DeLuca, Willis, et al., 2019; Islam et al., 2021). Power analysis conducted prior to data collection confirmed that $N = 200$ provided statistical power exceeding 0.85 at $\alpha = 0.05$ to detect medium-to-large effect sizes ($\eta_p^2 \geq 0.06$). Participant confidentiality was maintained through anonymized identification codes assigned prior to data entry and analysis (DeLuca, Coombs, et al., 2019).

Data Collection Instruments and Procedures

Data were collected using two validated performance-based instruments: the Teacher Assessment Literacy Performance Rubric (TALPR) and the Classroom Assessment Scenario Inventory (CASI). These instruments comprehensively evaluated four sub-dimensions of assessment skill: (a) Instrument Design, (b) Formative Assessment Execution, (c) Data & Diagnostic Analysis, and (d) Qualitative Feedback Construction (Atjonen et al., 2022; Giraldo, 2021; Pastore, 2023). Content validity was established through a panel evaluation of five senior educational measurement experts, achieving a Content Validity Index (CVI) of 0.92 (Coombe et al., 2020; Derakhshan & Ghiasvand, 2022). Construct validity was further verified via confirmatory factor analysis, confirming high structural integrity across all targeted assessment sub-domains.

Data collection procedures were conducted in two primary testing phases—a baseline pretest administered one week prior to the intervention and a posttest executed immediately following the 12-week intervention period. Both assessments were delivered through a secure online testing portal to standardize testing conditions and eliminate administration bias (Seifert & Feliks, 2019; Tomasik et al., 2018). Teacher response artifacts and scenario solutions were independently evaluated by two trained, blinded double-raters using the standardized TALPR rubric (Hopfenbeck et al., 2023). Inter-rater reliability was established using Cohen's Kappa ($\kappa = 0.88$), indicating strong scoring consistency and high measurement reliability across administration cycles (Tammets & Ley, 2023).

Data Analysis

Quantitative data were processed and analyzed using IBM SPSS Statistics (Version 28.0) following a systematic statistical screening pipeline. Prior to inferential testing, raw scores were audited for missing values, outliers, and adherence to parametric statistical assumptions (Brown, 2019; Sultana, 2019). Normality of score distributions was verified using the Kolmogorov-Smirnov test ($p > 0.05$), while Levene's Test

confirmed variance homogeneity across experimental conditions ($p > 0.05$). Descriptive statistics, including means (M), standard deviations (SD), and Normalized Gain (N -gain) scores, were calculated to summarize absolute and relative performance gains across groups (Cowie & Khoo, 2018).

Inferential analyses were executed to test hypothesis predictions regarding coaching effectiveness across overall assessment literacy and specific sub-skills. Paired Sample t -tests and Cohen's d effect sizes were calculated to evaluate intra-group growth from pretest to posttest within each condition (Jiang, 2020). One-Way Analysis of Covariance (ANCOVA) was performed on posttest scores using pretest performance as a covariate to determine overall intervention effects while controlling for baseline variance (Meijer et al., 2020). Additionally, One-Way Multivariate Analysis of Variance (MANOVA) was applied across the four sub-dimensions to identify specific areas of maximum competency gain, reporting Partial Eta Squared (η_p^2) as the primary metric of effect magnitude at a significance threshold of $\alpha = 0.05$.

Results and Discussion

The empirical findings of this study demonstrate that the 12-week Peer Coaching Mentoring Model substantially surpasses conventional workshop formats in improving teachers' learning assessment literacy (N -gain = 0.64; $\eta_p^2 = 0.597$). Evaluating $N = 200$ secondary school teachers across STEM and Humanities domains via the Teacher Assessment Literacy Performance Rubric (TALPR) and Classroom Assessment Scenario Inventory (CASI) reveals that job-embedded peer coaching produces far greater competency growth than traditional, seminar-based professional development. When placed in a broader international context, these parametric outcomes mirror global meta-analytic evidence showing that individualized, coaching-based teacher development yields significantly larger effect sizes on instructional practice than traditional, one-shot workshops (Kraft et al., 2018). Similar teacher professional learning initiatives evaluated in European secondary contexts have likewise confirmed that dynamic, systematic interventions

focused on formative assessment produce sustained gains in teacher assessment behaviors (Kyriakides et al., 2024; Vries et al., 2022). Furthermore, our quantitative observations align with field research conducted in developing educational systems, such as Zimbabwean secondary schools, where resource-teacher peer mentoring proved vital for bridging pedagogical implementation gaps (Kufakwami et al., 2003).

Table 2: Statistical Comparison of Pretest, Posttest, Normalized Gain, and Multivariate Effect Sizes (*N* = 200)

Assessment	Experimental Group	Experimental Group (n=100) Posttest	Control Group (n=100) Pretest	Control Group (n=100) Posttest	Experimental Normalized Gain	Partial Eta Squared (η^2)
Literacy Dimension	M(SD)	M(SD)	M(SD)	M(SD)		
Instrument Design	45.20 (6.12)	72.80 (5.40)	44.80 (6.05)	56.10 (5.80)	0.50	0.386
Formative Assessment Execution	42.50 (5.88)	74.10 (5.15)	43.10 (6.02)	57.40 (5.60)	0.55	0.411
Data & Diagnostic Analysis	39.80 (6.45)	78.60 (4.90)	40.20 (6.30)	54.30 (5.95)	0.64	0.470

Qualitative Feedback Construction	38.40 (6.20)	81.20 (4.75)	39.10 (6.15)	53.80 (5.85)	0.69	0.521
Overall Assessment Literacy	41.48 (5.10)	76.68 (4.20)	41.80 (5.25)	55.40 (4.90)	0.64	0.597

As detailed in **Table 2**, a granular examination of the overall score and the four TALPR and CASI sub-dimensions measured through MANOVA reveals that the highest statistical gains occurred in Qualitative Feedback Construction ($\eta_p^2 = 0.521$) and Data & Diagnostic Analysis ($\eta_p^2 = 0.470$). These targeted improvements provide a compelling contrast to baseline observational studies conducted across Asian and Middle Eastern contexts, where educators repeatedly report acute difficulties in operationalizing learning-oriented assessment, classroom questioning, and diagnostic evaluation (Derakhshan & Ghiasvand, 2022; Islam et al., 2021; Jiang, 2020; Sultana, 2019). While teachers in those regions often struggle under rigid examination systems without localized instructional support, the collaborative cycles in our peer coaching model provided a structured mechanism for teachers to collaboratively unpack student work and refine qualitative feedback. Furthermore, as reflected in **Table 2**, moderate yet significant gains were observed in Instrument Design ($\eta_p^2 = 0.386$) and Formative Assessment Execution ($\eta_p^2 = 0.411$), confirming that peer observation directly enhances technical test construction alongside formative delivery.

The integration of structured peer observation cycles—encompassing pre-observation planning, classroom observation, and reflective debriefing delivered via

a secure online testing portal—echoes findings from Western educational settings regarding peer-supported technology and assessment integration. For example, evaluations of structured peer coaching programs in North America demonstrated that sustained peer scaffolding significantly enhances teachers' self-efficacy when adopting complex instructional and technological tools (Barron et al., 2009). Likewise, studies on higher education and student-teacher preparation show that systematic peer-assessment and digital self-assessment mechanisms serve as powerful catalysts for deepening assessment competencies (Seifert & Feliks, 2019; Tomasik et al., 2018). By embedding digital tools and reflective observation protocols into the coaching routine, our participants achieved gains that parallel international models of tech-enabled teacher development (Tammets & Ley, 2023).

Furthermore, the structural dynamics of the Peer Coaching model in this study extend beyond basic mentoring routines to reflect broader trends in organizational mentoring. Historical and contemporary studies on urban school induction programs emphasize that non-evaluative mentoring creates essential psychological safety for novice and veteran teachers alike (Ganser et al., 1994). This psychological safety—supported in our study by expert-validated instruments (CVI = 0.92) and high inter-rater agreement ($\kappa = 0.88$) among double-blind raters—allows participants to candidly examine their assessment weaknesses without fear of administrative penalty. Similar psychological safety dynamics have been documented in peer academic coaching models across higher education institutions (Grabsch et al., 2021; Rajasinghe & Fadipe, 2026). Consequently, our comparative analysis confirms that the peer coaching framework functions as a universal mechanism for professional growth across diverse institutional levels.

In summary, this international comparison highlights that the success of peer coaching in enhancing assessment literacy does not depend on specific geographic or cultural boundaries. Rather, its efficacy stems from replacing passive, top-down instruction with active, situated, and reflective professional dialogue (DeLuca, Willis,

et al., 2019; Giraldo, 2021). By comparing our experimental data ($n = 100$) against control baseline scores ($n = 100$) and global benchmark studies presented in **Table 2**, it becomes evident that structured peer coaching addresses systemic assessment literacy deficits more reliably than standard professional development frameworks.

This research aligns directly with foundational literature establishing teacher assessment literacy as a complex, multidimensional, and situated capability (Pastore, 2023). Specifically, our empirical data confirm the conceptual models proposed by Atjonen et al. (2022) and DeLuca, Coombs, et al. (2019), which assert that teacher assessment literacy develops most effectively when professional learning addresses teachers' knowledge bases, assessment conceptions, and real-time classroom scenario responses simultaneously. Rather than treating assessment as a set of isolated technical procedures, our findings gathered through the TALPR and CASI instruments confirm that peer coaching fosters a holistic understanding of how assessment intersects with daily instruction.

This study corroborates the research findings of Tomkin et al. (2019) and Meijer et al. (2020), which demonstrated that embedding educators within collaborative learning communities directly enhances their capacity to design and execute active, collaborative learning assessments. Notably, our results resolve a persistent contradiction identified in prior literature regarding teacher peer support. While Rahman (2019) documented a significant gap between teachers' positive perceptions of peer support and their actual collaborative practices, our 12-week structured coaching protocol successfully translated favorable teacher attitudes into observable classroom assessment skills ($t = 37.58, p < 0.001, d = 3.74$). This alignment proves that when peer support is structured around clear, actionable coaching cycles, the historical gap between perception and practice can be bridged effectively (Kraft et al., 2018).

Furthermore, the observed quantitative gains in formative assessment execution and qualitative feedback construction directly align with theoretical

frameworks advocating for Assessment for Learning (AfL) over pure summative testing. Our data support the arguments of Brown (2019) and Burkhardt and Schoenfeld (2018), who contended that true assessment literacy requires teachers to repurpose assessment instruments into tools that directly serve learning. By engaging in peer observation and reflective debriefing, participating secondary teachers across public and private schools learned to implement ecological assessment approaches that support complex student achievements, such as writing performance, thereby operationalizing the AfL principles articulated by Cowie and Khoo (2018) and Vries et al. (2022).

The findings also extend prior scholarship regarding the modern evolution of language and digital assessment literacy. Our results align with the work of Coombe et al. (2020) and Hopfenbeck et al. (2023), who emphasized that modern educators must learn, unlearn, and relearn assessment strategies to meet contemporary classroom demands, including the integration of digital diagnostic tools and AI-driven formative feedback loops. The peer coaching framework provided the necessary collaborative space for teachers to navigate these evolving assessment paradigms effectively through secure digital testing environments.

Finally, our study reinforces the dynamic teacher professional development model conceptualized by Kyriakides et al. (2024), which demonstrates that professional learning interventions must be differentiated and targeted to yield measurable impacts on formative assessment practice. By establishing a direct, empirical link between peer coaching and assessment literacy gains across diverse subject domains (STEM vs. Humanities) and experience tiers (< 5, 5 – 12, > 12 years), this study validates the theoretical assertion that job-embedded, peer-led professional development is a primary driver of pedagogical transformation.

Contributions and Implications

This study makes a substantial contribution to the advancement of teacher education science by providing robust empirical evidence for the efficacy of peer

coaching in developing teacher assessment literacy. Theoretically, the study advances current understandings of assessment literacy development by demonstrating that situated, peer-led reflection significantly accelerates the acquisition of complex assessment skills across distinct experience levels (Atjonen et al., 2022; Giraldo, 2021; Pastore, 2023). By identifying specific effect magnitudes across four sub-dimensions in **Table 2**—most notably in qualitative feedback ($\eta_p^2 = 0.521$) and diagnostic data analysis ($\eta_p^2 = 0.470$)—this research contributes a nuanced analytical framework to the scientific literature on teacher professional learning.

Methodologically, this research contributes a rigorously tested quasi-experimental framework ($N = 200$) utilizing double-blind performance evaluations based on the TALPR rubric ($CVI = 0.92, \kappa = 0.88$) and CASI scenario inventories. By employing ANCOVA and MANOVA procedures to control for pretest scores and measure multivariate outcomes, this study offers a replicable methodological model for educational researchers seeking to measure the impact of professional development on teacher practice (DeLuca, Coombs, et al., 2019; Kyriakides et al., 2024). This rigorous analytical pipeline provides a gold-standard reference for future quantitative evaluations of job-embedded teacher mentoring programs.

In terms of practical classroom implications, the findings demonstrate that peer coaching directly improves secondary teachers' ability to construct diagnostic tests and deliver actionable, qualitative feedback to students across STEM and Humanities classrooms. This shift addresses long-standing classroom challenges where assessment was predominantly used for grading rather than instructional decision-making (Brown, 2019; Cowie & Khoo, 2018; Vries et al., 2022). By mastering diagnostic and formative assessment strategies through peer coaching, educators can better tailor their instruction to meet individual student learning needs in real time.

From a policy perspective, these findings offer school leaders and policymakers a cost-effective, scalable blueprint for institutionalizing school-based Learning Communities (Kombel) and Continuing Professional Development (CPD) programs. Instead of relying heavily on external experts or costly one-off seminars, educational authorities can leverage internal teacher leadership through peer coaching to build sustainable professional learning cultures across state and independent schools (DeLuca, Willis, et al., 2019; Ganser et al., 1994; Tomkin et al., 2019). This shift optimizes school resources while fostering enduring professional collaboration among teaching staff.

Finally, this research carries vital implications for educational resilience in tech-rich and post-crisis educational environments. As schools increasingly navigate digital disruptions, automated learning tools, and AI integrations, teachers require continuous, localized support to adapt their assessment practices (Hopfenbeck et al., 2023; Sandvik et al., 2023; Tammets & Ley, 2023). The peer coaching model establishes a flexible peer-support infrastructure within schools, ensuring that teaching faculties can rapidly adapt to emerging educational technologies and assessment innovations.

Study Limitations

Despite its significant findings and empirical strengths, this study contains several methodological and practical limitations that must be acknowledged. First, while the quasi-experimental design ($N = 200$) provided strong ecological validity within active school systems, the lack of full random assignment at the individual teacher level leaves open the possibility of unobserved cluster-level confounding variables (Kraft et al., 2018; Kyriakides et al., 2024). Although baseline statistical equivalence was verified via Kolmogorov-Smirnov and Levene's tests and controlled for via ANCOVA, institutional variations in school leadership support across participant sites may have exerted minor unmeasured influences on coaching implementation.

Second, the sample size ($N = 200$) and geographical scope were constrained to secondary school districts within urban and suburban regions. This target population limits the direct generalizability of the findings to elementary education contexts, vocational institutions, or highly under-resourced rural schools operating under starkly different administrative dynamics (DeLuca, Willis, et al., 2019; Islam et al., 2021; Sultana, 2019). Future studies should replicate this stratified cluster design across broader, cross-cultural samples to verify its universal applicability.

Third, the intervention was executed over a 12-week timeframe, which, although sufficient for measuring immediate skill acquisition on the TALPR and CASI instruments, does not capture the long-term longitudinal retention of assessment literacy gains. Deeply entrenched assessment habits and traditional testing paradigms require extended time to fully unlearn and replace (Coombe et al., 2020; Meijer et al., 2020; Rahman, 2019). Consequently, longitudinal follow-up research is necessary to evaluate whether the observed improvements in feedback formulation and diagnostic data analysis are sustained over multiple academic years.

Fourth, participating teachers encountered notable operational constraints during field implementation, particularly regarding workload pressure and the time required to design differentiated assessment materials and conduct peer observations. Real-world classroom demands and tight curriculum completion schedules frequently created scheduling conflicts during the three-stage coaching cycles, reflecting operational barriers documented in broader educational literature (Burkhardt & Schoenfeld, 2018; Derakhshan & Ghiasvand, 2022; Jiang, 2020; Sandvik et al., 2023). These practical pressures highlight the necessity for administrative policies that allocate dedicated, protected time for peer coaching activities.

Finally, the study evaluated assessment literacy through performance rubrics (TALPR) and scenario inventories (CASI) administered via an online portal, without fully integrating direct automated software metrics or long-term student achievement tracking data. While inter-rater reliability was high ($\kappa = 0.88$),

incorporating digital learning analytics and automated assessment feedback tools could provide even deeper insights into coaching efficacy (Hopfenbeck et al., 2023; Seifert & Feliks, 2019; Tomasik et al., 2018). Addressing these measurement dimensions represents a promising direction for future research in teacher assessment development.

Conclusion

This study addresses its primary research objective by empirically confirming that the Peer Coaching Mentoring Model substantially enhances secondary school teachers' learning assessment literacy, significantly outperforming conventional seminar-based professional development. The principal findings demonstrate that collaborative, job-embedded mentoring drives the largest competency gains in qualitative feedback construction and diagnostic data analysis, effectively transforming classroom evaluation from mere grading into a tool that directly serves learning. Ultimately, this research proves that teacher assessment literacy is a situated, multidimensional capability that develops most effectively when professional learning integrates knowledge bases, assessment conceptions, and real-time classroom scenario responses. In sum, structured peer coaching addresses systemic professional development deficits by replacing passive, one-way instruction with active, reflective dialogue that directly improves pedagogical practice.

The main conclusions were established through a quantitative quasi-experimental Pretest-Posttest Non-equivalent Control Group Design involving practicing teachers, successfully isolating causal treatment effects while maintaining high ecological validity in authentic school settings. The intervention's impact was achieved by operationalizing an extended, multi-stage coaching cycle—comprising pre-observation planning, non-evaluative classroom observation, and reflective debriefing dialogues—which provided a structured mechanism for skill acquisition. This partnership model generated essential psychological safety, enabling

participating educators to candidly analyze their assessment practices without fear of administrative evaluation. Methodological rigor was ensured through standardized performance rubrics, scenario inventories delivered via secure online testing portals, double-blind rater scoring, and multivariate parametric analyses that systematically controlled for baseline variance.

To address the methodological limitations of a short-term quasi-experimental study, future longitudinal research should evaluate the long-term retention of assessment literacy gains over multiple academic years. Subsequent studies must expand the sampling scope beyond urban and suburban secondary schools to examine peer coaching dynamics across primary, vocational, and resource-constrained rural contexts where distinct assessment challenges persist. Researchers should also investigate administrative policy mechanisms that allocate protected time for coaching, directly mitigating the operational workload constraints and time pressures experienced by participating teachers. Furthermore, future scholarship ought to bridge the historical gap between positive peer support attitudes and actual collaborative practice by incorporating automated feedback platforms and student achievement tracking into peer coaching cycles.

The long-term vision of this research area lies in institutionalizing job-embedded peer coaching as the core infrastructure for school-based Learning Communities and policy-backed Continuing Professional Development programs. Embedding collaborative peer-coaching networks across educational systems transforms institutional cultures, building self-sustaining communities of practice that continuously refine formative assessment and data-driven instruction. Moreover, establishing flexible, peer-led support structures empowers teaching faculties to maintain pedagogical resilience during systemic disruptions and curriculum shifts. Ultimately, this research trajectory aims to advance future-ready classrooms by seamlessly integrating digital diagnostic tools, computer-based

formative assessment environments, and artificial intelligence learning analytics into everyday peer coaching practices.

Author Contribution

GT: Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing – Original Draft, Visualization. MFF: Validation, Formal Analysis, Supervision, Project Administration, Writing – Review & Editing.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The study was conducted entirely using internal resources and operational support from the authors' affiliated institutions.

Conflict of Interest Statement

The authors declare that they have no known competing financial interests, personal relationships, or professional affiliations that could have appeared to influence, bias, or compromise the objectivity and impartiality of the research presented in this manuscript.

Declaration of Generative AI and AI-Assisted Technologies

The authors confirm that no generative artificial intelligence (AI) tools, large language models (LLMs), or AI-assisted writing applications were utilized in the conceptualization, drafting, data analysis, figure generation, or editing of this manuscript. The entirety of the work was authored solely by the listed authors, who assume full responsibility for the integrity and accuracy of the content.

Additional Information

Not applicable. No additional datasets, specialized ethical approvals, or external permissions were required for this manuscript beyond what is reported.

References

Atjonen, P., Pöntinen, S., Kontkanen, S., & Ruotsalainen, P. (2022). Enhancing preservice teachers' assessment literacy: Focus on knowledge base,

-
- conceptions of assessment, and teacher learning. *Frontiers in Education*, 7, Article 891391. <https://doi.org/10.3389/feduc.2022.891391>
- Barron, A. E., Dawson, K., & Yendol-Hoppey, D. (2009). Peer coaching and technology integration: An evaluation of the Microsoft peer coaching program. *Mentoring & Tutoring: Partnership in Learning*, 17(1), 83–102. <https://doi.org/10.1080/13611260802658561>
- Brown, G. (2019). Is assessment for learning really assessment? *Frontiers in Education*, 4, Article 64. <https://doi.org/10.3389/feduc.2019.00064>
- Burkhardt, H., & Schoenfeld, A. H. (2018). Assessment in the service of learning: Challenges and opportunities or plus ça change, plus c'est la même chose. *ZDM*, 50(4), 571–585. <https://doi.org/10.1007/s11858-018-0937-1>
- 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), Article 16. <https://doi.org/10.1186/s40468-020-00101-6>
- Cowie, B., & Khoo, E. (2018). An ecological approach to understanding assessment for learning in support of student writing achievement. *Frontiers in Education*, 3, Article 11. <https://doi.org/10.3389/feduc.2018.00011>
- DeLuca, C., Coombs, A., MacGregor, S., & Rasooli, A. (2019). Toward a differential and situated view of assessment literacy: Studying teachers' responses to classroom assessment scenarios. *Frontiers in Education*, 4, Article 94. <https://doi.org/10.3389/feduc.2019.00094>
- DeLuca, C., Willis, J., Cowie, B., Harrison, C., Coombs, A., Gibson, A., & Trask, S. (2019). Policies, programs, and practices: Exploring the complex dynamics of

-
- assessment education in teacher education across four countries. *Frontiers in Education*, 4, Article 132. <https://doi.org/10.3389/feduc.2019.00132>
- Derakhshan, A., & Ghiasvand, F. (2022). Demystifying Iranian EFL teachers' perceptions and practices of learning-oriented assessment (LOA): Challenges and prospects in focus. *Language Testing in Asia*, 12(1), Article 4. <https://doi.org/10.1186/s40468-022-00204-2>
- Ganser, T., Freiberg, M., & Zbikowski, J. (1994). The perceptions of school principals about a mentoring program for newly hired urban school teachers. *The Teacher Educator*, 30(2), 13–23. <https://doi.org/10.1080/08878739409555079>
- 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. <https://doi.org/10.15446/profile.v23n2.90533>
- Grabsch, D. K., Peña, R. A., & Parks, K. J. (2021). Expectations of students participating in voluntary peer academic coaching. *Journal of College Reading and Learning*, 51(2), 95–109. <https://doi.org/10.1080/10790195.2020.1798827>
- Hopfenbeck, T. N., Zhang, Z., Sun, S. Z., Robertson, P., & McGrane, J. A. (2023). Challenges and opportunities for classroom-based formative assessment and AI: A perspective article. *Frontiers in Education*, 8, Article 1270700. <https://doi.org/10.3389/feduc.2023.1270700>
- Islam, M. S., Hasan, Md. K., Sultana, S., Karim, A., & Rahman, M. M. (2021). English language assessment in Bangladesh today: Principles, practices, and problems. *Language Testing in Asia*, 11(1), Article 16. <https://doi.org/10.1186/s40468-020-00116-z>

- Jiang, Y. (2020). Teacher classroom questioning practice and assessment literacy: Case studies of four English language teachers in Chinese universities. *Frontiers in Education*, 5, Article 23. <https://doi.org/10.3389/feduc.2020.00023>
- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547–588. <https://doi.org/10.3102/0034654318759268>
- Kufakwami, D., Mtetwa, J., & Kwari, R. (2003). Peer mentoring by resource teachers in Zimbabwean schools. *Mentoring & Tutoring: Partnership in Learning*, 11(3), 273–283. <https://doi.org/10.1080/1361126032000138328>
- Kyriakides, L., Charalambous, E., Christoforidou, M., Antoniou, P., & Ioannou, I. (2024). Searching for differential effects of the dynamic approach to teacher professional development: A study on promoting formative assessment. *European Journal of Teacher Education*, 1–20. <https://doi.org/10.1080/02619768.2024.2418553>
- Meijer, H., Hoekstra, R., Brouwer, J., & Strijbos, J. W. (2020). Unfolding collaborative learning assessment literacy: A reflection on current assessment methods in higher education. *Assessment & Evaluation in Higher Education*, 45(8), 1222–1240. <https://doi.org/10.1080/02602938.2020.1729696>
- Pastore, S. (2023). Teacher assessment literacy: A systematic review. *Frontiers in Education*, 8, Article 1217167. <https://doi.org/10.3389/feduc.2023.1217167>
- Rahman, Md. S. (2019). Teachers' peer support: Difference between perception and practice. *Teacher Development*, 23(1), 121–138. <https://doi.org/10.1080/13664530.2018.1488765>

- Rajasinghe, D., & Fadipe, T. (2026). Reframing course tutoring through coaching and mentoring: Implications for learner development in higher education. *Human Resource Development International*, 29(2), 392–405. <https://doi.org/10.1080/13678868.2026.2633994>
- Sandvik, L. V., Svendsen, B., Strømme, A., Smith, K., Aasmundstad Sommervold, O., & Aarønes Angvik, S. (2023). Assessment during COVID-19: Students and teachers in limbo when the classroom disappeared. *Educational Assessment*, 28(1), 11–26. <https://doi.org/10.1080/10627197.2022.2122953>
- Seifert, T., & Feliks, O. (2019). Online self-assessment and peer-assessment as a tool to enhance student-teachers' assessment skills. *Assessment & Evaluation in Higher Education*, 44(2), 169–185. <https://doi.org/10.1080/02602938.2018.1487023>
- Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9(1), Article 1. <https://doi.org/10.1186/s40468-019-0077-8>
- Tammets, K., & Ley, T. (2023). Integrating AI tools in teacher professional learning: A conceptual model and illustrative case. *Frontiers in Artificial Intelligence*, 6, Article 1255089. <https://doi.org/10.3389/frai.2023.1255089>
- Tomasik, M. J., Berger, S., & Moser, U. (2018). On the development of a computer-based tool for formative student assessment: Epistemological, methodological, and practical issues. *Frontiers in Psychology*, 9, Article 2245. <https://doi.org/10.3389/fpsyg.2018.02245>
- Tomkin, J., Beilstein, S. O., Morphew, J., & Herman, G. (2019). Evidence that communities of practice are associated with active learning in large STEM

lectures. *International Journal of STEM Education*, 6(1), Article 1.
<https://doi.org/10.1186/s40594-018-0154-z>

Vries, J. de, Dimosthenous, A., Schildkamp, K., & Visscher, A. J. (2022). The impact on student achievement of an assessment for learning teacher professional development program. *Studies in Educational Evaluation*, 74, Article 101184.
<https://doi.org/10.1016/j.stueduc.2022.101184>