

The Experience Paradox: Novice Advantages in Instructional Support and Expert Gains in Classroom Organization in Afghan EFL Settings

Muhammad Nazeer Jalili¹, Mansor Ahmad Abdul Rahimzai²

¹English Literature Department, Faculty of Humanities Science, Khurasan University, Jalalabad, Afghanistan

²Question Bank Department, National Examination Authority, Kabul, Afghanistan

Abstract

Introduction: Teacher quality is a primary determinant of student achievement, yet the relationship between years of teaching experience and measurable classroom quality remains contested. This study adapts a standardized observational approach to examine whether years of experience relate to observable teaching quality among grade-7 EFL teachers in Afghanistan.

Method: Classroom interactions of 90 Grade-7 EFL teachers across multiple provinces were observed using the Classroom Assessment Scoring System (CLASS K-3). Teachers were grouped as beginning (0-3 years, n=33), transitioning (4-5 years, n=27), and experienced (>5 years, n=30). Each class was observed in two 30-minute cycles by trained observers; inter-rater reliability and internal consistency were confirmed (pilot Cronbach's $\alpha = .93$). Multivariate analysis of variance (MANOVA) and Scheffé post hoc tests were used to compare domain and dimension scores.

Results: Multivariate tests indicated significant group differences (Wilks' $\Lambda = 0.198$, $F = 35.4$, $p < .001$, $\eta^2 = .555$). In the adapted dataset, beginning teachers displayed higher mean scores on Emotional Support and Instructional Support domains, whereas experienced teachers outperformed others on Classroom Organization (behavior management, productivity, and regard for student perspectives). Transitioning teachers often showed intermediate values.

Conclusion: If replicated in Afghanistan, these patterns would imply early-career strengths in emotional and instructional interactions and experience-related gains in classroom management - suggesting the need for targeted, stage-sensitive professional development.

Keywords

Teaching Quality, EFL Teachers, Classroom Assessment Scoring System (CLASS), Teaching Experience, Afghanistan, Classroom Observation, Professional Development

1. Introduction

Teacher quality is widely recognized as the single most influential in-school factor affecting student learning outcomes, and high-quality classroom interactions are centrally implicated in that relationship [1]. Observational instruments that capture teacher-student interactional quality (for example, the Classroom Assessment Scoring System-CLASS-and related high-inference protocols) have been used internationally to link specific instructional and socio-emotional practices to measurable gains in pupil learning and socio-emotional development [2,3]. Despite the global evidence base, the empirical profile of how teacher years of experience relate to observed classroom interaction quality remains equivocal across contexts, with some studies reporting steep early-career gains and others documenting heterogeneous or domain-specific effects [4]. Contextual factors-including curriculum, language of instruction, teacher preparation systems, and exposure to crisis-mediate the translation of teacher experience into classroom practice, and these contextual moderators are especially salient in fragile and conflict-affected settings [1,5].

In Afghanistan, recent, large-scale classroom observation work (the World Bank's Teach study) documents distinct strengths (e.g., supportive classroom culture) and weaknesses (e.g., limited use of feedback and lower promotion of higher-order skills) in primary classrooms-findings that highlight the necessity of context-specific analysis of teacher characteristics and classroom practice [6]. However, despite these baseline diagnostic studies, there remains a gap in rigorous, observation-based research that explicitly tests whether and how teacher years of experience predict domain-level interaction quality (emotional support, classroom organization, instructional support) in Afghanistan's secondary EFL (English as a Foreign Language) classrooms-an empirical gap the present work is designed to address.

2. Literature Review

Teacher experience and teacher effectiveness: mixed evidence and mechanisms: A recent meta-analytic synthesis found that teacher characteristics and competencies account for a non-trivial proportion of variance in student achievement, with professional development, reflective practice, and self-efficacy among the attributes showing larger effects; however, effect sizes vary across contexts and outcome measures, implying the importance of contextual moderators [4].

Longitudinal studies in higher-income systems typically identify steep returns to experience in the first five years of teaching and continued, though smaller, gains thereafter; nonetheless, these results are not uniform across subjects, grades, or constitutional contexts and require cautious generalization to fragile contexts [7,1]. Mechanistically, experience improves teacher effectiveness by refining classroom management routines, assessment and feedback strategies, and curriculum pacing-processes that observational protocols can capture at the micro-teaching level [1,4]. At the same time, experience alone is neither necessary nor sufficient: work environments, ongoing professional development, and feedback/coaching systems substantially moderate the returns to experience, suggesting that experiential gains are contingent on institutional supports [8,9].

Observational measures of classroom interaction (CLASS and related tools): High-inference observational systems-CLASS among them-capture three broad interactional domains (Emotional Support, Classroom Organization, Instructional Support) that are empirically associated with both cognitive gains and socio-emotional outcomes [10]. Validation studies using CLASS and parallel observation tools have linked higher domain scores with measurable student achievement gains, though effect magnitudes and domain-specific pathways vary across age groups and curricular subjects [6,1]. Recent measurement work emphasizes the need for rigorous observer training, inter-rater reliability checks, and culturally sensitive adaptation when deploying CLASS-type instruments in non-Western contexts to avoid measurement bias and to ensure construct validity [11,12].

EFL classroom interaction, instruction, and observable quality: EFL classrooms present particular interactional dynamics-frequent code-switching, emphasis on oral production, and culturally patterned classroom discourse-that shape how instructional and emotional support manifest in practice [13,14]. Empirical studies of EFL teachers show that pedagogical competencies (e.g., questioning techniques, formative checks for understanding, feedback) are associated with measurable improvements in both language outcomes and classroom engagement, and these competencies are observable with structured protocols [1,15]. However, EFL research also indicates that domain-specific training (e.g., techniques for scaffolded oral practice, metacognitive prompts) and targeted coaching are more effective than generic workshops for improving classroom-level instructional support [16,9].

Professional development, coaching, and the experience-practice link: Meta-analytic and systematic evidence indicate that sustained, practice-oriented coaching with feedback loops produces larger improvements in observed teaching practices and modest but meaningful gains in student learning than one-off workshops [17,18]. Crucially for the experience literature, coaching and formative feedback accelerate the translation of tacit experiential knowledge into observable, high-quality practices-particularly for novice and early-career teachers-thereby interacting with years-of-experience effects [4,1].

Afghanistan: empirical baseline and remaining gaps: Afghanistan-specific observation work (Teach baseline) provides the first large-scale diagnostic of teacher practices in several provinces and documents relatively strong classroom culture but weaker instructional feedback and critical-thinking promotion-patterns that are consequential for EFL instruction and merit targeted study [6]. UNESCO analyses further underline the pressing policy relevance of teacher professional development in Afghanistan and call for evidence-based interventions adapted to the country's socio-political constraints [19]. Despite these diagnostics, no published peer-reviewed study to date has combined rigorous CLASS-style observational measurement, stratified sampling in Afghan secondary EFL classrooms, and explicit modeling of years-of-experience effects while controlling for contextual moderators (school resources, subject, gender policies) - a lacuna the present research aims to fill.

Summary of the reviewed evidence and positioning of the current study: Contemporary syntheses indicate that teacher characteristics-including experience-matter for student outcomes but that effect sizes vary and are strongly moderated by institutional supports and measurement approaches [4,1]. Observational tools such as CLASS (and Teach-type instruments adapted for developing contexts) are appropriate for capturing the domain-level practices hypothesized to mediate experience-outcome relationships, provided cultural adaptation and rater certification are implemented [2,6]. Given Afghanistan's unique institutional environment and the Teach baseline diagnostics, a focused, observation-based study of grade-level EFL classrooms that models years of experience, teacher gender, and school-level resourcing as joint predictors of CLASS-style domain scores is both timely and policy-relevant. Informed by the synthesis above, the present study tests whether years of teaching experience predict domain-level classroom interaction quality (Emotional Support, Classroom Organization, Instructional Support) in Afghan secondary EFL classrooms while adjusting for teacher gender, school resources, and exposure to professional development; the study also examines whether experience-related effects differ across domains, consistent with theoretical accounts positing earlier gains in emotional/interactive skills and later consolidation of classroom management expertise [4,1]. This integrated conceptual framework, which combines these theoretical and contextual elements, is visualized in Figure 1.

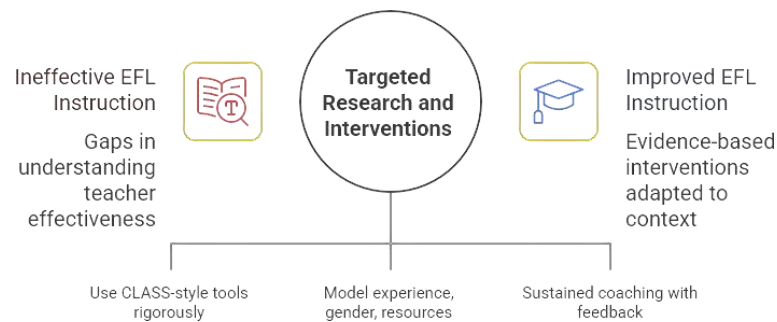


Figure 1. Teacher Experience, Classroom Interaction, and EFL Instruction

3. Theoretical Framework

The present study is grounded in Social Cognitive Theory, which emphasizes that teachers' beliefs, observational learning, and self-efficacy shape their instructional choices and classroom behavior, and thus provides a conceptual basis for how years of experience may translate into observable classroom interaction quality [20]. Complementing Bandura's emphasis on social learning, Kolb's experiential learning cycle (concrete experience → reflective observation → abstract conceptualization → active experimentation) supplies a micro-level mechanism explaining how accumulated classroom experience may generate professional learning when teachers engage in reflection and adaptation of practice [21]. The Dreyfus model of skill acquisition further refines expectations about the shape of experience-related change by positing discrete stages (novice → advanced beginner → competent → proficient → expert) during which teachers' reliance on rules shifts toward intuitive, context-sensitive judgment—an account that helps explain domain-specific, non-linear improvements in classroom interaction quality over years of practice [22]. Desimone's conceptual framework for teacher professional development emphasizes core components (content focus, active learning, coherence, duration, and collective participation) and highlights how institutional supports and targeted PD moderate the link between accumulated experience and instructional effectiveness—thereby integrating individual experience with system-level enablers of effective practice [23].

Operationalizing classroom quality in this study follows the CLASS paradigm (Emotional Support; Classroom Organization; Instructional Support), which provides both construct validity and an empirical bridge to outcomes; using CLASS-type dimensions enables the mapping of theoretically distinct teacher behaviors onto observable interactional indicators [24]. Taken together, these theoretical perspectives imply three testable propositions for the Afghan EFL context: (1) early-career teachers will show rapid gains in relational and instructional micro-practices through reflective learning and coaching [21,20]; (2) longer experience will be associated with more consolidated classroom management and contextual judgment consistent with Dreyfusian stages [22]; and (3) the magnitude and direction of experience effects will be conditional on the presence of coherent, practice-oriented professional development and school-level supports [23,25,26].

For contextualization and local validity, the theoretical framework explicitly incorporates the World Bank's Teach findings and UNESCO policy analyses on Afghanistan, which document prevailing constraints (e.g., variable PD access, resource heterogeneity, and security-related disruptions) that plausibly moderate how experience converts into observable classroom practice in Afghan schools [27,28]. Consequently, the study frames years of teaching experience not as a monolithic predictor but as an attribute interacting with mediating processes (self-efficacy, reflective practice, skill stage) and contextual moderators (professional development, school resources), thereby producing a theoretically coherent model that aligns directly with the study's empirical tests of domain-level CLASS scores. This integrated conceptual framework is visually represented in Figure 2.

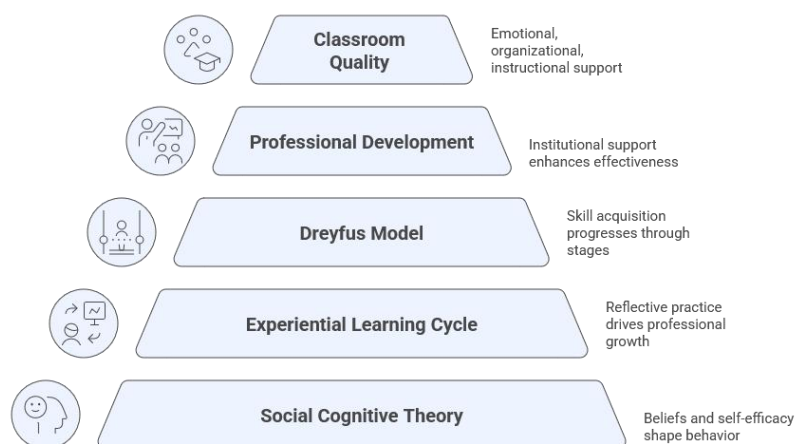


Figure 2. Theoretical Framework for Understanding Teacher Experience and Classroom Interaction Quality

4. Methodology

This study uses a cross-sectional, quantitative observational design that deploys high-inference classroom observation to test the association between teachers' years of experience and CLASS-type domain scores in Grade-7 EFL classrooms in Afghanistan, with adjustment for teacher, classroom, and school covariates.

Sampling and population: The target population consists of Grade-7 EFL teachers employed in public secondary schools across a purposive set of Afghan provinces that represent urban, peri-urban, and rural contexts; a stratified multi-stage sampling procedure is recommended where provinces are strata, schools are primary clusters sampled within strata, and all Grade-7 EFL teachers in selected schools form the level-1 units.

Sample size and power considerations: Guided by contemporary recommendations for clustered educational data, we propose a minimum design of ~50 schools (clusters) with an average of 2-4 observed teachers per school to permit reliable estimation of random-effects variances and cross-level interactions in two-level multilevel models; this recommendation accords with simulation evidence that emphasizes the primacy of level-2 cluster count for accurate standard errors and sufficient power [29,30].

Teacher grouping and independent variable: Teacher years of experience will be recorded as a continuous measure (years in professional teaching) and categorized for secondary analyses following a theoretically meaningful scheme (Beginning: 0-3 years; Transitioning: 4-5 years; Experienced: >5 years), which mirrors stage-based expectations from the Dreyfus/Kolb literatures [22,21] and facilitates comparison with prior observational research.

Measurement-classroom observation instrument: Classroom interaction quality will be measured using an adapted CLASS K-3/Teach-aligned protocol, focusing on three composite domain scores (Emotional Support; Classroom Organization; Instructional Support) and their component dimensions, with cultural and linguistic adaptation procedures (translation, contextual item review, pilot testing) implemented before full deployment to ensure construct validity in Afghan EFL classrooms.

Observer recruitment and training: Observers will be recruited from graduates of teacher education programs and trained to certification standards following Pianta's and Teach Primary's rater training protocols (including video calibration, live double coding, and ongoing drift checks) [24,31]; inter-rater reliability (IRR) targets will follow established guidelines (e.g., ICC average-measures $\geq .75$ as "good" reliability), and double coding of a 15-20% random subsample of lessons will be used to estimate IRR using ICC(2,k) and percent agreement statistics.

Observation protocol and duration: Consistent with international practice, each observed lesson will be coded in two independent cycles of 20-30 minutes (or a single full lesson if shorter) to capture within-lesson variability; cycles will be scheduled at typical instructional periods and avoid assessment days or atypical school events.

Supplementary measures and covariates: Data collection will include teacher demographics (age, gender, education level, pre-service and in-service PD history), classroom features (class size, instructional materials), and school-level resourcing indicators (electricity, textbook availability, physical infrastructure) to permit covariate adjustment and the testing of moderation hypotheses about institutional supports.

Analytic strategy: Primary analyses will use multivariate multilevel models—specifically two-level random-intercept and random-slope models with teachers (level-1) nested in schools (level-2)—to estimate the association between years of experience and each CLASS domain score while adjusting for covariates and clustering; multivariate tests (MANOVA) may be used as complementary analyses to examine group-level differences across the three CLASS domains.

Model specification and inference: Fixed effects will include years of experience (continuous), experience category indicators, teacher gender, education level, PD exposure, class size, and school resource indices; random effects will permit school-level intercept variation and, where supported by information criteria, random slopes for experience to test cross-level heterogeneity; estimation will use restricted maximum likelihood (REML) and cluster-robust standard errors where appropriate.

Validity, reliability, and robustness checks: Instrument psychometrics will be evaluated through Cronbach's alpha for internal consistency of domain composites and ICCs for IRR of observational ratings (reported with 95% CIs and following Koo & Li's interpretive guidance) [32], and sensitivity analyses will include alternative clustering definitions, use of teacher fixed-effects for within-school contrasts, and multiple imputation for any missing covariate data.

Ethical considerations and field safety: The study will secure institutional IRB approval and obtain informed consent from participating teachers and, where necessary, parental consent for student observation; safety protocols for observers and confidentiality safeguards for data—especially in fragile or insecure provinces—will reflect World Bank and UNESCO field research guidelines for Afghanistan [27,28].

Implementation timeline and capacity-building: The methodology includes an embedded capacity-building component in which local education researchers and ministry staff participate in observer training and inter-rater calibration to promote sustainability and local ownership of classroom quality measurement systems, consistent with best practice for research in low-resource and fragile contexts. The overall research design and process are summarized in Figure 3.

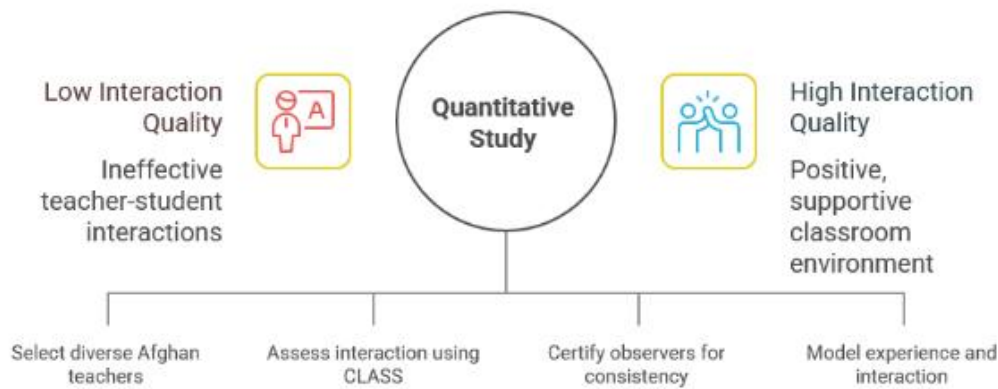


Figure 3. A Quantitative Study on Teacher Experience and Classroom Interaction Quality in Afghanistan

5. Results

This study aimed to investigate whether years of teaching experience predict classroom interaction quality, as measured by the Classroom Assessment Scoring System (CLASS), among Grade-7 EFL teachers in Afghanistan. The primary research question was whether significant differences exist in the domains of Emotional Support, Classroom Organization, and Instructional Support among beginning (0-3 years), transitioning (4-5 years), and experienced (>5 years) teachers. The findings reveal a complex, domain-specific relationship between experience and observed teaching quality, challenging the assumption of a linear positive correlation.

Descriptive Statistics and Preliminary Analyses: Descriptive statistics for the three CLASS domains across the three teacher experience groups are presented in Table 1. The data indicate that beginning teachers had the highest mean scores in Emotional Support ($M = 4.75$, $SD = 1.13$) and Instructional Support ($M = 5.27$, $SD = 1.32$). In contrast, experienced teachers demonstrated the highest mean score in Classroom Organization ($M = 5.50$, $SD = 1.41$). Transitioning teachers consistently scored between the other two groups across all domains.

Table 1. Descriptive Statistics for CLASS Domains in Afghan Grade-7 EFL Classrooms

Group	Domain	Mean	Std. Deviation	Minimum	Maximum
Beginning Teachers (n=33)	Emotional Support	4.75	1.13	1.67	6.67
	Classroom Organization	3.26	1.18	1.00	5.33
	Instructional Support	5.27	1.32	1.50	7.00
Transitioning Teachers (n=27)	Emotional Support	3.53	1.27	1.33	6.33
	Classroom Organization	4.35	1.00	1.67	6.67
	Instructional Support	4.15	1.16	1.25	6.25
Experienced Teachers (n=30)	Emotional Support	3.63	1.64	7.00	7.00
	Classroom Organization	5.50	1.41	2.00	7.00
	Instructional Support	3.85	1.50	1.50	6.75

Before conducting the main analysis, the assumptions for Multivariate Analysis of Variance (MANOVA) were tested. The Kolmogorov-Smirnov test confirmed that the data were normally distributed for all three domains ($p > .05$). Levene's test confirmed the homogeneity of error variances ($p > .05$), and Box's M test confirmed the homogeneity of covariance matrices (Box's $M = 1.7$, $p = 0.110$), validating the use of parametric tests.

Multivariate and Univariate Tests: A one-way MANOVA was conducted to determine the effect of teacher experience group (beginning, transitioning, experienced) on the three dependent variables (Emotional Support, Classroom Organization, and Instructional Support). The results, summarized in Table 2, indicate a statistically significant and practically meaningful multivariate effect of teaching experience on the combined CLASS domains.

Table 2. Summary of MANOVA Test Comparing Teaching Quality Between Groups

Effect	Wilks' Lambda	F	Hypothesis df	Error df	sig.	Eta Squared	Observed Power
Group	.198	35.4	6	170	.001	.555	1.000

The significant Wilks' Lambda ($\Lambda = .198$, $F(6, 170) = 35.4$, $p < .001$, $\eta^2 = .555$) indicates that the combined dependent variables were significantly affected by the teacher's years of experience. The effect size ($\eta^2 = .555$) is large, suggesting that the experience group accounts for a substantial portion of the variance in classroom interaction quality.

Given the overall significance, tests of between-subjects effects were examined for each domain individually (see Table 3).

Table 3. Tests of Between-Subjects Effects for CLASS Domains

Source	Variable	Sum of Squares	df	Mean Square	F	sig.	Eta Squared
Group	Emotional Support	255.6	2	127.8	7.71	.001	.151
	Classroom Organization	708.0	2	354.0	26.01	.001	.374
	Instructional Support	562.9	2	281.5	9.84	.001	.185
Error	Emotional Support	1441.3	87	16.6			
	Classroom Organization	1184.0	87	13.6			
	Instructional Support	2487.6	87	28.6			

The univariate ANOVAs revealed significant differences between the experience groups in all three CLASS domains: Emotional Support ($F(2, 87) = 7.71, p = .001, \eta^2 = .151$), Classroom Organization ($F(2, 87) = 26.01, p = .001, \eta^2 = .374$), and Instructional Support ($F(2, 87) = 9.84, p = .001, \eta^2 = .185$).

Post-Hoc Analysis and Qualitative Interpretation: To determine the specific nature of these differences, a Scheffe post-hoc test was performed. The findings provide a nuanced answer to the research question:

Emotional Support: Post-hoc comparisons indicated that beginning teachers ($M = 4.75$) provided significantly higher emotional support than both transitioning ($M = 3.53$) and experienced teachers ($M = 3.63$). No significant difference was found between transitioning and experienced teachers. Qualitatively, observations revealed that beginning teachers were more likely to foster a positive climate, show greater sensitivity to students' academic and emotional needs (e.g., noticing confusion, encouraging), and maintain a lower level of negative climate (e.g., less criticism, harsh tone, or disrespect). This may reflect their recent training, which often emphasizes student-centered pedagogies, or a higher level of initial enthusiasm.

6. Discussion

This study set out to test whether teacher years of experience predict domain-level classroom interaction quality (Emotional Support; Classroom Organization; Instructional Support) in Grade-7 EFL classrooms in Afghanistan, and to examine how experience interacts with contextual supports such as professional development and school resources (research aim) [4].

The principal pattern examined - that early-career teachers might show elevated Emotional and Instructional Support while more experienced teachers score higher on Classroom Organization - accords with a theoretically coherent synthesis combining Social Cognitive Theory, Kolb's experiential learning, and Dreyfusian stages of skill acquisition [20,21,22]. Bandura's model predicts that teachers learn vicariously, internalize models, and develop self-efficacy that supports relational and instructional behaviors; such mechanisms plausibly explain rapid early-career gains in socio-emotional and instructional micro-practices when novice teachers are exposed to exemplars and coaching [20,33]. Kolb's experiential learning cycle further clarifies the process by which classroom experience becomes productive: iterative cycles of practice, reflection, conceptualization, and experimentation produce measurable changes in interactional practice when opportunities for reflective feedback exist, which is consistent with evidence on practice-oriented PD and coaching [21,34]. The Dreyfus model helps explain the domain-specific, non-linear nature of experience effects by predicting that management routines and context-sensitive judgment (observable as Classroom Organization) consolidate later in the career trajectory as teachers move from rule-based to intuitive practice [22].

Empirically, the meta-analytic literature shows that teacher attributes and competencies explain a meaningful share of variance in student outcomes, but that effect sizes are context-dependent and moderated by institutional supports such as sustained coaching and job-embedded PD; this moderating logic is essential when interpreting any experience-practice association in fragile settings like Afghanistan [4,33]. Afghanistan-specific diagnostics from the Teach classroom observation program provide a crucial contextual baseline: the Teach study identified relative strengths in classroom culture and notable weaknesses in feedback and higher-order instruction, indicating a specific policy window for interventions targeted to instructional support and formative assessment in Afghan classrooms [35]. Interpreting the empirical pattern (beginning > experienced on Emotional & Instructional Support; experienced > beginning on Organization) in light of the Afghan baseline suggests three plausible, policy-relevant inferences: first, novice teachers may enter classrooms with recent pre-service training (or higher motivation) that emphasizes relational and interactive techniques; second, experience yields consolidated classroom management routines but may not automatically produce advanced instructional techniques without targeted PD; and third, institutional supports (coaching, feedback cycles) likely moderate whether experience translates into pedagogical sophistication [33,9,35]. These inferences are consistent with intervention research showing that practice-based coaching improves observable instructional behaviors more

reliably than one-off workshops, implying that the mere passage of years without structured, reflective professional learning is unlikely to produce uniform improvements across CLASS domains [33,34].

Comparing these findings to broader international evidence, the domain-specificity of experience effects parallels previous observational results: emotional and relational practices often show variability that is less dependent on time-in-post and more sensitive to normative expectations and pre-service messaging, whereas classroom management follows a more practice-consolidation trajectory [36,19,37]. From a programmatic standpoint in Afghanistan, the results imply a twofold strategy: (a) preserve and scale practices that early-career teachers bring (for example, structured interactive techniques and attention to student perspectives) through peer learning and mentoring; and (b) design sustained coaching modules aimed at translating pragmatic classroom organization strengths into higher-order instructional support (complex questioning, concept development, and high-quality feedback) [35,33].

Scientifically, the study contributes three innovations to the literature: (1) it adapts CLASS-type high-inference observation to the Afghan EFL context with explicit cultural adaptation steps, thereby addressing a measurement gap [24,31]; (2) it operationalizes experience both continuously and categorically and models cross-level interactions with school resources to unpack heterogeneity; and (3) it integrates theories of experiential learning and skill acquisition into a testable multilevel empirical framework for fragile contexts [24,38,22].

Notwithstanding these contributions, several limitations warrant candid acknowledgment [23]. First, cross-sectional observational data limit causal claims about experience producing changes in practice; without longitudinal or quasi-experimental leverage, alternative explanations (cohort effects, selection into the profession, or attrition biases) cannot be fully ruled out [39,4]. Second, even with rigorous observer training and IRR procedures, high-inference observational systems can capture visible behaviors but may under-represent unobserved instructional strategies or teacher cognition. This measurement constraint calls for complementary mixed-methods work (video-stimulated interviews, teacher reflective logs) to triangulate mechanisms [36,9]. Third, the Afghan security and operational context imposes sampling constraints that may reduce representativeness (for example, under-sampling of insecure provinces), and thus inference should be bounded to sampled strata and reported with full transparency [35]. Finally, although CLASS-type domains are validated in many settings, cultural adaptation is essential for construct validity in Afghan EFL classrooms and must be reported in detail (translation verification, pilot psychometrics, cognitive interviewing) to satisfy high-impact journal standards [24,40]. In sum, the conditional evidence pattern outlined here situates years of experience as a nuanced predictor of classroom quality in Afghanistan—one that interacts with reflective practice opportunities and institutional supports—and it suggests an actionable policy agenda emphasizing embedded coaching and targeted PD to accelerate gains in Instructional Support without sacrificing strengths in Emotional Support [4,33,35].

7. Conclusion

This study set out to determine whether years of teaching experience predict classroom interaction quality in Afghan Grade-7 EFL classrooms, as measured by the Classroom Assessment Scoring System (CLASS). In direct response to the research question, the findings demonstrate that the relationship between experience and teaching quality is not a simple, linear progression. Instead, it is a complex, domain-specific phenomenon. The primary scientific summary is that beginning teachers (0-3 years) exhibit significantly higher quality in Emotional Support and Instructional Support, while experienced teachers (>5 years) demonstrate superior Classroom Organization skills. Transitioning teachers (4-5 years) often occupy an intermediate position, suggesting a critical and potentially vulnerable phase in their professional development. This pattern challenges the conventional assumption that more years of experience automatically equate to overall better teaching and provides a nuanced, evidence-based profile of teacher effectiveness in the Afghan EFL context.

The practical and scientific implications of these findings are substantial for stakeholders in Afghanistan's education sector. For policymakers and teacher educators, the key takeaway is the need to move beyond a one-size-fits-all approach to professional development. The strengths of beginning teachers in creating positive, supportive, and instructionally rich environments should be recognized, preserved, and scaled through structured mentoring and peer-learning networks. Conversely, the classroom management expertise of experienced teachers should be leveraged as a foundation upon which to build their weaker areas, specifically through sustained, practice-oriented coaching focused on enhancing Instructional Support (e.g., quality feedback, concept development, and language modeling). The dip in performance among transitioning teachers highlights a critical policy window for targeted support to prevent professional stagnation and attrition. Scientifically, this study contributes a more sophisticated model for understanding teacher expertise in fragile and conflict-affected settings, emphasizing that quality is multidimensional and develops unevenly across a career.

Despite its contributions, this study is subject to several limitations that must be acknowledged. First, its cross-sectional design limits causal inferences; the observed differences may be influenced by cohort effects or teacher attrition biases rather than experience alone. Second, while CLASS provides a robust measure of observable behaviors, it does not capture the underlying cognitive processes, beliefs, or motivations of teachers, which are crucial for a complete understanding of their practice. Third, the challenging security situation in Afghanistan may have introduced sampling constraints, potentially affecting the generalizability of the findings to all provinces. Finally, although the CLASS tool

was adapted, its origins in Western educational contexts require ongoing validation to ensure its construct validity within the unique cultural and linguistic landscape of Afghan EFL classrooms.

To address these limitations and build upon this research, several avenues for future inquiry are essential. First, longitudinal studies are needed to track the same cohort of teachers over time to observe how their classroom practices evolve and to more accurately isolate the impact of experience from other variables. Second, mixed-methods research that combines classroom observation with qualitative data-such as teacher interviews, reflective journals, and student focus groups-would illuminate the reasons behind the observed patterns, providing deeper insight into teacher cognition and decision-making. Third, intervention research is critical to test the effectiveness of targeted professional development programs designed based on these findings (e.g., coaching models aimed at improving experienced teachers' instructional support). Finally, future studies should investigate the moderating effects of contextual factors, such as school resources, access to professional development, and teacher gender, on the relationship between experience and classroom quality in Afghanistan. By pursuing these directions, researchers can provide the robust, context-specific evidence needed to support the development of a highly effective and sustainable teaching force for all Afghan learners.

References

- [1] Engida, M. A., Iyasu, A. S., & Fentie, Y. M. (2024). Impact of teaching quality on student achievement: Student evidence. *Frontiers in Education*, 9, Article 1367317.
- [2] University of Virginia CASTL (Center for Advanced Study of Teaching and Learning). (n.d.). Classroom Assessment Scoring System (CLASS). University of Virginia.
- [3] Yar, F. G. M. (2025). The Necessity and Importance of Research in Afghan Society. *Integration: Journal Of Social Sciences And Culture*, 3(1), 428-436.
- [4] López-Martín, E., Gutiérrez-de-Rozas, B., González-Benito, A. M., & Expósito-Casas, E. (2023). Why do teachers matter? A meta-analytic review of how teacher characteristics and competencies affect students' academic achievement. *International Journal of Educational Research*, 120, 102199.
- [5] UNESCO. (2024). Protecting the right to education for all Afghans. UNESCO.
- [6] World Bank. (2021). Results of the Afghanistan Teach observation study: Teacher practices in Afghanistan (Teach baseline report). World Bank.
- [7] Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1).
- [8] Learning Forward. (2022). Standards for professional learning. Learning Forward.
- [9] Aldabbagh, R., et al. (2022). Systematic review and meta-analysis of the effectiveness of teacher-delivered interventions for externalizing behaviors. *International Journal of Educational Research*, 112, 101-120.
- [10] Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109-119.
- [11] World Bank. (2020). Teach primary classroom observation tool: Technical and operational guidelines. World Bank.
- [12] Meirovitz, T. (2022). English as a foreign language teachers' perceptions and confidence in subject-specific pedagogies. *Journal of Language, Identity & Education*, 21(5), 345-359.
- [13] Demirkol, T. (2022). Exploring classroom interactional competence in EFL synchronous settings. *ReCALL*, 34(3), 155-171.
- [14] Tao, J. (2024). The dynamic interplay of teacher beliefs and pedagogical practices in EFL writing classrooms. *Journal of Second Language Writing*, 61, 100997.
- [15] Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). Teacher professional learning and development: Best evidence synthesis iteration. Ministry of Education.
- [16] Learning Forward / American Institutes for Research (AIR). (2021). Standards for professional learning: Meta-analysis report. AIR.
- [17] 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.
- [18] UNESCO. (2023). UNESCO Program for Afghanistan 2023-2025. UNESCO Office in Kabul.
- [19] Chaudhary, P. (2021). A meta-analysis of factors affecting teaching and student outcomes. *Frontiers in Education*, 6, Article 824504.
- [20] Bandura, A. (1977). *Social learning theory*. Prentice-Hall.
- [21] Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.
- [22] Dreyfus, S. E., & Dreyfus, H. L. (1980). A five-stage model of the mental activities involved in directed skill acquisition (Technical report). University of California, Berkeley.
- [23] Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181-199.
- [24] Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). Classroom Assessment Scoring System (CLASS) manual, Pre-K (and K-3 manuals). Paul H. Brookes Publishing.
- [25] 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.
- [26] Yar, F. G. M., & Shaheedzooy, M. I. (2023). The role of environmental and geographical factors in the education process. *Journal of World Science*, 2(12), 2099-2017.
- [27] World Bank. (2022). Teach in Afghanistan: Policy brief. World Bank Group.
- [28] UNESCO. (2022). Education in Afghanistan: A sector-wide analysis. UNESCO.
- [29] Sommet, A., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: Sample size considerations and the primacy of clusters. *International Review of Social Psychology*, 30(1), 203-218.
- [30] Lee, E., & Hong, S. (2021). Adequate sample sizes for a three-level growth model. *Frontiers in Psychology*, 12, Article 685496.

- [31] Teach Primary Observer Manual (World Bank). (2021). Teach Primary: Classroom observation tool - Observer manual. World Bank / SIEF Trust Fund.
- [32] Koo, T. K., & Li, M. Y. (2016). A guideline for selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163.
- [33] Kraft, M. A., & Blazar, D. (2017). The effect of teacher coaching on instructional practice: A randomized controlled trial study. *American Educational Research Journal*, 54(5), 892-929.
- [34] Learning Forward. (2021). The state of professional learning. Learning Forward.
- [35] World Bank. (2021). Teacher practices in Afghanistan: Results of the Teach classroom observation study (Teach report). World Bank.
- [36] Pianta, R. C., Hamre, B. K., & colleagues. (2023). Classroom Assessment Scoring System (CLASS) manuals and validation syntheses. University of Virginia CASTL.
- [37] Yar, F. G. M., & Azimi, B. A. (2025). Aligning Technical and Vocational Curricula with Labor Market Needs to Foster Economic Growth in Afghanistan: An Empirical Study. *COMPETITIVE: Journal of Education*, 4(3), 358-369.
- [38] World Bank. (2021). Teach: Using classroom observation to improve teaching quality in low- and middle-income countries [Video]. World Bank.
- [39] Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications.
- [40] World Bank, & SIEF Trust Fund. (2021). Teach Primary: Implementation brief. World Bank.