

The Impact of In-Service Teacher Training Programs on EFL Teachers' Beliefs and Practices of Formative Assessment in K-12 Classrooms

Lý Thị Bích Phương (Corresponding author)

Faculty of Foreign Languages, Can Tho University
Campus II, 3/2 Street, Ninh Kiều Ward, Can Tho City, Vietnam
E-mail: ltbphuong@ctu.edu.vn

Trần Ngọc Bảo Châu

Faculty of Foreign Languages, Can Tho University, Vietnam

Nguyễn Thị Khánh Đoan

Faculty of Foreign Languages, Can Tho University, Vietnam

This study is funded by Can Tho University, Code: CTCS2024-07-05

Abstract

This study investigates the impact of an in-service formative assessment (FA) training program, organized by Can Tho University, on the beliefs and classroom practices of Vietnamese EFL teachers across primary, lower-secondary, and upper-secondary levels. A structured questionnaire combining closed- and open-ended items was administered to 114 K-12 English teachers who had completed the training; this paper reports both the closed-ended quantitative findings and a thematic analysis of the six open-ended items. The questionnaire measured training effectiveness, FA beliefs, technique-use frequency before and after training, perceived student impact, and contextual factors, with excellent internal consistency (Cronbach's $\alpha = .895-.935$). Paired-samples *t*-tests showed statistically significant increases in reported use of all eleven FA techniques after training ($p < .001$), with moderate-to-large effect sizes (Hedges' $g = .60-.78$), with the largest raw gains for exit tickets, observation checklists, error-correction codes, and feedback techniques. Teachers rated trainers' expertise and materials as most valuable, while lack of time and large class sizes were the leading barriers—corroborated by open-ended responses describing large classes as a key constraint. Qualitatively, teachers described the training mainly as expanding their practical technique repertoire rather than shifting underlying assessment beliefs. These findings are consistent with in-service training serving as an effective mechanism for shifting EFL teachers' reported FA practices under Vietnam's General Education Curriculum 2018, pointing to the need for follow-up coaching and institutional support to sustain these gains.

Keywords: in-service teacher training, formative assessment, EFL teachers, teacher beliefs, K-12 education, Vietnam

DOI: 10.7176/JEP/17-9-08

Publication date: September 30th 2026

1. Introduction

Testing and assessment occupy a central place in Vietnam's ongoing reform of English language teaching (ELT). Decision 2080/QĐ-TTg (2017) on the National Foreign Languages Project identifies testing and assessment as one of eight priority areas for reform, and a succession of Ministry of Education and Training (MOET) circulars—Circular 26/2020/TT-BGDĐT, Circular 27/2020/TT-BGDĐT, and Circular 22/2021/TT-BGDĐT—have progressively required school teachers to move beyond summative testing toward assessment that is embedded in day-to-day teaching and learning. Under the General Education Curriculum 2018 (English Curriculum), formative assessment (FA) is explicitly framed as a tool for providing ongoing feedback on students' communicative competence, guiding both student learning and teachers' instructional decisions, and it is expected to combine teacher assessment, peer assessment, and self-assessment in varied, continuous forms.

Realizing this policy vision, however, depends heavily on teachers' capacity to translate FA principles into classroom techniques—a capacity that in-service training is meant to build. While a growing international

literature suggests that in-service training can improve teacher efficacy, pedagogical skill, and even student achievement (Popova et al., 2016; Öztürk, 2019; Karagiorgi & Symeou, 2007; Ulla, 2018; Mugarura et al., 2022), few studies in Vietnam have examined how in-service training specifically shapes EFL teachers' formative assessment beliefs and practices. This gap motivated the present study, which reports quantitative findings from a survey of 114 K-12 EFL teachers who completed a formative-assessment-focused in-service training program organized by Can Tho University.

2. Literature Review

2.1 Formative Assessment in the Vietnamese EFL Context

In the Vietnamese context, MOET has issued several documents to operationalize FA in the classroom, including Circular 5333/BGDĐT-GDTrH (2014) on competency-based testing and Circular 3333/BGDĐT-GDTrH (2016) on proficiency test specifications. These documents encourage a diverse range of FA techniques, including observation, academic records, oral and written comments, peer assessment, and self-assessment, which align with the communicative, learner-centered orientation of the 2018 English Curriculum.

2.2 Conceptualizing Formative Assessment

Formative assessment is broadly defined as assessment carried out during the learning process with the explicit purpose of generating feedback that teachers and students can use to adjust teaching and learning while they are still happening (Black & Wiliam, 1998). In their influential review, Black and Wiliam (1998) synthesized evidence from over 250 studies and concluded that classroom-based formative assessment practices produce some of the largest measurable gains in student achievement of any known educational intervention. Black and Wiliam (2009) later refined this definition, framing formative assessment as a process in which evidence about learning is elicited and interpreted, and used by teachers, learners, or their peers to make decisions about next instructional steps that are better, or better founded, than the decisions they would have made without that evidence. Heritage (2010) further emphasized that effective formative assessment requires teachers to possess not only assessment literacy but also a clear understanding of learning progressions, so that feedback can be targeted precisely to where a learner currently stands. These theoretical accounts underline that FA is not a fixed set of techniques but an evidence-based instructional cycle, a distinction that motivates the present study's focus on both teachers' underlying beliefs about FA and their observable frequency of technique use.

2.3 Teacher Beliefs and the Beliefs–Practice Relationship

Teacher beliefs have long been recognized as a powerful, if often implicit, filter through which teachers interpret training input and decide how to act in the classroom (Jordan, 2018; Grapin et al., 2024; Manuel & Elder, 2025; Jacobson et al., 2022). Beliefs are far more influential than knowledge in shaping how teachers plan, interpret, and evaluate their teaching (Jordan, 2018; Hanna et al., 2019; Herrington et al., 2016; Wijesundera & Wijethunga, 2021), precisely because beliefs are typically formed early, are resistant to change (Valcke et al., 2010; Moghadam, 2015), and operate largely outside conscious awareness (Bakshi, 2020). Borg (2003), in a comprehensive review of teacher cognition research in language teaching, similarly concluded that what language teachers believe about teaching and assessment substantially shapes what they do in the classroom, even when that relationship is not always linear or immediate. More recently, Fives and Buehl (2012) proposed that beliefs serve three key functions for teachers, filtering new information, framing problems, and guiding action, which helps explain why training programs that succeed in shifting beliefs are more likely to produce durable changes in practice than those that target technique acquisition alone. This body of work provides the theoretical rationale for measuring both beliefs and self-reported practice as distinct but related constructs, rather than treating practice change as a straightforward proxy for belief change.

Internationally, Popova et al. (2022) synthesized evidence that in-service teacher training can improve student learning outcomes, finding that programs which “link participation to career incentives, have a specific subject focus, incorporate lesson enactment in the training, and include initial face-to-face training tend to show higher student learning gains” (Herrington et al., 2016). Öztürk (2019) argued that identifying which features of in-service training interventions are most effective is a prerequisite for understanding their impact on student learning. Mohammadi and Moradi (2017) surveyed 86 EFL teachers in Iran and found that professional development workshops could shift teachers' beliefs about continuous professional development, particularly when programs were tailored to teachers' actual needs. Debreli (2016) similarly demonstrated that teachers' beliefs and practical knowledge can change through structured training and practicum experiences. Within Vietnam specifically, Nguyen et al. (2020) found that professional development programs positively affected

both teacher efficacy and student achievement, while Ha et al. Duong (2021) reported that the training program improved teachers' pedagogical skills, including their ability to design assessments.

2.4 In-Service Training Design and Effectiveness

Beyond documenting that in-service training can shift teacher practice, a parallel strand of research has sought to specify which features of training make it effective. Desimone (2009) identified five core features of high-quality professional development—content focus, active learning, coherence with existing standards and policies, sustained duration, and collective participation—that have become widely accepted benchmarks for evaluating training programs. Kennedy (2016) similarly argued that the design features of professional development matter more than its structural form, emphasizing the importance of content relevance and sustained opportunities for practice over surface characteristics such as program length or delivery mode. These design-focused frameworks provide a lens for interpreting not only whether the present training program changed teachers' reported practice, but why: as reported in Section 6.4, teachers rated trainer expertise, theoretical content, and practical examples as the most valuable aspects of the program, a pattern that aligns closely with Desimone's and Kennedy's emphasis on content focus and active, example-driven learning.

Building on this body of theoretical and empirical work, the present study contributes statistically validated, effect-size-based evidence on the extent to which EFL teachers' perceived beliefs and self-reported use of specific formative assessment techniques change following in-service training in the Vietnamese context.

3. Theoretical and Conceptual Framework

This study is guided primarily by Clarke and Hollingsworth's (2002) Interconnected Model of Professional Growth (IMPG), which conceptualizes teacher professional development as movement across four interrelated domains: the External Domain (sources of information, stimulus, or support external to the teacher, such as a training program or the school environment), the Personal Domain (teachers' knowledge, beliefs, and attitudes), the Domain of Practice (professional experimentation—the techniques teachers actually try out in the classroom), and the Domain of Consequence (salient outcomes, such as perceived student engagement and learning gains). Change in one domain is theorized to translate into change in another through the mediating processes of enactment (translating belief or knowledge into action) and reflection (deriving belief or knowledge from the outcomes of action). Unlike simpler linear models, the IMPG explicitly allows for change to travel in multiple directions—for example, a positive change in student outcomes can reinforce or further shift teacher beliefs—which aligns with Guskey's (2002) observation that belief change often follows, rather than precedes, evidence of improvement in practice or outcomes.

Mapped onto the present study, the in-service training program constitutes the External Domain; teachers' reported beliefs about formative assessment (Part III of the questionnaire) represent the Personal Domain; teachers' self-reported frequency of using specific FA techniques before and after training (Part IV.A) represent the Domain of Practice; and teachers' perceptions of the training's impact on students (Part IV.D) represent the Domain of Consequence. This framework motivates treating belief change and practice change as related but empirically distinct outcomes of training, rather than assuming that an increase in technique use is sufficient evidence that underlying beliefs have shifted.

4. Research Objectives and Questions

This study investigates the impact of in-service teacher training on EFL teachers' beliefs and practices of formative assessment in K-12 classrooms in the context of Can Tho University's teacher training program. Specifically, it addresses the following research questions:

1. How do EFL teachers perceive the effectiveness of the in-service training program in enhancing their formative assessment practices and professional development?
2. To what extent did the training program change the frequency with which teachers use specific formative assessment techniques?
3. What do teachers identify as the most valuable aspects of the training, and what challenges do they face in implementing formative assessment in their classrooms?

5. Methodology

5.1 Research Design

This paper reports findings from a single structured questionnaire examining the impact of in-service teacher training on formative assessment beliefs and practices. Alongside its closed-ended Likert-scale and multiple-response items, the questionnaire included six open-ended questions inviting teachers to describe their experiences with formative assessment and the training program in their own words. Sections 6.1–6.6 report the closed-ended, quantitative findings; Section 6.7 reports a thematic content analysis of the open-ended responses as a complementary qualitative component of the same instrument.

5.2 Participants

Participants were selected using three criteria: (a) currently teaching English at the primary, lower-secondary, or upper-secondary level in a Vietnamese public school; (b) having completed the Can Tho University in-service formative assessment training under the National Foreign Languages Project, verified via a screening item in Part I of the questionnaire that retained only affirmative responses; and (c) voluntary, informed consent, with anonymity and the right to withdraw assured. No further exclusions were applied by gender, age, experience, or education. Because the training itself had been delivered across the Mekong Delta and south-central regions, the achieved sample was naturally concentrated in the eight provinces and cities served by the program rather than through deliberate geographic sampling. Applying these criteria yielded 114 eligible respondents (100% of the analytic sample); full demographic characteristics are reported in Section 6.1 and Table 1.

5.3 Instrument

Data were collected using a structured, bilingual (Vietnamese/English) questionnaire developed by the research team, organized into seven parts. Following a demographic section (Part I), the instrument comprised six Likert-scale subscales and several closed-ended, multiple-response items: (a) perceived effectiveness of the training program (Part II; 10 items, 5-point agreement scale, including the multiple-response item on the most valuable aspects of the training); (b) beliefs about formative assessment (Part III; 11 items, 5-point agreement scale); (c) frequency of use of 11 formative assessment techniques—exit tickets, think-pair-share, classroom questioning, peer assessment, self-assessment, feedback techniques, error analysis, digital tools, error-correction codes, observation checklists, and portfolio assessment—rated separately for before and after the training (Part IV.A; 5-point frequency scale, “never” to “always”); (d) perceived impact of formative assessment on students (Part IV.D; 8 items, 5-point agreement scale); (e) contextual factors affecting implementation (Part V; 8 items, 5-point agreement scale); and (f) multiple-response items on implementation challenges and support needs (Part IV.C), technology use (Part VI), and future professional development needs (Part VII). The questionnaire also included six open-ended questions: suggestions for improving the training program (Part II, Q12); how teachers’ understanding of formative assessment had changed (Part III, Q12); a concrete example of formative assessment’s impact on students (Part IV.D, Q23); how teachers’ school environment influenced their implementation of FA techniques (Part V, Q9); a successful example of using technology for FA (Part VI, Q4); and how teachers’ experience with FA had shaped their overall teaching philosophy (Part VII, Q8). These are analyzed thematically in Section 6.7 as a qualitative complement to the quantitative findings reported in Sections 6.1–6.6.

5.4 Data Collection Procedure

The questionnaire was administered to teachers who had completed the in-service training program organized by Can Tho University. Participation was voluntary and anonymous, and respondents were informed that their answers would be used solely for research purposes.

5.5 Data Analysis

Data were analyzed using SPSS. Descriptive statistics (frequencies and percentages) characterized the sample and the multiple-response items. Internal consistency of each Likert-scale subscale was assessed using Cronbach’s alpha. Paired-samples t-tests compared teachers’ self-reported frequency of using each of the 11 formative assessment techniques before and after the training program, with Cohen’s *d* and Hedges’ correction reported as effect sizes.

Responses to the six open-ended questions were analyzed using inductive thematic content analysis. Responses were first read in full in the original Vietnamese; recurring ideas were then grouped into candidate themes, and theme frequencies were verified through systematic keyword searches across the full response set for each question. Representative responses were translated into English for reporting. This analysis was conducted by a single coder, without an independent second coder or formal inter-rater reliability check; findings should therefore be read as a systematic but single-coder content analysis rather than a fully verified thematic analysis.

6. Results

6.1 Demographic Characteristics of Participants

As shown in Table 1, respondents were predominantly female ($n = 81$, 71.1%), with male teachers comprising 28.9% ($n = 33$) of the sample. The sample was fairly evenly distributed across teaching levels, with primary school teachers making up 31.6% ($n = 36$), lower-secondary teachers 33.3% ($n = 38$), and upper-secondary teachers 35.1% ($n = 40$). In terms of age, respondents were concentrated in the mid-to-late career range: 46.5% ($n = 53$) were aged 41–50, 32.5% ($n = 37$) were aged 31–40, and 14.9% ($n = 17$) were over 50, with only a small proportion under 30 (6.2% combined). This age profile is mirrored in years of teaching experience, where the majority of respondents (53.5%, $n = 61$) had more than 20 years in the classroom, while only 11.4% ($n = 13$) had fewer than 5 years of experience. Nearly all participants held a bachelor's degree (92.1%, $n = 105$), with the remainder (7.9%, $n = 9$) holding a master's degree. All 114 respondents (100%) confirmed having attended the Can Tho University in-service formative assessment training program. Geographically, respondents were drawn from eight provinces and cities in the Mekong Delta and south-central regions, with the largest concentrations from Ca Mau (32.5%), Binh Phuoc (19.3%), and Can Tho (16.7%), and smaller representations from Ninh Thuan (15.8%), Soc Trang (7.0%), Hau Giang (4.4%), Tra Vinh (2.6%), and Ben Tre (1.8%), reflecting the uneven geographic reach of the training program noted in the Limitations section.

Table 1. Demographic Characteristics of Participants (N = 114)

Variable	Category	N	%
Gender	Male	33	28.9
	Female	81	71.1
Teaching level	Primary	36	31.6
	Lower secondary	38	33.3
	Upper secondary	40	35.1
Age	Under 25	2	1.8
	25–30	5	4.4
	31–40	37	32.5
	41–50	53	46.5
	Over 50	17	14.9
Teaching experience	Under 5 years	13	11.4
	5–10 years	14	12.3
	11–15 years	18	15.8
	16–20 years	8	7
	Over 20 years	61	53.5
Education	Bachelor's degree	105	92.1
	Master's degree	9	7.9

6.2 Reliability of the Measurement Scales

All six Likert-scale subscales demonstrated good-to-excellent internal consistency, as presented in Table 2. Internal consistency was assessed using Cronbach's alpha (α), a widely used measure of reliability that evaluates the degree to which items within a subscale are interrelated (Herrington et al., 2016; Jordan, 2018). Values equal to or above .70 were considered satisfactory (Grapin et al., 2024; Manuel & Elder, 2025; Jacobson et al., 2022), consistent with established psychometric standards Hanna et al., 2019). For scales with fewer items, alpha values between 0.80 and 0.90 have been considered "ideal" (Herrington et al., 2016).

As shown in Table 2, the Cronbach's alpha coefficients for the six subscales ranged from .895 to .935, indicating that all subscales exceeded the .80 threshold for good reliability and that four of the six surpassed the .90 threshold for excellent reliability. Specifically, the perceived impact of formative assessment on students subscale yielded the highest reliability ($\alpha = .935$), followed by perceived effectiveness of training ($\alpha = .928$), frequency of FA technique use before training ($\alpha = .917$), and frequency of FA technique use after training ($\alpha = .912$). The beliefs about formative assessment subscale ($\alpha = .898$) and contextual factors subscale ($\alpha = .895$) also demonstrated good internal consistency. Internal consistency, which reflects the degree of interrelatedness of items within a measurement instrument or subscale (Herrington et al., 2016), was assessed using Cronbach's alpha (α), the most widely used estimate of reliability in education, public health, and the social and behavioural sciences (Jordan, 2018; Grapin et al., 2024). A Cronbach's alpha of 0.70 or higher is generally considered acceptable (Herrington et al., 2016; Manuel & Elder, 2025; Jacobson et al., 2022; Hanna et al., 2019), while values above 0.90 indicate excellent internal consistency (Herrington et al., 2016; Wijesundera & Wijethunga, 2021). However, reliability coefficients exceeding 0.90 may also indicate item redundancy in measuring the intended construct (Valcke et al., 2010). The consistently high reliability coefficients across all subscales confirm that the items within each scale measured their respective constructs with a high degree of internal coherence, thereby supporting the trustworthiness of the quantitative findings derived from the instrument.

Table 2. Reliability (Cronbach's Alpha) of the Measurement Scales

Scale	No. of items	Cronbach's α
Perceived effectiveness of training	10	.928
Beliefs about formative assessment	11	.898
Frequency of FA technique use – before training	11	.917
Frequency of FA technique use – after training	11	.912
Perceived impact of FA on students	8	.935
Contextual factors	8	.895

6.3 Changes in Formative Assessment Practices Before and After Training

Paired-samples t-tests compared teachers' self-reported frequency of use (1 = never to 5 = always) of 11 formative assessment techniques before and after the training. As shown in Table 3, the frequency of use increased significantly for every technique (all $p < .001$, $df = 113$), with the mean composite frequency rising from 3.15 before training to 4.04 after training, an average gain of 0.89 scale points. Effect sizes (Hedges' g) ranged from .60 to .78, indicating moderate-to-large practical significance across all techniques.

Ranked by raw mean difference (Δ), the largest gains were observed for exit tickets ($M = 2.82$ to 3.90 , $\Delta = 1.08$, $g = .757$), observation checklists ($M = 2.99$ to 4.03 , $\Delta = 1.04$, $g = .625$), error-correction codes for writing ($M = 2.99$ to 3.98 , $\Delta = 0.99$, $g = .674$), and structured feedback techniques ($M = 3.24$ to 4.21 , $\Delta = 0.97$, $g = .604$). Notably, error analysis showed a more moderate raw gain ($M = 3.40$ to 4.18 , $\Delta = 0.77$), but the single largest standardized effect size among the 11 techniques ($g = .779$), reflecting a comparatively small standard deviation on this item rather than a larger absolute change. Classroom questioning techniques showed the smallest, though still statistically significant, gain ($M = 3.54$ to 4.28 , $\Delta = 0.75$, $g = .651$), likely reflecting a ceiling effect, as questioning already had the highest pre-training mean of the 11 techniques.

Table 3. Paired-Samples t-Test Results for Frequency of FA Technique Use Before and After Training (N = 114, df = 113)

Technique	M before	M after	Mean diff.	t	Hedges' g	p
Exit tickets	2.82	3.90	-1.08	-15.27	.757	< .001
Think-pair-share	3.16	4.01	-0.85	-12.63	.722	< .001
Classroom questioning	3.54	4.28	-0.75	-12.26	.651	< .001
Peer assessment	3.15	3.91	-0.76	-12.97	.630	< .001
Self-assessment	3.26	4.10	-0.83	-13.97	.639	< .001
Feedback techniques	3.24	4.21	-0.97	-17.28	.604	< .001
Error analysis	3.40	4.18	-0.77	-10.62	.779	< .001
Digital tools	3.05	3.99	-0.94	-13.69	.735	< .001
Error-correction codes	2.99	3.98	-0.99	-15.76	.674	< .001
Observation checklists	2.99	4.03	-1.04	-17.74	.625	< .001
Portfolio assessment	3.02	3.84	-0.83	-13.17	.671	< .001

6.4 Valuable Aspects of the Training Program

When asked which aspects of the training they found most valuable (multiple selections permitted), teachers most frequently cited the input from trainers (75.4%), followed closely by the theoretical knowledge of formative assessment and the training materials provided (64.0% each), practical examples of FA techniques (61.4%), discussion with peers (58.8%), and hands-on practice opportunities (53.5%). Only 2.6% selected "other."

Table 4. Most Valuable Aspects of the Training Program (Multiple Response, N = 114)

Aspect	N	%
Input from trainers	86	75.4
Theoretical knowledge of FA	73	64
Training materials provided	73	64
Practical examples of FA techniques	70	61.4
Discussion with peers	67	58.8
Hands-on practice opportunities	61	53.5
Other	3	2.6

6.5 Challenges in Implementing Formative Assessment

The most commonly reported challenge to implementing formative assessment was lack of time (71.1%), followed by large class size (64.0%), student resistance/non-cooperation (44.7%), examination pressure (40.4%), and technological limitations (36.0%). Fewer teachers cited curriculum constraints (25.4%), limited resources (21.1%), lack of confidence (15.8%), parent expectations (13.2%), lack of support from school administration (8.8%), and insufficient training (3.5%).

Table 5. Challenges to Implementing Formative Assessment (Multiple Response, N = 114)

Challenge	N	%
Lack of time	81	71.1
Large class size	73	64
Student resistance	51	44.7
Examination pressure	46	40.4
Technological limitations	41	36
Curriculum constraints	29	25.4
Limited resources	24	21.1
Lack of confidence	18	15.8
Parent expectations	15	13.2
Lack of support from school administration	10	8.8
Insufficient training	4	3.5

6.6 Future Professional Development Needs

Regarding topics teachers wished to learn more about, language-related assessment techniques were the most requested (68.4%), followed by strategies for differentiating assessment for students of varying levels (60.5%), integrating formative with summative assessment (51.8%), and digital tools for FA (51.8%). About half of respondents wanted more guidance on using assessment data to inform instruction (50.9%) and on student self-assessment (50.0%). Only 5.3% reported no further learning needs.

Table 6. Topics Teachers Wish to Learn More About (Multiple Response, N = 114)

Topic	N	%
Language-focused assessment techniques	78	68.4
Differentiating assessment for diverse students	69	60.5
Digital tools for formative assessment	59	51.8
Integrating formative and summative assessment	59	51.8
Using assessment data to inform instruction	58	50.9
Student self-assessment	57	50
Balancing curriculum requirements with FA	48	42.1
Theoretical foundations of FA	42	36.8
Peer assessment	36	31.6
Providing effective feedback	31	27.2
None	6	5.3

6.7 Qualitative Findings from Open-Ended Responses

To complement the quantitative findings above, responses to the questionnaire's six open-ended items were analyzed using the content-analysis procedure described in Section 5.5. Findings are reported by item below; illustrative quotes have been translated from the original Vietnamese into English.

6.7.1 Suggestions for Improving the Training Program

Asked what they would change about the training program, the largest group of teachers (44%) offered no suggestion, typically because they considered the program's content already adequate (e.g., *"The program is good, so I have no further suggestions"*). Among teachers who did suggest changes, the most common request (11%) was for more hands-on practice time, echoing the quantitative finding that hands-on practice opportunities were among the most valued aspects of training (Table 4). A second cluster of suggestions (9%) concerned scheduling, with several teachers requesting that training be held during the summer break to avoid conflicts with teaching duties (e.g., *"Training should be held during summer so it doesn't overlap with teaching time"*). A smaller group (8%) asked that materials be provided in advance so they could prepare before sessions.

6.7.2 Changes in Understanding of Formative Assessment

When asked how their understanding of formative assessment had changed, the most common response (25%) described acquiring a broader repertoire of specific techniques (e.g., *"I learned more assessment methods"*), and a further 11% specifically reported a heightened recognition of formative assessment's importance or value. A small minority (4%) reported little or no change, generally because they already valued formative assessment before the training (e.g., *"Not much change, since I was already using formative assessment"*). Many responses (37%) framed the change in terms of benefits to students rather than to teachers' own knowledge (e.g., *"It helps me assess students better"*), suggesting that for many teachers the training's value was experienced through its downstream effect on learners rather than as an abstract shift in belief.

6.7.3 Examples of Formative Assessment's Impact on Students

Teachers most often illustrated formative assessment's impact through concrete feedback loops (36%), describing how giving students detailed, timely feedback helped them notice and correct errors (e.g., *"Giving detailed verbal feedback helps students improve their learning"*). A further quarter of responses (25%) described group or presentation work as the vehicle for visible student progress, and 17% specifically cited peer- or self-assessment activities. These examples corroborate the quantitative finding that feedback techniques and error-correction codes showed some of the largest measured gains in reported use (Table 3).

6.7.4 Influence of School Environment

Consistent with the quantitative barriers reported in Table 5, teachers' open-ended responses about their school environment most frequently referenced class size and workload: 15% explicitly described large classes as a constraint on implementation (e.g., *"An overly large class is an obstacle"*), and a further 11% cited time constraints more generally. At the same time, the single largest group of responses (31%) described the school environment in positive terms such as supportive leadership, adequate facilities, or a collaborative atmosphere (e.g., *"A supportive school environment, with cooperation between teachers and school leadership"*). About one in five teachers (21%) reported no meaningful environmental influence either way, and 11% cited limitations in facilities or technology.

6.7.5 Successful Examples of Technology Use

Reported technology use for formative assessment was dominated by quiz- and game-based platforms (35%), especially Kahoot, Quizizz, Azota, and Google Forms, typically used for quick vocabulary or comprehension checks with immediate feedback (e.g., *"I use Quizizz to check vocabulary after each lesson; students engage enthusiastically, and seeing results immediately lets me adjust my teaching promptly"*). Video or audio recording tools were the second most common category (16%), used mainly for speaking assessment. However, a substantial minority of teachers (17%) reported not having successfully used technology for formative assessment, citing limited digital skills, poor internet connectivity, or insufficient time—consistent with technological limitations being cited by 36% of teachers as a broader implementation barrier (Table 5).

6.7.6 Influence on Teaching Philosophy

The largest group of responses (34%) described a general positive influence on teaching without further elaboration (e.g., *"Positive influence"*). Among more elaborated responses, 22% described using formative assessment evidence to actively adjust their teaching methods and choice of activities, and 15% specifically framed the shift as moving toward more learner-centered, competency-based teaching (e.g., *"My experience with formative assessment helped me recognize the importance of continuous feedback in learning; as a result, I adjusted my teaching toward a more learner-centered approach"*). A small minority (7%) reported no

substantial change to their teaching philosophy, most often because they had already been using a communicative, learner-centered approach before the training.

Taken together, these open-ended responses reinforce and add texture to the quantitative results: they corroborate large class size and limited time as salient implementation barriers (Section 6.5), illustrate concretely how feedback- and error-correction-oriented techniques translate into visible changes in student behavior, and suggest that for many teachers, the training's primary perceived effect was less a change in abstract beliefs about formative assessment than an expansion of their practical technique repertoire and a shift toward evidence-based instructional adjustment.

7. Discussion

The findings provide strong quantitative evidence that the in-service training program organized by Can Tho University changed how surveyed EFL teachers reported using formative assessment in their classrooms. Every one of the 11 FA techniques examined showed a statistically significant and practically meaningful increase in reported frequency of use after training, echoing prior evidence that well-designed in-service training can shift classroom practice (Karagiorgi & Symeou, 2007; Ulla, 2018; Mugarura et al., 2022) and specifically supporting Ha et al.'s (2021) finding that in-service training in Vietnam can improve teachers' ability to design and use assessments.

The pattern of gains is informative. Techniques with the lowest pre-training use—exit tickets, observation checklists, and error-correction codes—showed the largest raw gains, suggesting the training was most transformative where prior practice was weakest, consistent with the aim of the National Foreign Languages Project reform to build FA capacity essentially from the ground up. Error analysis is a partial exception to this pattern: despite a comparatively high pre-training baseline ($M = 3.40$) and a moderate raw gain, it produced the largest standardized effect size of all 11 techniques ($g = .779$), suggesting the training also achieved unusually consistent gains across teachers for this technique specifically, independent of how novel the practice was beforehand. In contrast, classroom questioning, which teachers already used relatively often before training ($M = 3.54$, the highest baseline of all 11 techniques), showed the smallest absolute gain—plausibly a ceiling effect rather than evidence that the training was less effective for this technique.

Teachers' ratings of what made the training valuable—trainer expertise, materials, and practical examples ranking highest—are consistent with Desimone's (2009) emphasis on content focus and active learning as core features of effective professional development, and with Kennedy's (2016) argument that the quality of content and opportunities for practice matter more than structural format. These findings also align with Mohammadi and Moradi's (2017) observation that customized, well-delivered professional development is perceived as particularly beneficial, and with the training-quality factors (relevance, materials, trainer quality, duration, and follow-up support) identified as central to evaluating in-service training effectiveness (Androniceanu et al., 2020). At the same time, the challenges teachers reported—above all lack of time and large class sizes—echo a persistent tension in the literature between policy ambitions for formative, continuous assessment and the day-to-day working conditions of teachers in large, exam-oriented classrooms. This suggests that training alone, however effective at building knowledge and initial practice, may be insufficient to sustain FA use without complementary institutional support, such as protected planning time, smaller class sizes where feasible, and continued mentoring or coaching (Herrington et al., 2016).

Read through the lens of Clarke and Hollingsworth's (2002) Interconnected Model of Professional Growth, the training program (External Domain) appears to have produced a clear shift in the Domain of Practice—the significant increases in FA technique use reported in Table 3—more readily than it can be assumed to have produced a corresponding shift in the Personal Domain of underlying beliefs, which was measured separately in Part III of the questionnaire but is not the focus of the present quantitative analysis. This is consistent with Herrington et al. (2016) and Borg's (2003) caution that the beliefs–practice relationship is rarely linear: teachers may adopt new techniques procedurally in response to training and institutional expectation before their beliefs about the value of those techniques fully catch up—or, per Guskey (2002), belief change may instead follow from teachers observing the classroom consequences of their new practice, rather than preceding it. The open-ended responses analyzed in Section 6.7 offer partial support for this reading: many teachers described their shift less as a change in abstract beliefs about formative assessment than as an expanded technique repertoire and a growing habit of adjusting instruction based on evidence from students—a pattern consistent with practice change preceding, or at least running ahead of, fully articulated belief change. Direct comparison of the beliefs, subscale data (Cronbach's $\alpha = .898$) with these qualitative patterns in future analyses would allow this to be

tested more rigorously, since the IMPG predicts that sustained practice change is most durable when reflection loops back to reinforce the Personal Domain rather than remaining confined to the Domain of Practice.

8. Implications

For teacher educators and program designers, the results suggest that in-service FA training should continue to prioritize experienced, well-prepared trainers and concrete practical examples, since these were the aspects teachers valued most. The consistently large gains for previously underused techniques (exit tickets, observation checklists, error-correction codes) indicate that training content addressing genuinely novel practices, rather than only reinforcing techniques teachers already use, yields the greatest return.

For school leadership and policymakers, the prominence of time constraints and class size as barriers points to the need for structural support, such as dedicated planning time or reduced administrative burden, to accompany training if gains in practice are to be sustained beyond the training period.

For future training design, the topics teachers most wanted to learn more about (language-focused assessment techniques, differentiation, and integrating formative with summative assessment) offer a concrete agenda for follow-on professional development.

9. Limitations and Directions for Future Research

Several limitations should be acknowledged. First, this study relies on a retrospective self-report design in which teachers rated their frequency of FA technique use both before and after training simultaneously in a single questionnaire administration, rather than through a true pre-post design with independent measurement occasions; this approach is susceptible to recall bias and social desirability effects. Second, the sample, while reasonably sized ($N = 114$), was drawn from eight provinces and cities with uneven geographic representation (e.g., Ca Mau 32.5% versus Ben Tre 1.8%) and from teachers who had already completed the training organized by Can Tho University, which may limit generalizability to other regions of Vietnam. Third, the absence of classroom observational data means that self-reported practice change cannot be independently verified; as Pajares (1992) and Borg (2003) have cautioned, the relationship between reported beliefs and actual practice is rarely linear, so the improvements captured here should be interpreted as changes in reported, rather than directly observed, practice. Fourth, although Cronbach's alpha values indicate strong internal consistency for all six subscales (Table 2), no information is provided regarding pilot testing, factor analysis, or content validation procedures beyond internal consistency, which limits what can be concluded about the instrument's construct validity. Fifth, the design lacks a comparison group of teachers who did not receive the training; consequently, alternative explanations for the observed increases, such as general professional experience, concurrent policy pressure to adopt formative assessment under the 2018 General Education Curriculum, or simple maturation over the training period, cannot be ruled out, and the reported increases should not be interpreted as unambiguously caused by the training program alone.

This paper reports both the closed-ended, quantitative items of the questionnaire and a thematic content analysis of its six open-ended questions (Section 6.7), which together provide richer insight into why particular techniques were more readily adopted than others and how teachers described the training's influence on their practice. The qualitative analysis, however, was conducted by a single coder without an independent second coder or formal inter-rater reliability check (Section 5.5), and relies on English translations of original Vietnamese responses, which may not fully preserve nuance; a more rigorous qualitative analysis—ideally involving multiple coders and, where feasible, follow-up interviews rather than brief open-ended survey items—would strengthen confidence in the themes reported here. Future research should also incorporate classroom observation data to triangulate self-reported practice change and employ longitudinal designs to assess the durability of training effects. Additionally, direct comparison of the beliefs subscale data (Cronbach's $\alpha = .898$; Table 2) with the practice data reported in Table 3 would allow empirical testing of the IMPG's prediction that sustained practice change requires reinforcement through the Personal Domain.

10. Conclusion

This study examined the impact of an in-service formative assessment training program, organized by Can Tho University, on the practices of 114 K-12 EFL teachers in Vietnam. Teachers reported significant increases in the frequency of use of all 11 formative assessment techniques examined after completing the training, with moderate-to-large effect sizes across the board. Trainer expertise and practical, well-prepared materials emerged

as the most valued components of the program, while lack of time and large class sizes were the most commonly cited barriers to implementation, a pattern corroborated by teachers' own open-ended accounts of their school environment. Thematic analysis of the questionnaire's open-ended items further suggested that teachers experienced the training primarily as an expansion of their practical technique repertoire and a shift toward evidence-based instructional adjustment, rather than as an explicit change in their underlying beliefs about assessment. These findings are consistent with in-service training serving as an effective lever for advancing formative assessment practice under Vietnam's General Education Curriculum 2018, while underscoring that sustained implementation will likely require institutional support, particularly around time and class size, alongside continued training. Reading through Clarke and Hollingsworth's (2002) IMPG, the study highlights the importance of distinguishing between practice change and belief change as empirically distinct outcomes of professional development, and points to the need for future research that traces the reflective pathways through which training-induced practice change may—or may not—consolidate into durable shifts in teachers' underlying assessment beliefs.

11. References

- Androniceanu, A.-M., Kinnunen, J., Georgescu, I., & Androniceanu, A. (2020). A multidimensional approach to competitiveness, innovation, and well-being in the EU using canonical correlation analysis. *Journal of Competitiveness*, 12(4), 5–21. <https://doi.org/10.7441/joc.2020.04.01>
- Bakshi, P. M. (2020). 'It is a dialect, not a language!'—Investigating teachers' beliefs about Mewati. *LI Educational Studies in Language and Literature*, 20, Running Issue, 1–24. <https://doi.org/10.17239/liesll-2020.20.01.06>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36(2), 81–109. <https://doi.org/10.1017/S0261444803001903>
- Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, 18(8), 947–967. [http://doi.org/10.1016/S0742-051X\(02\)00053-7](http://doi.org/10.1016/S0742-051X(02)00053-7)
- Debreli, E. (2016). Pre-service teachers' belief change and practical knowledge development during the course of practicum. *Journal of Education and Training Studies*, 4(7), 37–46. <http://doi.org/10.11114/jets.v4i7.1513>
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199. <http://doi.org/10.3102/0013189X08331140>
- Fives, H., & Buehl, M. M. (2012). Spring cleaning for the “messy” construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook. Vol. 2: Individual differences and cultural and contextual factors* (pp. 471–499). Washington, DC: American Psychological Association. <https://doi.org/10.1037/13274-019>
- Grapin, S. E., Plumley, C. L., Banilower, E. R., Mahon, A. J. S., Craven, L., Malzahn, K. A., Pasley, J. D., Schwenger, A., Haas, A., & Lee, O. (2024). Development of a questionnaire on teachers' beliefs, preparedness, and instructional practices for teaching NGSS science with multilingual learners. *Science Education*, 109(1), 128–156. <https://doi.org/10.1002/scs.21905>
- Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, 8(3), 381–391. <http://doi.org/10.1080/135406002100000512>
- Hanna, F., Oostdam, R., Severiens, S., & Zijlstra, B. J. H. (2019). Domains of teacher identity: A review of quantitative measurement instruments. *Educational Research Review*, 27, 15–27. <https://doi.org/10.1016/j.edurev.2019.01.003>
- Ha, H. T. L., Pham, A. T. K., Nguyen, H. T., & Duong, H. T. T. (2021). Training pedagogical skills: Evaluation of lecturers and teacher training students at educational universities in Vietnam. *EURASIA Journal of Mathematics, Science and Technology Education*, 17(12), 20–54. <https://doi.org/10.29333/ejmste/11418>
- Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Corwin Press.
- Herrington, D. G., Yezierski, E. J., & Bancroft, S. F. (2016). Tool trouble: Challenges with using self-report data to evaluate long-term chemistry teacher professional development. *Journal of Research in Science Teaching*, 53(7), 1055–1081. <https://doi.org/10.1002/tea.21323>
- Jacobson, E., Francis, D. C., Willey, C., & Wilkins-Yel, K. G. (2022). Race, gender, and teacher equity beliefs: Construct validation of the attributions of mathematical excellence scale. *AERA Open*, 8. <https://doi.org/10.1177/23328584221130999>

- Jordan, A. (2018). The Supporting Effective Teaching Project: 2. The Measures. *Exceptionality Education International*, 28(3). <https://doi.org/10.5206/eei.v28i3.7770>
- Karagiorgi, Y., & Symeou, L. (2007). Teachers' in-service training needs in Cyprus. *European Journal of Teacher Education*, 30(2), 175–194. <https://doi.org/10.1080/02619760701275487>
- Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, 86(4), 945–980. <http://doi.org/10.3102/0034654315626800>
- Manuel, L. D., & Elder, A. C. (2025). Relationships among implicit theories of intelligence, attitudes toward inclusion, and teacher self-efficacy beliefs for inclusive practices. *Journal of Research in Special Educational Needs*, 26(1). <https://doi.org/10.1111/1471-3802.70048>
- Moghadam, M. S. (2015). The discrepancy between teachers' belief and practice, from the sociocultural perspective. *Studies in English Language Teaching*, 3(3), 252. <https://doi.org/10.22158/selt.v3n3p252>
- Mohammadi, M., & Moradi, K. (2017). Exploring change in EFL teachers' perceptions of professional development. *Journal of Teacher Education for Sustainability*, 19(1), 22–42. <http://doi.org/10.1515/jtes-2017-0002>
- Mugarura, P., Ssempala, F., & Nachuha, S. (2022). Role of in-service teacher training as a tool for the student's performance in selected public secondary schools in Kisoro District. *International Journal of Educational Policy Research and Review*, 9 (1), 1–13. <https://doi.org/10.15739/IJEPRR.22.001>
- Nguyen, V. B. H., Vu, T. M. H., Hoang, T. K. H., & Nguyen, T. M. N. (2020). Vietnamese education system and teacher training: Focusing on science education. *Asia-Pacific Science Education*, 6(1), 179–206. <https://doi.org/10.1163/23641177-BJA10001>
- Öztürk, M. (2019). An evaluation of an innovative in-service teacher training model in Turkey. *International Journal of Higher Education*, 8(1), 23–36. <https://doi.org/10.5430/ijhe.v8n1p23>
- Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2022). Teacher professional development around the world: The gap between evidence and practice. *The World Bank Research Observer*, 37(1), 107-136. <http://doi.org/10.1093/wbro/lkab006>
- Ulla, M. B. (2018). In-service teachers' training: The case of university teachers in Yangon, Myanmar. *Australian Journal of Teacher Education (Online)*, 43(1), 66-77. <https://doi.org/10.14221/ajte.2018v43n1.4>
- Valcke, M., Sang, G., Rots, I., & Hermans, R. (2010). Taking prospective teachers' beliefs into account in teacher education. 622–628. <https://doi.org/10.1016/b978-0-08-044894-7.00668-0>
- Wijesundera, S., & Wijethunga, K. (2021). The association between teachers' mathematical beliefs, teacher characteristics, and student achievements at the junior secondary level of education in Sri Lanka. *Sri Lanka Journal of Social Sciences*, 44(2), 237. <https://doi.org/10.4038/sljss.v44i2.8260>