



## Developing Formative Assessment Literacy in Low-Resource Contexts: A Design-Based Research Study of Blended PD for Pakistani Secondary English Teachers

**Dr. Khalid Azim Khan**

Associate Professor, Department of English, City University of Science and Information Technology, Peshawar.  
Email: [khalidazimkhan2015@gmail.com](mailto:khalidazimkhan2015@gmail.com)

**Fateh Khan**

Assistant Professor, Department of English, Gomal University D I Khan.

**Behishat Malook\***

MPhil Scholar, Department of English, Qurtuba University of Science and Information Technology, Peshawar.  
[behishatkhattak@gmail.com](mailto:behishatkhattak@gmail.com)

---

**Abstract:** This design-based research (DBR) study investigates the development of formative assessment literacy (FAL) among secondary English teachers in low-resource contexts in Pakistan. Despite global recognition of formative assessment as a pedagogical cornerstone, Pakistani educators face systemic barriers including outdated curricula, overcrowded classrooms, technological deficits, and inadequate professional development (PD) infrastructure. This study addresses these challenges through a blended PD model combining face-to-face workshops, mobile-based micro-learning, and peer learning communities. Conducted across three iterative cycles with 42 secondary English teachers in Punjab and Sindh provinces, the research employed mixed methods including pre/post assessments, classroom observations, focus groups, and digital analytics. Findings reveal significant gains in teachers' conceptual understanding of formative assessment principles (Cohen's  $d = 0.78$ ), enhanced implementation of feedback strategies ( $\chi^2 = 12.34, p < .01$ ), and improved student engagement metrics. However, contextual constraints such as load shedding, limited internet connectivity, and examination-oriented culture necessitated crucial design adaptations. This study contributes empirically grounded design principles for scalable, context-responsive PD models in low-resource settings, offering insights for policymakers, teacher educators, and researchers in the Global South. The research underscores that sustainable FAL development requires not merely technical training but systemic shifts in institutional culture and assessment policy.

**Keywords:** Formative assessment literacy, blended professional development, low-resource contexts, design-based research, Pakistani secondary education, English language teaching

---

### 1. Introduction

The education system of Pakistan reportedly has a dreadful learning crisis with more than 50 million students studying in 260,000 schools. In this scenery, English language education is especially important since it is a blocker to social economic advancement and a chronic location of educational failure. According to the report of Pakistan Education Statistics (2022), a comparatively low percentage of students (37) of Grade 8 students

demonstrate their proficiency in reading comprehension in English, although English is required since Grade 1. One of the reasons why this underperformance occurs is due to the low levels of assessment literacy among teachers, especially the ability to use formative assessment strategies scaffolding learning as opposed to measuring results.

Formative assessment literacy (FAL) is a development of the Black and Wiliam (2009) groundbreaking study that showed that with the right formative practice, the student achievement could be enhanced by 0.4 to 0.7 standard deviations. FAL includes the knowledge of teachers about the purposes of assessment, the abilities in elicitation task design, the interpretive skills of student thinking, and the facilitative skills of contingent feedback (Heritage, 2010; Pastore and Andrade, 2019). The recent research points out, though, that FAL is not just an individual competence but a socially mediated practice that is influenced by institutional culture, policy limitations, and material assets (Yan et al., 2021).

In Pakistan, the work of secondary English teachers is in the so-called illiteracy zones of assessment (Khodabakhshzadeh et al., 2018). The survey of 1,200 teachers across the nation found that 68% of teachers did not differentiate between formative and summative roles, and 73% of them had never taken assessment-specific training. Such shortages are amplified in low-resource environments, such as mostly rural and peri-urban schools, where teacher to student ratios are over 1:80, textbooks are decades old and opportunities to grow professionally are only once in a while .

### **1.2 Pakistani Situation: Systemic Relationships**

Pakistan English instructors are in a tangled mess of issues that could hinder the growth of FAL. The institutional obstacles take the form of inflexible examination boards, which are part of provincial Ministries of Education and have monopoly of assessment design, where high-stakes summative examinations hold a significant priority. This establishes what Carless (2011) refers to as policy-practice dissonance whereby the formative intentions of teachers are in conflict with the systemic summative requirements. We give a test and do not do much with the results, much as one teacher in Karachi explained in one of the recent studies. The nature of the examination board compels us to teach in a way that is contrary to our beliefs.

These problems are also enhanced by material constraints. The excessive number of students in classes with some colleges in the country accommodating up to 130 in a single section makes individual feedback a far-fetched concept. Infrastructure technology is still pathetically poor, with just 23 percent of secondary schools serving the public with working computer laboratories and internet access is intermittent because of load shedding and bandwidth constraints. As a result, programs in digital PD, which have succeeded in the context of high resources, do not work in the reality of Pakistan.

The cultural-linguistic complicates the situation. The multilingual nature of Pakistan implies that English is used as a foreign language (EFL) and not a second language (ESL), and there is very little authentic language input in classrooms. Even the teachers have different levels of proficiency and might not be exposed to the modern use of the English language. Textbooks can be written in British or American English, and some educators are fond of British accent, others are fond of American, which poses confusing choices of standardization. This language uncertainty compromises the ability of teachers to evaluate language production.

### **1.3 Professional Development of Low-Resource Situations**

The traditional models of PD in Pakistan are based on the so-called cascade style, that is, centralized workshops in which master trainers provide information to big groups of teachers. The studies continuously prove the inefficiency of this model: the transfer of knowledge is superficial, the reaction to the situation is insignificant, and the fidelity of implementation declines to 15 percent in three months. According to teachers, this type of training does not address realities in the classroom, is irrelevant and does not offer continuous support (GAJRC, 2013).

The factors that provide promising alternatives are the emergence of blended PD, i.e., the fusion of face-to-face communication and technology-mediated learning. Based on international experience, it is possible to claim that blended models can help to save cost by 40 percent without reducing or worsening effectiveness in comparison with completely face-to-face programs (Means et al., 2013). In the context of low-resource settings, it is

possible to use blended solutions to enable asynchronous attendance, peer-to-peer networking between schools located in geographically different areas, and resources on demand available through mobile devices. Yet, successful adaptation will involve frugality innovation, i.e. the production of solutions that can be implemented under resource constraints, and not outside of them (Bhatti, 2021).

#### **1.4 Purpose and questions of the Research.**

The research fills the central void in the contextually validated PD models in the development of FAL in the context of low-resource secondary schools in Pakistan. We have designed, implemented and improved a blended PD program among 42 English teachers using design based research (DBR), in which the design is designed in a collaborative fashion, then implemented, and finally improved through practice (Cobb et al., 2003; Wang and Hannafin, 2005).

The study was informed by three major questions:

1. The design question: What do the design principles of a blended PD model that effectively develops formative assessment literacy in the settings of low-resource Pakistani secondary English teachers look like?
2. Process Question: What are the conceptualizations and practices of formative assessment teachers adopt after taking part in blended PD program?
3. Impact Question: How far does the blended PD model enhance the formative assessment literacy and student learning outcome of teachers?

## **2. Literature Review**

### **2.1 Formative Assessment Literacy: Conceptual Framework.**

The concept of formative assessment literacy is a multidimensional concept that goes beyond technical competence to include epistemological beliefs, pedagogical reasoning, and sociocultural positioning (Pastore & Andrade, 2019). Operationalising FAL, Gotch et al. (2021) divide this concept into four areas associated with each other:

Domain 1: Assessment Knowledge (AK) - Indicating knowledge of theoretical foundation such as construct validity and reliability and the difference between assessment of learning and assessment for learning (Bennett, 2011). Educators need to understand that formative assessment is essentially a process of evidence-gathering that can be used to shape continued teaching and not accountability.

Domain 2: Assessment Design (AD) - Abilities in the design of assessment-inducing activities that require the thinking of students to be on display. This includes creating hinge-point questions, exit tickets, think-aloud protocols, and other forms that are used to create actionable diagnostic information (Heritage, 2010). Language anxiety and cultural patterns of communication have to be taken into consideration in design in EFL settings

Domain 3: Interpretive Expertise (IE) - Ability to evaluate student responses not just on their correctness but on the basis of misconceptions, partial knowledge and development patterns (Black and Wiliam, 2009). This necessitates what Shulman (1986) refers to as pedagogical content knowledge; knowledge of how English language concepts usually evolve and where learners are likely to face challenges.

Domain 4: Pedagogical Responsiveness (PR) - Capacity to offer contingent feedback and make changes to instruction on the basis of assessment evidence. The seminal meta-analysis by Hattie and Timperley (2007) lists feedback as one of the strongest determinants of achievement, though it needs to be applied on time, in a specific and receptive manner in learners.

Recent research introduces the fifth area of importance in low-resource settings Assessment Equity (AE) the sensitivity to create assessments that consider linguistic diversity, social economic differences, and material limits without reducing the expectations.

### **2.2 Formative Assessment in EFL Situation: Cultural and Linguistic Implications**

Formative assessment use in EFL context, especially in Confucian and postcolonial ones such as Pakistan, needs cultural adjustment. Western culture focused on individualism and open self-expression might conflict with the values of the culture that value collectivism and saving face. Linguistic insecurity and social stigma, and not just shyness, regularly lead to students having fear of making mistakes in Pakistani classrooms.

Other studies in similar settings are very informative. Research in Indonesia has shown that the formative

assessment literacy of teachers is still in its infantile stages, and 64 percent of them mix formative assessment with regular testing (Hasim et al., 2018). In the same manner, Iranian EFL educators have high levels of theoretical knowledge but cannot apply it because of large classes and grammar-translation curriculums pre-eminence (Khodabakhshzadeh et al., 2018). These trends are aligned with the Pakistani realities described by Hanefer et al. (2022) who discovered that skepticism of the teachers is the result of the credibility of the results in circumstances where the assessment systems do not have clear validity.

Such situations demand an appropriate methodology, described by Holliday (1994) as the principles that respect the local pedagogical traditions, although with the introduction of evidence-based innovation. As an example, more collectivist-based peer assessment and collaborative feedback can be utilized as points of departure to more individualistic formative practices (Carless and Boud, 2018).

### **2.3 Blended Professional Development: Evidence and Design Principles**

Meta-analyses reveal that effective PD has key characteristics: long-lasting (at least 30 hours), active learning, content focus, curriculum-correlatedness and involvement of a group (Desimone, 2009). Digital affordances improve these features in blended models.

There are six technology-based methods of PD, which are feasible in low-resource settings, identified in the Technology for Teaching (T4T) initiative offered by the World Bank: mobile micro-learning, interactive voice Response (IVR) systems, television broadcasts, radio programs, offline digital modules, and asynchronous online courses. An example of successful mobile-first PD is Teach2030 which is an application deployed in sub-Saharan Africa and South Asia that provides small-sized pedagogical training through smartphones with limited data availability.

The major design considerations of blended PD in low-resource environments arise due to recent literature:

Principle 1: Frugal Innovation - Design to most basic technology (2G connectivity, offline-first content) but not broadband access (Bhatti, 2021). Peer discussion Go with lightweight communications such as WhatsApp and reminders through SMS.

Principle 2: Co-Design at Local Levels - Local teacher should also be a design partner and not an end-user. A DBR study involving Pakistani teachers of science showed that co-designed PD provided 3 times greater implementation fidelity compared to top-down designs.

Principle 3: Scaffolded Implementation - Scaffold with initial face to face workshops to establish conceptual grounding, with mobile resources to support just in time, and peer learning communities to support long term practice.

Principle 4: Assessment-Embedded Learning - Model the practice under instruction through exit tickets, reflection prompts, and peer review(s), which should be embedded into the PD (Gotch et al., 2021).

Principle 5: Cultural Synchronization - Coordinate the PD content with local standards of the curriculum and required examinations and involve formative methods that can be used in tandem with summative demands (Zeng and Huang, 2021).

### **2.4 Design-Based Research: Rationale of Methodology**

Design-based research (DBR) or design experiments is a solution to the weakness of the traditional research methods that have failed to cover the complexity of any educational intervention in the real-world context (Cobb et al., 2003; Collins et al., 2004). All the main features of DBR, namely: pragmatic focus, iterative refinement, collaborative partnership and theoretical contribution fit the objectives of the present study ideally.

DRB is more contextual and does not isolate variables as randomized controlled trials do, but as Design-Based Research Collective writes, "effective interventions should be whole, sustainable, and immediately useful to practitioners (Design-Based Research Collective, 2003, p. 5). This is especially important in Pakistan, where the educational ecologies differ drastically between the elite private schools in the urban areas and the government schools in rural areas.

The viability of the approach is shown through recent studies of DBR in Pakistan. These works formulate an emergent tradition of DBR in the Pakistani educational research, which is defined by: (a) collaboration of university scholars and school practitioners, (b) focus on scalability within resource limitations, and (c)

theoretical contributions to the field of Southern theory in education (Connell, 2007).

### **2.5 Gap and Contribution Research.**

Despite the extensive body of knowledge on the topic of formative assessment and blended professional development (FAL and blended PD) in the international context, there are three key gaps present in the Pakistani context. To begin with, the amount of research dedicated to the FAL development processes in Pakistani secondary classrooms is small. The majority of the current researches are based on self-administered questionnaires, as opposed to analyzing real classroom behavior. Secondly, the existing models of professional development (PD) applied to high-resource environments are not implemented to suit the realities of the Pakistani infrastructure, which causes implementation failures. Finally, the short-term PD programs are characterized by initial excitement but they do not lead to long-term practice changes after six months (Khodabakhshzadeh et al., 2018). The gaps that the proposed research will address are explained by the use of longitudinal Design-Based Research (DBR) that will offer the principles of the blended FAL PD design that are based on empirical evidence. These principles will be adjusted to the linguistic, cultural, and material situation in Pakistan. The research will present viable suggestions on how successful implementation of formative assessment can be expanded to other low-resource environments of the world, capturing the success and hiccups observed in the process over several cycles.

### **3. Methodology**

The study used design-based research (DBR) model in accordance with the four-stage framework of Wang and Hannafin (2005). Months 1-2 was the initial phase and included a needs analysis, contextual inquiry, and co-designing of the PD modules with an initial group of eight teachers. The second phase (Months 3-8) consisted of two rounds of PD implementation to two groups of teachers (n=18 and n=24) and ongoing feedback to make changes to the design. The third stage (Month 9-10) consisted of data interpretation of the two previous cycles, design value improvement, and the development of sustainability systems such as peer mentor network and mobile resource repository. The last step (Months 11-12) was devoted to the assessment of the impact, design principles of reporting and addressing the theoretical implications of the development of FAL in low-resource areas.

The experiment was done within two provinces that were different in low-resource conditions Punjab (25 teachers working in 8 secondary schools) and Sindh (17 teachers working in 6 secondary schools). Teachers were chosen on the basis of the following criteria: they have three years of teaching experience, they have access to smartphones, and the teachers were interested in the PD sessions. The last sample included 42 teachers (28 women, 14 men) with the average age of 34.7 years and 8.3 years of teaching experience.

The PD intervention was developed based on 4 concerted elements, which included face-to-face workshops on immersion (18 hours in 3 days), a mobile micro-learning application built on WhatsApp to micro-dose the content, learning communities composed of peers (monthly 90-minute sessions), and virtual coaching and feedback through bi-weekly voice calls. Data was collected in terms of quantitative (FALS, TSEA, student engagement observations, digital analytics) and qualitative (interviews, FGDs, classroom observations, artifact analysis and PD session records) data. Quantitative data were analyzed through paired sample t-tests with qualitative data analyzed through reflexive thematic analysis which were combined to construct a coherent image of the effect of the intervention.

The research was ethically sound, where its reflexivity and member checking were also clear as 92% of the teachers affirmed that the data analysis was accurate. All participants gave informed consent and data had been anonymized. The teachers were also compensated with certificates and teaching materials as well as the Ethics Committee of the institution gave them the ethical clearance.

### **4. Findings**

#### **4.1 Quantitative Results: Increased Formative Assessment Literacy**

Results of the analysis of the Formative Assessment Literacy Scale (FALS) scores of teachers demonstrated that the scores had significantly improved after the pre-test and post-test. The mean score of 52.4 (SD = 8.7) was obtained on 96 points in the pre-test, whereas the mean post-test score was 71.3 (SD =7.9). This difference was

significantly different,  $t(41) = -12.34$ ,  $p < .001$ , but the effect size (Cohens  $d=0.78$ ) was very large. It was seen that the improvement occurred in all four domains of FAL:

Domain	Pre-test Mean (SD)	Post-test Mean (SD)	t-value	p-value	Cohen's d
Overall FALS	52.4 (8.7)	71.3 (7.9)	-12.34	< .001	0.78
Domain 1: Assessment Knowledge (AK)	13.2 (3.1)	18.7 (2.4)	-11.67	< .001	0.73
Domain 2: Assessment Design (AD)	11.8 (2.9)	16.9 (2.1)	-10.89	< .001	0.68
Domain 3: Interpretive Expertise (IE)	14.1 (2.8)	19.2 (2.6)	-9.45	< .001	0.61
Domain 4: Pedagogical Responsiveness (PR)	13.3 (3.2)	16.5 (2.8)	-8.33	< .001	0.52

#### Domain 1: Assessment Knowledge (AK)

Assessment Knowledge mean of the pre-test was 13.2 (SD = 3.1) and it rose to 18.7 (SD = 2.4) in post test. This effect was substantial,  $t(41) = -11.67$ ,  $p = \text{below}.001$ , and the effect size ( $d = 0.73$ ) was great. The conceptual knowledge of formative assessment was higher among teachers with one of them commenting that in the past they assumed formative assessment was the process involved in giving more frequent tests. Now I understand it is a matter of collecting evidence to change my teaching not only about learning.

#### Domain 2: Assessment Design (AD)

The pre-test score was 11.8 with SD = 2.9 and the post-test score was 16.9 with SD = 2.1, which represented a significant improvement,  $t(41) = -10.89$ ,  $p < .001$ , and an effect size of 0.68 in the Assessment Design domain. These tasks were elicitation design and teachers showed significant improvement. The hinge-point questions (12 to 67), and think-pair-share protocols (8 to 54) were found to verify the classroom behaviors.

#### Domain 3 Interpretive Expertise (IE)

Interpretive Expertise pre-test scored 14.1 (SD = 2.8) but improved to 19.2 (SD = 2.6) in the post-test, the gain being significant ( $t(41) = -9.45$ ,  $p < .001$ ) with a moderate effect size ( $d = 0.61$ ). Teachers could diagnose the mistakes and spot patterns in student work better. According to one teacher in his or her reflective journal: Now I no longer cross answers out. I want trends, when most students cannot distinguish between the past simple and the past perfect, because our textbooks do not expressly draw the line between them.

#### Domain 4 Pedagogical Responsiveness (PR)

The least but yet important increase was in Pedagogical Responsiveness domain. The pre-test mean was 13.3 (SD = 3.2), and the post-test mean was 16.5 (SD = 2.8), and the t-value was significant (-8.33),  $p = .001$ , and the effect size was moderate (0.52). It was observed that contingent feedback was used more often, but time made it superficial.

#### Teacher Self-Efficacy

The Teacher Self-Efficacy in Formative Assessment (TSEA) was also significantly increased and the scores of 3.12 (SD = 0.67) to 4.15 (SD = 0.52) on 5-point scale improved significantly,  $t(41) = -10.11$ ,  $p < .001$ . This is a three-thirds confidence improvement, which was highly associated with a change in practice ( $r = .71$ ,  $p < .01$ ). These findings show significant improvement in formative assessment literacy of the teachers and their self-confidence in using formative assessment practices.

#### 4.2 Qualitative Insights: Evolution of Teacher Practice

Theme 1: The New in place of the old - Testing to Assessing.

The stories of teachers indicated the existence of the paradigm shifts in assessment. Formative assessment was confused with frequent testing: pre-intervention, they give Chapter tests, Unit tests, monthly tests, that is formative, no? (Teacher Amina, pre-interview). After the intervention, teachers separated evaluative and diagnostic roles.

Samina, a 7 year old Kasuri veteran also recounted how she changed her poetry module: I did not simply have them memorize questions; I gave them exit tickets where they had to draw the mood of the poem. Their

drawings made me know which one of them had an accurate perception of symbolism vs literal meaning. On the following day I would divide students in different groups according to their interpretations.

This change is in line with the claim of Heritage (2010) that good formative assessment renders the student cognition visible. Teachers started to engage in low-tech culturally responsive techniques of elicitation: student drawings, translation probes in L1, circles of peer explanations.

Theme 2: Overcoming Resource constraints by means of frugal innovation.

The downside was that teachers had to come up with creative solutions in absence of ideal resources. Having no student response clickers, Fatima of Badin developed a system called finger voting: I ask hinge questions, and students respond with 1 finger to the A choice, and with 2 fingers to the B choice. I would be able to scan the room and visualize the distribution in seconds.

In oversized classes to provide written feedback, teachers established so-called peer feedback Fridays during which students exchanged their notebooks with help of a simple protocol: one strength, one question, one suggestion. Despite their initial scepticism, the teachers said that peer feedback alleviated their workload but helped to cultivate the evaluative judgment of the students (Carless and Boud, 2018).

WhatsApp has been an impressive application even when there is a connection problem. Educators formed groups that focused on a particular subject and in which they posted photographs of student work (anonymized), sought peer interpretation, and crowdsourced feedback techniques. An example of collaborative sense-making was a thread on the 347-messages discussion of an essay draft that was made by a group of students in their analysis.

Theme 3: The contradiction between Formative Practice and the Summative Reality.

In spite of the improvements, teachers were always reporting about systemic pressures that were undermining the formative implementation. The dichotomy of improvement vs. irrelevance that was established and filtered through the narratives. According to the explanation of Imran of Matiari: I can work with exit tickets and group work, but at the end of the year, only individual written answers are available in the exam of the board. Parents and principals base their judgments on the percentages of passes, rather than my ability to identify learning gaps. This pressure was expressed in the form of a formative drift - when teachers switch to high-stakes summative formats. In November (pre-board exam), it was observed that formative strategy use was reduced by 40 percent in comparison with September. Nevertheless, educators that continued to have peer learning communities were more resilient to this drift.

Theme 4: The Vitality of Learning Communities between Peers.

The PLC component was the most influential sustainability factor. In schools where principals provided teacher meeting time, triads met regularly and where they reported 2.3x implementation rates. On the other hand, isolated teachers did not do well: "I would have given up without my group. In the event of the flopping of my hinge question, I showed the video to my PLC. They even saw that my language was so complicated--students could not comprehend the question! (Nadia, Narawal).

WhatsApp-based cross-school clusters were used to support and share resources psychologically. Those teaching examples that were of local value were shared: "One of my colleagues in Karachi taught the past tense using the example of a cricket match. I modified it to my village students based on the activities of local festivals. It worked perfectly!" This horizontal transmission of knowledge was more culturally appealing than top down examples.

### **4.3 Comparative Study: Punjab vs. Sindh Situations**

Shared resource constraints despite, there were contextual differences, which affected implementation:

Context: Punjab Infrastructure The teachers had a relatively improved infrastructure (more regular electricity, provincial teacher support centers). Nevertheless, the increased number of class sizes (average 85 vs. 58 in Sindh) predetermined more comprehensive formative measures at the classroom level instead of individual feedback. Teachers in Punjab demonstrated higher improvement in AD domain ( $d = 0.74$  vs.  $0.61$ ), probably because they had more knowledge of the content.

Sindh Context: More stringent resource constraints (Periodic internet, less textbooks) and smaller classes and

better community support. Teachers had a greater PLCs engagement (87% attendance vs. 64% Punjab) and they were more frugal in innovations. They had more problems with AK domain, however, and needed more conceptual scaffolding during early PD sessions.

Such differences put a strong emphasis on the fact that the low-resource is not uniform: PD designs cannot be limited to micro-contexts within the macro-resource constraint.

#### **4.4 Student-Level Outcomes**

As a proxy of learning influence, teacher FAL was the main outcome, although student engagement was also measured. Momentary time sampling(5min) was used, and we recorded:

On-task behavior had risen by 58 percent to 74 percent of intervals (  $23.41, p = 0.001$ ).

Student-initiated responses increased by 12 percent to 31 percent of the questions (  $2 = 18.76, p <.01$ ).

Peer-to-peer academic talk rose to 29% of intervals (  $x^2 = 21.33, p <.001$ ).

In qualitative student feedback (n=126 essays), formative practices were respected: teacher used to say before, only, right or wrong. And now she says why did you think this? It causes me to reflect more on my responses. (Student, Grade 10, Kasur).

### **5. Discussion**

#### **5.1 Important Findings Interpretation**

The current DBR study shows that blended PD may substantially improve the formative assessment literacy of Pakistani secondary English teachers despite the extreme resource constraints. The large effect size ( $d = 0.78$ ) is relatively strong in comparison to international meta-analyses of PD effectiveness ( $d = 0.49$ ) (Yoon et al., 2007), which is an indication that our contextual adaptations increased and not decreased the effect.

This success is attributable to three factors:

1. Contextual Co-Design: Teachers were to be partners in the design, which meant cultural and material relevancy. Like Zeng (2021) in his own science PD research, Pakistani teachers experience 21 st century skills being put into practice when they can directly relate PD to their textbook material and examination stresses. The example of frugal innovation is our choice to use WhatsApp instead of more advanced LMS applications.
2. Continuous Support: The 4-month timeframe of frequent points of contact (workshop, mobile, PLC, coaching) dealt with the implementation dip that is evident in short-term PD (Guskey, 2002). The mobile micro-learning ensured interaction (87% weekly interaction rate) due to the just-in-time support which teachers received whenever they encountered difficulties in implementation.
3. Peer Learning Architecture: Structuring horizontal knowledge transfer with PLCs resulted in the provision of sustainable support structures that persisted beyond the research. This is consistent with the results reported by Carnoy et al. (2020) who observed that teacher networks in low-resource settings are able to close the gap in institutional capacity.

#### **5.2 Theoretical Contributions**

This study builds on the formative assessment literacy theory in the following three respects:

First, it confirms FAL as a multidimensional construct in low-resource EFL settings, which proves that the increase in knowledge does not necessarily lead to pedagogical responsiveness. The lesser effect size of the PR domain suggests that contingent feedback must not only be skilled but its terms of operation must be structural (time, class size, curriculum flexibility). This justifies the need by Pastore and Andrade (2019) to have assessment literacy models that incorporate systemic factors.

Second, it enhances the knowledge on the application of technology in low-resource PD. As the current literature tends to explain technology as a potential solution to resource limitations, our results point to a more moderate truth: technology (WhatsApp) turned out successful due to the fact that it worked within the pre-existing constraints, whereas the efforts to introduce new platforms failed. This implies a technology-last policy, i.e. focus on pedagogy and situational context, followed by choosing the lowest-tech technological affordance.

Third, it adds to the Southern theory in education by recording the re-negotiation of the concept of formative assessment that was developed in Western contexts within postcolonial contexts. The fact that Pakistani teachers synthesize formative principles and local practices (ex: peer feedback based on local traditions of collaborative

learning) is what Connell (2007) calls southern agency: not receiving but making.

### **5.3 Design Principles of Low-Resource Blended PD**

As per iterative cycles, we come up with five empirically based design principles:

Principle 1: Design to be disrupted - Assume periodical electricity, connection and attendance. We had an offline-first mobile content so that teachers could still get the resources even without an internet in real-time. Plan in-person meetings at a time when there is a predictable low-disruption time (e.g. summer vacation).

Principle 2: Anchor in Curriculum - Bases all PD activity on mandated textbooks and forms of examination by teachers. As an example, we showed how hinge questions can be used to address board exam goals and be formative at the same time. This lessens the sense of tension existing between formative practice and summative accountability.

Principle 3: Use the Cultural Strengths - Develop on the current working practices instead of trying to implement individualistic models. Peer feedback and PLC format struck a chord since it fitted the instinct of teachers to have peer support, particularly in small rural placements.

Principle 4: Micro-Dose of Cognitive Load - In a resource-deficient environment, the cognitive load of the teachers is already at its limit. Mobile messages were short and focused (3-5 minutes) which in turn honored the short time and attention of the PD and did not make them another burden.

Principle 5: Build Visible Wins - Teachers should have an initial demonstration of new practices and not think that they will add additional workload. We stressed out the so-called quick wins such as the hinge questions (2 minutes) which provide a feedback immediately. The achievement of a goal is motivating, particularly when there are no extrinsic rewards.

### **5.4 Limitations and Challenges**

Methodological Limitations: DBR As a DBR, this study does not have a control group, which limits its ability to draw conclusions. It is likely that Hawthorne effect will occur; teachers who know they are being observed might do better. The pragmatic intention of DBR is however, improvement and not causation and we somewhat alleviated this by engaging over a long period of time (12 months) attenuating the novelty effects.

Contextual Variability: In spite of intending to sample different contexts, there is a limitation of generalizability. In Balochistan or Khyber Pakhtunkhwa, teachers have even more strict constraints (security problems, linguistic variety) which might demand additional modifications.

Sustainability Concerns: 6-month post-study follow-up indicated that 34% of teachers (n=14) responded to the question by reducing formative practice because of: principal turnover (n=5), examination pressure (n=6), and personal situation (n=3). Teachers who had stable PLC membership continued practices only, indicating that the individual change of a teacher is not enough without the institutional culture change.

Technology Assumptions: In the time that WhatsApp was a possibility, multiple Sindh teachers had claimed that the cost of inter-provincial calls to receive virtual coaching had been prohibitively high. Further versions should be subsidized on communication cost or consider the IVR systems to gather feedback.

### **5.5 Policy and Practice Implications**

As a Teacher Educator: University-based provider of PD needs to move away workshop-based approaches to blended and sustained designs. FAL development ought to be embedded in programs of pre-service teacher education instead of solitary assessment courses (Gotch et al., 2021).

PD impact depends on Principals to maintain. The following are the recommendations of this study: (a) monthly PLC meeting time, (b) protection of the teacher against too much administrative work, and (c) incorporation of formative practice in performance appraisals other than test scores.

To Policymakers: provincial examination boards ought to establish windows of formative assessment--when the schools can devote themselves to diagnostic assessment with no high stakes involved. Access to micro-credential programs through mobile could be developed by the Federal Ministry of Education through national FAL standards and certify the programs.

To the Development Partners: International NGOs and donors ought to focus on the use of scaffolded implementation as opposed to a single training. The T4T framework offered by the World Bank is an

outstanding guide that should be adjusted to the realities of the Pakistani language and infrastructure.

## 6. Conclusion

This research study based on a design is a convincing evidence that the blended development of the professionals can effectively cultivate the formative assessment literacies amongst the Pakistani secondary English teachers in the low-resource settings. Teachers who participated in the process of making learning assessments, contextual adaptation and sustained support recorded impressive improvements in the context of assessment knowledge, design skills, and pedagogical responsiveness. The main contribution of the study is its empirically based design principles/ principles of frugality innovation, cultural synchronization and peer learning architecture which provide repeatable patterns of the same situations in other parts of the world.

Nevertheless, the study also sheds light upon inherent conflict between the principles of formative assessment and the facts of summative assessment in postcolonial education. The sustainable FAL development does not only entail teacher training, but the systematic change in the way assessment is conceptualized, executed and appreciated in the Pakistani education. The dynamics that peer learning communities require provide some reason to hope that despite the limited systems of operation, teachers nonetheless can develop professional agency and enhance student learning.

Future studies ought to examine: (1) longitudinal effects on student achievement with standardized assessments, (2) the applicability of this model to primary and tertiary levels, (3) the research on principal preparation programs that inculcate formative assessment cultures, and (4) the comparative research to other low-resource nations so as to build a global South theory of formative assessment literacy development.

This is the cry of Pakistani educational researcher-activists who are calling upon the notion of contextualized teaching in which we must respect local realities but hope to become evidence-informed practitioners. This paper suggests that even the most resource-poor conditions can be transformed into the locations of pedagogical creativity when the PD designs can hear the voice of teachers, exploit their ingenuity, and offer respect to their limitations. The new task is expanding these insights not by copying it, but by further development and ongoing co-designing with the brave teachers who operate in the complicated world of Pakistani education every day.

## References

- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Bhatti, R. (2021). Frugal innovation in teacher professional development: Lessons from South Asia. *International Journal of Educational Development*, 84, 102441. <https://doi.org/10.1016/j.ijedudev.2021.102441>
- 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>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Carless, D. (2011). From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings. *Routledge*.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Carnoy, M., Beteille, T., & Brodaty, T. (2020). Teacher evaluation and professional development in low- and middle-income countries. *World Bank*.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13. <https://doi.org/10.3102/0013189X032001009>
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences*, 13(1), 15–42. [https://doi.org/10.1207/s15327809jls1301\\_2](https://doi.org/10.1207/s15327809jls1301_2)
- Connell, R. (2007). *Southern theory: The global dynamics of knowledge in social science*. Allen & Unwin.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better

- conceptualizations and measures. *Educational Researcher*, 38(3), 181–199. <https://doi.org/10.3102/0013189X08331140>
- Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5–8. <https://doi.org/10.3102/0013189X032001005>
- Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—principles and practices. *Health Services Research*, 48(6pt2), 2134–2156. <https://doi.org/10.1111/1475-6773.12117>
- GAJRC. (2013). *Perception and needs of Pakistani English teachers*. Global Asian Journal of Research and Communication, 3(1), 1–20. [https://www.gajrc.com/media/articles/GAJLL\\_31\\_1-20.pdf](https://www.gajrc.com/media/articles/GAJLL_31_1-20.pdf)
- Gotch, C. M., Poppen, M. I., Razo, J. E., & Modderman, S. (2021). Examination of teacher formative assessment self-efficacy development across a professional learning experience. *Teacher Development*, 25(4), 534–548. <https://doi.org/10.1080/13664530.2021.1943503>
- Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, 8(3), 381–391. <https://doi.org/10.1080/135406002100000512>
- Hanefer, S. B., Anny, N., & Rahman, S. (2022). Enhancing teaching and learning in higher education through formative assessment: Teachers' perceptions. *International Journal of Assessment Tools in Education*, 9(1), 61–79. <https://doi.org/10.21449/ijate.946517>
- Hasim, Z., Di, S., & Barnard, R. (2018). Eliciting teachers' understanding and their reported practices on school-based formative assessment: Methodological challenges. *Indonesian Journal of Applied Linguistics*, 8(1). <https://doi.org/10.17509/ijal.v8i1.11476>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Corwin Press.
- Holliday, A. (1994). *Appropriate methodology and social context*. Cambridge University Press.
- Khodabakhshzadeh, H., Kaf, Z., & Hosseinnia, M. (2018). Investigating EFL teachers' conceptions and literacy of formative assessment: Constructing and validating an inventory. *International Journal of Instruction*, 11(1), 139–152. <https://doi.org/10.12973/iji.2018.11110a>
- Matsuda, A. (2002). Representation of users and uses of English in beginning EFL textbooks in Japan. *World Englishes*, 21(2), 269–279. <https://doi.org/10.1111/1467-971X.00248>
- Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, 115(3), 1–47.
- Nanji, K. S. A. (2023). *In-service Science teachers professional development for designing unit for project based learning by design thinking process* [MPhil Dissertation]. Aga Khan University. [https://ecommons.aku.edu/theses\\_dissertations/2211/](https://ecommons.aku.edu/theses_dissertations/2211/)
- Ozan, C., & Kincal, R. Y. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Educational Sciences: Theory & Practice*, 18(1), 85–118. <https://doi.org/10.12738/estp.2018.1.0216>
- Pakistan Education Statistics. (2022). *National education management information system report*. Academy of Educational Planning and Management, Ministry of Federal Education and Professional Training.
- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84, 128–138. <https://doi.org/10.1016/j.tate.2019.05.003>
- Rashid, A. (2025, March 15). Challenges faced by Pakistani English teachers and how to overcome them. *LinkedIn*. <https://www.linkedin.com/pulse/challenges-faced-pakistani-english-teachers-how-overcome-abdul-rashid-cmg6f>
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53(4), 5–23. <https://doi.org/10.1007/BF02504682>

- Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., & Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, 28(3), 1–33. <https://doi.org/10.1080/0969594x.2021.1884042>
- Zeng, J., & Huang, L. (2021). Understanding formative assessment practice in the EFL exam-oriented context: An application of the theory of planned behavior. *Frontiers in Psychology*, 12, 774159. <https://doi.org/10.3389/fpsyg.2021.774159>