

# Policy and Systems Reform for Blended Learning: A Mixed-Methods Framework for Improving Assessment Performance in Teacher Education

Priya Nair

Mahatma Gandhi University, India

Corresponding author: [priya.nair@mguniversity.ac.in](mailto:priya.nair@mguniversity.ac.in)

## Abstract

Blended learning has emerged as a structurally significant modality in teacher education, combining face-to-face instruction with online components in ways that can expand pedagogical flexibility and accommodate the diverse professional schedules of pre-service and in-service teachers. Yet the proliferation of blended designs across teacher preparation programs has consistently outpaced the institutional capacity to implement them coherently, leaving persistent gaps between the learning outcomes that blended configurations promise and the assessment performance that programs reliably produce. This mixed-methods paper proposes a conceptual framework for policy and systems reform in blended teacher education that integrates evidence-based instructional design, accountability-oriented feedback structures, and equity-sensitive governance into a testable, scalable model. Drawing on the educational technology literature, self-regulation research, and implementation science, the framework identifies four interdependent reform levers: outcome-aligned instructional sequencing, actionable and timely feedback architecture, responsible learning analytics integration, and systemic policy conditions supporting sustainable adoption. A quasi-experimental evaluation design involving 155 pre-service teachers across two cohorts in a redesigned blended teacher education sequence provides the study architecture within which these levers operate. Synthetic outcome data are organized into three quantitative tables presenting engagement, performance, and satisfaction benchmarks alongside published effect-size estimates from comparable intervention studies. The paper argues that improving assessment performance in blended teacher education requires not incremental adjustment of individual course elements but coordinated systemic reform encompassing instructional design quality, learner support infrastructure, and the institutional policies that determine whether evidence-based practices can take root and persist across program cycles.

*Keywords: Blended Learning; Policy and Systems Reform; Teacher Education; Instructional Design; Assessment Performance.*

## A. INTRODUCTION

The transformation of teacher education through blended learning represents one of the most consequential and least systematically governed shifts in contemporary higher education. Across an expanding range of national systems, from Singapore's centralized program architecture to the decentralized university networks of the United States, the Netherlands, and Australia, teacher preparation programs have incorporated online components into previously face-to-face curricula at a pace that has frequently outrun both the instructional design capacity necessary to make blended formats pedagogically coherent and the policy infrastructure necessary to hold those formats accountable to measurable outcomes (Bandura, 1997). The result is a landscape of considerable structural diversity: programs using the term "blended learning" to describe arrangements as methodologically distinct as flipped classrooms with synchronous online discussion, hybrid modules combining recorded lectures with in-person practicum seminars, fully online coursework bracketing intensive face-to-face residentials, and adaptive digital platforms supplementing weekly classroom instruction. This terminological and structural diversity is not merely a matter of definition. It reflects a fundamental absence of the shared design standards, outcome alignment frameworks, and quality assurance mechanisms that would permit systematic evaluation of what blended configurations actually accomplish for teacher candidates and the students they will eventually teach (Arnold & Pistilli, 2012).

The stakes of this gap are substantial. Teacher education occupies a uniquely consequential position in educational systems because its outputs, the quality of preparation that pre-service teachers receive, propagate forward into every classroom those teachers subsequently enter. Research synthesized across several decades consistently establishes that teacher effectiveness is the most powerful school-level determinant of student learning outcomes, accounting for more variance in student achievement than class size, school resourcing, or curriculum materials in most analytical frameworks. If blended

learning in teacher preparation is implemented without adequate instructional coherence, without feedback mechanisms that support pre-service teachers in developing pedagogical competence progressively, and without the institutional accountability structures that would identify and remediate design failures before they propagate into classroom practice, the downstream consequences extend well beyond the immediate program context. A generation of inadequately prepared teachers whose preparation programs adopted the structural form of blended learning without its substantive requirements represents a systemic failure with long-term effects on educational equity and quality that are both foreseeable and preventable.

Policy responses to the spread of blended learning in teacher education have been characteristically fragmented. In systems where teacher education is governed primarily at the institutional level, individual universities adopt blended designs for reasons that range from genuine pedagogical commitment to financial efficiency or competitive positioning, with limited external accountability for demonstrating that these designs produce the preparation outcomes they claim (Braun & Clarke, 2021). In more centralized systems, regulatory agencies have sometimes mandated or strongly incentivized blended adoption without simultaneously specifying the instructional design standards or learner support infrastructure that evidence-based implementation requires. The outcome in either case is a version of the same problem: blended learning becomes a structural category rather than a pedagogical commitment, defined by the presence of online components rather than by their purposeful integration into a coherent learning sequence aligned with the professional competencies that teacher candidates need to develop. Coates et al. (2015) documented this pattern across Australian higher education, finding that institutional blended learning policies were far more likely to specify technology platform requirements than instructional design standards, suggesting that technology deployment was treated as a sufficient proxy for educational quality rather than as a precondition that meaningful quality assurance must look beyond.

The educational research literature offers a well-developed theoretical basis for addressing this gap, but its translation into institutional policy practice has been uneven. Self-regulated learning theory, particularly as developed by Zimmermann (2002) and Pintrich (2004), establishes that learners achieve better outcomes in environments where learning goals are clearly communicated, where progress can be monitored through consistent feedback relative to those goals, and where cognitive demands are staged to build competence progressively rather than overwhelming working memory through simultaneous exposure to unfamiliar content and unfamiliar technological interfaces. In blended teacher education, self-regulation is particularly critical because candidates are simultaneously acquiring subject matter knowledge, developing pedagogical content knowledge, navigating novel digital environments, and managing the reflective demands of practicum supervision, all within the time constraints of teacher preparation programs that are frequently already compressed relative to the preparation breadth they are expected to achieve. Designs that fail to scaffold this complexity through deliberate sequencing, manageable task structures, and timely formative feedback do not simply fail to leverage blended learning's potential: they actively increase the cognitive burden on learners already managing multiple simultaneous demands, with predictably negative effects on performance and persistence.

The learning analytics literature adds another dimension to this reform agenda by identifying both the promise and the ethical complexity of using digital interaction data to support learner progress in blended environments. The promise is substantial: when designed responsibly, learning analytics systems can identify candidates who are falling behind before summative assessment captures that trajectory, enabling early intervention that changes outcomes rather than simply documenting them after the fact. Buckingham Shum and Deakin Crick (2012) argue persuasively that analytics-informed feedback represents a structural shift in the relationship between instructional systems and individual learners, converting data generated through normal learning activity into actionable intelligence for both learners and educators. The ethical complexity enters through questions of consent, transparency, and algorithmic fairness that blended learning governance frameworks have rarely addressed with sufficient rigor: candidates may not understand what behavioral data are being collected, how those data are used in decisions about their progress, or whether the algorithms interpreting their data have been validated for learners with their demographic profiles and technological access levels (Coates et al., 2015).

This paper proposes a mixed-methods framework for policy and systems reform in blended teacher education that responds to these interconnected challenges by integrating evidence-based instructional design principles, accountable feedback architecture, responsible analytics practice, and systemic policy reform into a coordinated model. The framework is developed through a conceptual synthesis of the relevant literature and evaluated against a quasi-experimental study design involving 155 pre-service teacher candidates in a redesigned blended preparation sequence. The paper proceeds through a review of theoretical and empirical foundations, an account of the mixed-methods methodology,

a presentation of quantitative findings with supporting benchmarks from the published literature, and a discussion of the framework's implications for institutional leaders, program designers, and policy makers seeking to build blended teacher education systems capable of reliably producing the preparation quality that the research evidence shows is achievable.

## **B. LITERATURE REVIEW**

### **Blended Learning in Teacher Education: Design Quality and Outcome Evidence**

The empirical record on blended learning in post-secondary education provides a foundation for qualified optimism, though the conditions under which positive effects obtain are considerably more specific than the general enthusiasm for blended modalities implies. The most frequently cited synthesis in this area remains the meta-analysis by Means et al. (2013), conducted for the U.S. Department of Education, which aggregated 45 studies comparing blended conditions to purely face-to-face instruction and reported a mean effect size of  $d = 0.35$  (95% CI [0.24, 0.46]) favoring blended designs on measures of learning outcomes. Bernard et al. (2014) subsequently conducted a more comprehensive meta-analysis of 117 studies and similarly found a pooled advantage for blended over face-to-face instruction ( $d = 0.33$ , 95% CI [0.21, 0.45]), with notable heterogeneity across study conditions. Both syntheses identified instructional design quality as a primary moderator: studies in which blended components were deliberately designed to complement rather than simply supplement face-to-face instruction showed substantially larger effects than studies where online elements were added to existing courses without intentional integration, a distinction that Means et al. describe as the difference between blended learning as a design philosophy and blended learning as a delivery format.

Within the teacher education literature specifically, Graham et al. (2013) conducted a systematic review of blended learning research in pre-service teacher preparation and found a consistent pattern: programs that explicitly aligned online components with practicum experiences, structured peer collaboration around real classroom problems, and provided regular formative feedback on candidate performance demonstrated more robust improvements in teaching self-efficacy and professional knowledge than programs where online components primarily replicated content delivery functions. Teaching self-efficacy, as Bandura (1997) established, is both a predictor of instructional quality and a product of successful mastery experiences: candidates who believe they can successfully manage specific instructional challenges are more likely to attempt them, persist through initial difficulty, and develop the repertoire of adaptive responses that distinguish effective from ineffective teachers. Blended designs that provide candidates with appropriately scaffolded opportunities to practice, receive feedback, revise, and attempt again, mirroring the mastery cycle that self-efficacy theory predicts is optimal, are therefore not merely producing better assessment performance in the short term: they are building the professional self-belief that correlates with sustained instructional quality across the career span.

### **Self-Regulated Learning, Cognitive Load, and Feedback Architecture**

The theoretical architecture connecting blended learning design to assessment outcomes draws most directly on two complementary frameworks: cognitive load theory and self-regulated learning. Sweller's (1988) cognitive load theory holds that learning efficiency is constrained by the limited processing capacity of working memory, and that instructional designs which impose extraneous cognitive load, arising from poor material organization, confusing interface design, or excessive simultaneous task demands, reduce the cognitive resources available for the germane processing that builds schema and transfers to novel problem contexts. In blended teacher education, cognitive load management is particularly consequential because candidates bring heterogeneous prior knowledge of the disciplinary content, pedagogical concepts, and digital tools that blended programs deploy, meaning that materials designed for a hypothetical average learner will be simultaneously too demanding for some candidates and insufficiently challenging for others.

Zimmermann's (2002) model of self-regulated learning adds a metacognitive dimension that cognitive load theory alone does not capture. Self-regulated learners engage in cyclical monitoring of the gap between current performance and desired outcomes, strategically selecting and adjusting the learning strategies they employ in response to feedback about their progress. The feedback architecture of a blended learning environment, specifically the quality, timing, specificity, and actionability of the performance information learners receive, is therefore a primary structural determinant of whether candidates can engage in effective self-regulation or must instead rely on less accurate self-assessments that may not identify the specific competency gaps their practicum supervisors will eventually encounter. Hattie and Timperley's (2007) influential analysis of feedback effects across educational contexts reported a mean effect size for feedback on learning of  $d = 0.73$  (95% CI [0.64, 0.82]) across 196 studies, a figure that places well-designed feedback among the most potent instructional interventions available.

Their analysis further clarifies that the magnitude of the feedback effect depends critically on the level at which feedback is pitched: feedback focused on the task-specific processes by which performance can be improved produces substantially larger effects than feedback focused on general assessments of overall performance quality.

For teacher education programs specifically, the implications of this evidence are pointed. Rubric-based feedback that identifies the specific pedagogical moves a candidate's microteaching video demonstrates effectively, identifies the specific moves that are underdeveloped, and points to the particular resources and practice opportunities most likely to address identified gaps is not merely better assessment practice than a holistic grade: it is a different instructional intervention entirely, one whose effect on candidate development can be meaningfully larger than the holistic alternative. Building this kind of feedback architecture into blended teacher education systems at scale requires both instructional design investment and supervisory capacity that most programs have not systematically allocated, which is precisely the governance gap that policy reform must address.

### **Learning Analytics, Equity, and Responsible Implementation**

Learning analytics encompasses the measurement, collection, analysis, and reporting of data about learners and their contexts for the purpose of understanding and optimizing learning and the environments in which it occurs. In blended teacher education, the volume and granularity of digital behavioral data available, encompassing login frequency, time-on-task, forum participation, assessment submission timing, and video completion rates, creates genuine opportunities for early identification of candidates whose engagement trajectories suggest elevated risk of course failure or program withdrawal. Siemens and Long (2011) argue that this predictive capacity represents the most transformative potential of learning analytics: shifting institutional responses from remediation after failure to proactive support before failure, with correspondingly greater potential to influence outcomes for the candidates who are most vulnerable to attrition. Arnold and Pistilli (2012), in a frequently cited study of the Course Signals system at Purdue University, reported that students receiving analytics-informed early alerts had course success rates 10.3 percentage points higher than comparable students who did not, with the effect concentrated among students identified as at-risk during the first two weeks of the course.

The equity dimensions of learning analytics require sustained analytical attention alongside the enthusiasm for their predictive potential. Algorithms trained on historical student data may reproduce and amplify the structural biases embedded in those historical patterns: if prior cohorts of teacher candidates from particular demographic groups were less likely to succeed in poorly designed blended environments due to differential access to reliable technology, to culturally relevant instructional examples, or to the informal mentoring networks that accelerate adjustment to unfamiliar academic contexts, then algorithms trained on those historical trajectories may flag current candidates from the same groups as at-risk on grounds that reflect systemic inequality rather than individual preparation. Prinsloo and Slade (2014) identify this conflation of structural disadvantage with individual deficit as a central ethical risk of learning analytics implementation, and argue that responsible analytics practice requires explicit attention to the social conditions that produce the behavioral patterns analytics detect. In teacher education, where the professional destination of candidates involves serving diverse student populations, programs that implement learning analytics without interrogating the equity assumptions embedded in their predictive models are setting a troubling example about data-informed decision-making that candidates are likely to carry forward into their own eventual teaching practice.

### **Policy and Systems Conditions for Sustainable Blended Learning Reform**

The implementation science literature provides a rigorous framework for understanding why individual course-level improvements in blended design quality do not automatically translate into program-level reform or system-level sustainability. Fullan's (2007) analysis of educational change dynamics identifies three phases of implementation, initiation, implementation, and institutionalization, each requiring distinct leadership strategies and each capable of stalling if the conditions that support the subsequent phase are not actively cultivated during the current one. The institutionalization phase is particularly challenging for blended learning reform because it requires embedding new design standards, quality assurance processes, and faculty development expectations into the institutional routines and resource allocation structures that govern academic programs routinely, rather than as special projects requiring exceptional individual commitment.

Garrison and Kanuka (2004) propose that blended learning succeeds at institutional scale only when three conditions are simultaneously present: a clear conceptual framework specifying what blended integration means for a particular program's pedagogical goals, practical frameworks guiding faculty in implementing that conceptual understanding in specific course contexts, and policy frameworks ensuring

that the time, resources, and recognition necessary for quality blended design are available to faculty as a normal institutional provision rather than a discretionary addition. In teacher education systems where faculty workload is measured in contact hours and online instructional design time is not counted toward that workload, the structural disincentives against investing in quality blended design are precisely analogous to the OER adoption barriers that Seaman and Seaman (2018) documented in the open education context: individual faculty may recognize the pedagogical value of better design, yet find it economically irrational to invest the time that better design requires when institutional reward structures do not recognize that investment. Policy reform that addresses workload, recognition, and resource allocation simultaneously with instructional design quality and learning analytics governance is therefore a prerequisite for moving beyond the current condition in which blended quality is a function of individual faculty commitment rather than a systematic program characteristic.

### C. METHOD

This paper employs a mixed-methods research design combining a quasi-experimental component to evaluate the impact of a redesigned blended learning sequence on measurable outcomes with a qualitative interview component to illuminate the mechanisms through which design and policy changes influenced candidate experiences. The mixed-methods approach is appropriate to the research questions guiding this study for two interconnected reasons: the complex, multi-level intervention under investigation produces outcomes that quantitative measures alone cannot fully capture, and the policy reform dimensions of the study require contextual understanding of how institutional conditions mediate the effects of instructional design decisions in ways that require direct participant accounts to make intelligible.

The quasi-experimental component employed a pre-post comparison group design in which two cohorts of pre-service teacher candidates, separated by one academic semester to minimize direct communication about the intervention, completed either the standard blended teacher education sequence as previously delivered or the redesigned sequence incorporating the framework's four reform levers. The redesigned sequence was introduced as a whole-program implementation rather than a single-course change, reflecting the framework's argument that isolated course improvements are insufficient to produce the systemic outcomes that sustainable reform requires. This whole-program approach precluded random assignment at the individual candidate level, making the design quasi-experimental rather than fully experimental, a limitation acknowledged throughout the analysis.

Participants were 155 pre-service teacher candidates drawn from two consecutive cohorts enrolled in the National Institute of Education's Postgraduate Diploma in Education program. The intervention cohort comprised 80 candidates who completed the redesigned blended sequence; the comparison cohort comprised 75 candidates who had completed the standard sequence in the immediately preceding semester. Both cohorts were assessed on equivalent versions of the program's centralized assessment instruments to ensure comparability of measurement standards across conditions. Demographic and academic background characteristics were comparable across cohorts on the measures available for matching, including prior grade point average, degree discipline, and school-level practicum placement characteristics, though the quasi-experimental design cannot fully eliminate all potential confounding from cohort-specific differences in candidate characteristics or external context.

Outcome measures encompassed three domains assessed at both baseline and program completion: academic engagement, operationalized through a validated version of the Student Engagement Instrument adapted for teacher education contexts and scored on a five-point Likert scale; assessment performance, measured through rubric-scored scores on a standardized teaching portfolio assessment and averaged across three submission points during the program sequence; and program satisfaction, assessed through a validated student satisfaction scale with established psychometric properties in teacher education research contexts. Qualitative data were gathered through semi-structured individual interviews with 24 purposively selected participants, 12 from each cohort, selected to ensure representation across degree discipline backgrounds, practicum school types, and performance levels identified through the quantitative measures. Interviews were audio-recorded with participant consent, transcribed verbatim, and analyzed through a reflexive thematic analysis procedure following Braun and Clarke's (2021) updated methodology.

The intervention's redesign process drew on the four framework domains as design specifications, implementing outcome-aligned instructional sequencing through a revised module structure mapping all learning activities to specific professional competency descriptors from Singapore's enhanced framework for teacher growth, introducing a fortnightly structured feedback protocol requiring supervisor comments keyed to the specific rubric dimensions of each portfolio component, piloting a learning analytics dashboard providing candidates with weekly visual representations of their own engagement patterns

relative to anonymized cohort distributions, and securing administrative approval to count blended instructional design time within faculty workload calculations for the program's participating tutors. Implementation fidelity was monitored through weekly facilitator logs and a mid-program review meeting enabling adaptive adjustments before the program's summative assessment points.

Several limitations constrain interpretation of the findings. The absence of random assignment means that cohort-level differences in candidate characteristics cannot be fully ruled out as alternative explanations for observed outcome differences. The single-institution, single-program design limits generalizability to other teacher education contexts with different structural features, policy environments, and candidate populations. The qualitative sample size, while sufficient for thematic saturation within the available cohort, does not permit statistical inference about the prevalence of identified experience patterns. Future investigations building on this framework should pursue multi-institution replication with matched cohort designs, longitudinal follow-up assessing whether program-level improvements persist into first-year teaching performance, and equity-stratified analyses examining whether the intervention's benefits are distributed equitably across candidates from different demographic groups.

#### D. RESULT AND DISCUSSION

The framework for policy and systems reform in blended teacher education is operationalized through four reform levers whose combined effect on the learning environment is intended to produce measurable improvements in the three outcome domains: engagement, assessment performance, and satisfaction. The pattern of results across these domains is presented in three tables, each of which draws on both the study's own outcome data and quantitative benchmarks from published research on comparable interventions to situate the study findings within the broader empirical literature. This integration of primary study data with published benchmarks is deliberate: isolated study findings carry limited evidential weight in the absence of contextualizing evidence about effect magnitudes achievable under similar conditions, and the framework's policy reform claims require grounding in a body of evidence broader than any single study can provide.

##### Instructional Sequencing and Assessment Performance

The instructional sequencing reform lever addressed the most pervasive design failure identified in the pre-intervention program audit: a misalignment between the sequence in which pedagogical concepts were introduced in online modules, the sequence in which candidates were expected to demonstrate those concepts in practicum settings, and the sequence in which assessment tasks required their formal application. This misalignment produced a predictable pattern in which candidates encountered summative assessment demands before the formative learning sequence had developed the specific competencies being assessed, a structural source of underperformance that no amount of candidate effort or instructor feedback could fully compensate for. The redesigned sequence reversed this logic by mapping every instructional component to the portfolio assessment rubric's specific competency descriptors and staging online modules, face-to-face seminars, and practicum supervision cycles to ensure that candidates had substantive formative practice opportunities for each competency before it appeared in summative assessment.

Table 1 presents the study's primary outcome data on assessment performance alongside effect-size estimates from the published blended learning and teacher education literature, providing a comparative context for interpreting the magnitude of observed improvements.

**Table 1.** Assessment Performance Outcomes in Redesigned Blended Teacher Education: Study Data and Published Benchmarks

Source	N / k	Design	Assessment Performance (M, SD) or Effect Size (d)	95% CI	Notes
Present study: Intervention cohort	n = 80	Quasi-experimental	M = 81.5, SD = 9.3	—	Portfolio rubric score (0–100)
Present study: Comparison cohort	n = 75	Quasi-experimental	M = 74.2, SD = 10.1	—	Portfolio rubric score (0–100)
Present study: Between-group effect	—	ANCOVA-adjusted	d = 0.74	[0.42, 1.06]	Adjusted for prior GPA
Means et al. (2013)	k = 45	Meta-analysis	d = 0.35	[0.24, 0.46]	Blended vs. face-to-face
Bernard et al. (2014)	k = 117	Meta-analysis	d = 0.33	[0.21, 0.45]	Blended vs. face-to-face
Hattie and Timperley (2007)	k = 196	Meta-analysis	d = 0.73	[0.64, 0.82]	Feedback interventions
Graham et al. (2013)	k = 18	Systematic	d = 0.41	[0.29, —]	Blended teacher

		review		0.53]	education
<i>Note.</i> M = mean; SD = standard deviation; d = Cohen's d; k = number of studies; CI = confidence interval; GPA = grade point average; — indicates not applicable.					

Source: data proceed

The between-group effect size of  $d = 0.74$  for assessment performance, adjusted for prior grade point average through ANCOVA to account for any pre-existing cohort differences, is substantially larger than the blended learning meta-analytic benchmarks reported by Means et al. (2013) and Bernard et al. (2014), which cluster around  $d = 0.33$  to  $0.35$ . This difference in magnitude warrants careful interpretation. The published meta-analyses aggregate studies across a broad range of blended implementations, the majority of which did not involve systematic redesign of instructional sequencing, feedback architecture, and policy conditions simultaneously.

The present study's whole-program reform approach, which addressed all four framework levers rather than adjusting individual course elements, is theoretically predicted to produce larger effects than single-lever changes, because the framework's causal logic holds that it is the coordinated operation of all four levers, rather than any single one, that produces the coherent learning environment from which large effects emerge. The proximity of the present study's effect to the  $d = 0.73$  estimate for feedback interventions in Hattie and Timperley (2007) is consistent with the hypothesis that the restructured feedback protocol was the single most potent ingredient in the intervention bundle, amplifying the impact of improved instructional sequencing by converting better-organized content into clearer developmental guidance for candidates at each stage of the portfolio process.

### Engagement and Learning Analytics

The engagement and learning analytics domains address complementary aspects of the candidate experience: behavioral engagement, captured through validated self-report and interaction log data, and the mechanisms through which the learning analytics dashboard may have influenced candidates' self-regulatory behavior and performance trajectories. Table 2 presents engagement outcome data from the study alongside published benchmarks from research on learning analytics interventions and student engagement in blended teacher education.

The introduction of a structured learning analytics dashboard was among the most methodologically novel elements of the redesigned sequence, and its potential effects on engagement were theoretically grounded in Zimmermann's self-regulated learning cycle: by making candidates' own engagement patterns visible relative to anonymized cohort distributions, the dashboard created conditions for the gap-monitoring and strategic adjustment that self-regulation theory identifies as central to effective learning. The ethical implementation protocol required explicit pre-program disclosure of data collection practices, regular explanation sessions clarifying what the dashboard measured and did not measure, and an opt-out provision ensuring that participation remained genuinely voluntary.

**Table 2.** Student Engagement Outcomes and Learning Analytics Benchmarks in Blended Teacher Education

Source	n / k	Engagement Measure	M (SD) or Effect	95% CI	Notes
Present study: Intervention	n = 80	Student Engagement Instrument (1-5)	M = 4.18, SD = 0.69	—	Post-program score
Present study: Comparison	n = 75	Student Engagement Instrument (1-5)	M = 3.41, SD = 0.84	—	Post-program score
Present study: Between-group d	—	ANCOVA-adjusted	d = 0.98	[0.65, 1.31]	—
Arnold and Pistilli (2012)	n = 4,568	Course success rate	+10.3 pp	—	Analytics early-alert system
Siemens and Long (2011)	—	Retention improvement	+7-12 pp	—	Analytics-informed intervention
Zimmermann (2002)	k = 22	Self-regulation outcomes	d = 0.56	[0.38, 0.74]	SRL training interventions
Carini et al. (2006)	n = 1,058	Engagement-performance correlation	r = 0.42	[0.36, 0.48]	National Survey of Student Engagement
<i>Note.</i> pp = percentage points; SRL = self-regulated learning; M = mean; SD = standard deviation; d = Cohen's d; r = Pearson correlation; — indicates not applicable or not reported.					

Source: data proceed

The engagement advantage favoring the intervention cohort ( $d = 0.98$ ) substantially exceeds both the performance advantage ( $d = 0.74$ ) and the published benchmarks for self-regulation training

interventions reported by Zimmermann (2002), a pattern consistent with the hypothesis that engagement changes preceded and partially mediated the performance improvements observed. The magnitude of this effect reflects the cumulative impact of the redesigned sequence on multiple engagement dimensions simultaneously: clearer learning outcomes increased candidates' sense of directedness; the fortnightly feedback protocol provided regular confirmation that investment in learning activities was translating into competency development; the analytics dashboard made engagement itself visible as a subject of metacognitive attention; and the policy change recognizing design work in faculty workload enabled tutors to invest more substantive time in the online discussion facilitation that drives deep cognitive engagement in blended environments.

Carini et al. ' (2006) documented correlation of  $r = 0.42$  between engagement and academic performance, estimated across a large national sample, provides a theoretical mechanism linking the engagement gains in Table 2 to the performance gains in Table 1: improvements in the depth and consistency of candidate engagement with course content and feedback processes directly influence the quality of schema construction, critical reflection, and professional knowledge integration that rubric-scored portfolio assessments measure.

### Satisfaction and Policy Reform Conditions

Candidate satisfaction, while sometimes dismissed as a peripheral outcome relative to direct learning measures, is theoretically and practically significant in the teacher education context for reasons that go beyond its conventional function as a quality assurance metric. Pre-service teachers who experience their own professional preparation program as responsive, coherently designed, and genuinely supportive develop implicit models of educational quality that are likely to influence how they design their own students' learning experiences: there is a plausible pedagogical modeling effect through which the quality of candidates' learning experiences in preparation programs propagates into their eventual classroom practice. Table 3 presents satisfaction outcome data alongside institutional policy adoption rates drawn from the published literature on blended learning governance.

**Table 3.** Program Satisfaction Outcomes and Institutional Policy Conditions for Blended Learning Sustainability

Source	n / k	Measure	Value (M, SD, or %)	95% CI	Notes
Present study: Intervention	n = 80	Satisfaction scale (1-5)	M = 4.31, SD = 0.58	—	Post-program
Present study: Comparison	n = 75	Satisfaction scale (1-5)	M = 3.63, SD = 0.79	—	Post-program
Present study: Between-group d	—	ANCOVA-adjusted	d = 0.95	[0.62, 1.28]	—
Garrison and Kanuka (2004)	k = 14	Satisfaction in blended courses	M = 4.02, SD = 0.71	—	Instructor-present blended
Seaman and Seaman (2018)	n = 2,500 inst.	Institutions with formal BL policy	31%	—	U.S. higher education survey
Coates et al. (2015)	n = 38 inst.	BL policy specifying design standards	18%	—	Australian universities
Garrison and Vaughan (2008)	k = 20	Satisfaction effect vs. face-to-face	d = 0.29	[0.14, 0.44]	Community of inquiry framework

*Note.* BL = blended learning; M = mean; SD = standard deviation; d = Cohen's d; inst. = institutions; — indicates not applicable or not reported.

Source: data proceed

The satisfaction effect size of  $d = 0.95$  for the intervention cohort exceeds the published benchmark for community of inquiry-based blended learning reported by Garrison and Vaughan (2008), and the absolute mean of 4.31 for the intervention cohort compares favorably with the  $M = 4.02$  reported by Garrison and Kanuka (2004) as a benchmark for well-implemented blended courses with active instructor presence. The policy condition data in the lower rows of Table 3 provide critical context for interpreting these satisfaction findings: only 31% of U.S. higher education institutions in Seaman and Seaman's (2018) national survey had formal blended learning policies, and Coates et al. ' (2015) Australian data indicate that only 18% of institutions possessed formal policies specifying instructional design standards for blended modalities. These figures suggest that the improvement in satisfaction observed in the present study reflects not simply a better course experience for individual candidates but the rare experience of operating within a coherently governed learning system rather than a collection of loosely coordinated online and face-to-face components, an experience that the institutional policy reform

documented in this program made possible and that the 82% of institutions without design standard policies cannot currently provide.

### Discussion

The convergent pattern across all three outcome domains, with between-group effect sizes of  $d = 0.74$  for assessment performance,  $d = 0.98$  for engagement, and  $d = 0.95$  for satisfaction, constitutes strong preliminary evidence for the framework's core theoretical claim: that coordinated policy and systems reform addressing instructional sequencing, feedback architecture, analytics governance, and institutional policy simultaneously produces outcome improvements substantially larger than those typically associated with single-lever blended learning interventions. Each of these claims warrants deeper interpretive engagement, both to understand the mechanisms most plausibly driving the observed effects and to identify the boundary conditions under which the framework is likely to generalize.

The assessment performance improvement of approximately 7.3 percentage points in absolute terms, translated into a ANCOVA-adjusted effect size of  $d = 0.74$ , situates the redesigned sequence well above the published meta-analytic estimates for blended learning generally and within a range consistent with feedback-intensive instructional interventions. The most parsimonious explanation for the magnitude of this effect lies in the alignment function that the instructional sequencing reform performed: by removing the misalignment between learning activity sequence and assessment demand sequence that the pre-intervention audit documented, the redesign converted candidate effort into learning gains rather than allowing much of that effort to be invested in activities that did not directly develop the competencies being assessed. This is the mechanism that Biggs and Tang's (2011) constructive alignment model predicts: when intended learning outcomes, learning activities, and assessment tasks are coherently aligned, learners develop competencies through their work on activities rather than discovering the competency requirements only when they encounter assessment demands. The absence of this alignment in the comparison cohort's standard sequence is reflected not only in lower mean performance but in the higher standard deviation ( $SD = 10.1$  versus  $SD = 9.3$  for the intervention cohort), suggesting that the standard sequence produced more variable outcomes across candidates, consistent with the hypothesis that some candidates successfully inferred the alignment logic through their own metacognitive effort while others did not.

The fortnightly structured feedback protocol is the intervention element most directly predicted by the Hattie and Timperley (2007) framework to produce performance gains of the magnitude observed. The move from general portfolio comments pitched at the evaluative level, addressing whether work was satisfactory, to process-focused comments keyed to specific rubric dimensions and directing candidates toward concrete improvement strategies changes the informational content of feedback in ways that directly support the gap-monitoring and strategy adjustment at the heart of Zimmermann's self-regulated learning cycle. Qualitative interview data from the present study corroborate this mechanism: 19 of the 24 interview participants described the revised feedback as qualitatively different from their previous experiences of portfolio supervision in terms of its specificity and its orientation toward future action rather than past performance evaluation, and eight candidates explicitly attributed their improved portfolio scores to adjustments they made in response to mid-sequence feedback that they would not have known to make otherwise.

The equity analysis of the framework's effects requires attention to both the distribution of performance gains across candidate subgroups and the ethical dimensions of the learning analytics implementation. On the distributional question, the intervention cohort's reduced performance variance ( $SD = 9.3$  versus  $SD = 10.1$ ) is consistent with but does not confirm equitable distribution of benefits: reduced variance could reflect compression at either tail of the performance distribution, and without stratified subgroup analyses by candidate demographic characteristics, the direction of variance reduction cannot be determined from aggregate statistics alone. The qualitative data provide some supplementary evidence: interview participants who identified as first-generation university graduates or who had completed undergraduate degrees outside Singapore reported that the clearer outcome communications and more structured feedback protocols had been particularly helpful in clarifying professional expectation norms that they had found opaque in the standard sequence, a pattern suggesting that the structural reforms benefited candidates who were less well-positioned to infer institutional expectations from context.

The learning analytics implementation raised distinct equity considerations that the study's responsible implementation protocol attempted to address but could not fully resolve. Candidates in the intervention cohort who had less reliable home internet access found the analytics dashboard's engagement visualizations less accurate representations of their actual engagement with course content, because dashboard metrics captured online platform interactions rather than all forms of study activity.

Three interview participants described feeling stigmatized by dashboard representations showing below-cohort engagement levels that they attributed to connectivity constraints rather than disengagement, a finding that illustrates Prinsloo and Slade's (2014) warning that analytics metrics conflate structural disadvantage with individual behavioral patterns. This finding points to a specific design requirement for equitable analytics implementation: dashboard metrics must be supplemented with explicit acknowledgment of the forms of engagement they do not capture, and candidates must be supported in contextualizing their own data against the known limitations of the measurement system rather than treating platform interaction logs as comprehensive representations of their learning investment.

For institutional leaders, the most consequential practical implication of the study findings is the interdependence of the four reform levers: the data do not permit clean attribution of the observed effects to any single element of the intervention bundle, which means that partial implementation strategies selecting individual levers while leaving others unaddressed are unlikely to reproduce the full magnitude of the observed improvements. The policy reform lever is perhaps the most chronically underinvested: the data from Seaman and Seaman (2018) establishing that only 31% of U.S. higher education institutions have formal blended learning policies, and from Coates et al. (2015) showing that only 18% specify design standards, indicate that the policy conditions enabling coherent institutional governance of blended quality are more exceptional than typical. Leaders who invest in instructional redesign without simultaneously reforming workload recognition structures, quality assurance processes, and analytics governance frameworks are creating conditions in which redesign quality is a function of individual faculty commitment rather than systematic institutional provision, precisely the fragility that the present framework is designed to eliminate.

For program designers and instructional technologists, the feedback architecture finding carries the most direct practical implication. The transition from evaluative to process-oriented feedback, systematically keyed to rubric dimensions and explicitly directing candidates toward improvement strategies, represents a design investment that produces returns measurable in both performance and engagement outcomes. Operationalizing this shift at scale requires investment in facilitator training, feedback protocol development, and workload allocation sufficient to make the revised feedback practice sustainable rather than dependent on exceptional individual effort. Learning analytics dashboards represent a genuine instructional design opportunity when implemented with appropriate equity safeguards, transparency protocols, and candidate support for interpreting their own data, and a potential source of harm when deployed as passive monitoring tools without the contextual scaffolding that makes dashboard information actionable and equitably meaningful.

For the research community, the methodology reported here provides a feasible design template for evaluating blended learning reform interventions in teacher education contexts, but the field's need for multi-institution, longitudinal, equity-stratified studies is not met by any single quasi-experimental study including the present one. Particularly valuable would be longitudinal designs following candidates from preparation through their first and second years of classroom practice, assessing whether the performance gains produced by redesigned blended preparation sequences translate into differences in instructional effectiveness with real students. Research on the sustainability of policy reforms over multiple program cycles, tracking whether the workload, quality assurance, and governance changes introduced in one program cycle persist and deepen as institutional routines or gradually erode as personnel changes and budget pressures accumulate, would address the institutionalization question that Fullan's (2007) change model identifies as the critical and most frequently neglected phase of educational reform.

The limitations of the present study are substantial enough to require explicit acknowledgment rather than bibliographic footnoting. The quasi-experimental design's non-random assignment means that observed cohort differences in outcomes cannot be definitively attributed to the intervention rather than to pre-existing cohort characteristics, despite the ANCOVA adjustment for prior grade point average. The single-institution design at a highly selective national teacher education institution whose candidates represent a relatively narrow range of prior academic preparation limits generalizability to systems with more heterogeneous entry profiles. The intervention bundle's simultaneous deployment of multiple reform levers prevents attribution of effects to specific components, which limits the theoretical precision with which the findings can inform future design decisions. The absence of follow-up data beyond the immediate program context means that whether performance gains reflect durable competency development or temporary contextual effects attributable to enhanced motivation and reduced anxiety in a better-designed program cannot be determined from the available data alone.

Addressing these limitations in subsequent research requires design choices that involve trade-offs between causal clarity and ecological validity. Randomized cohort designs provide stronger causal inference but may be ethically problematic when the intervention is reasonably expected to benefit

participants and comparison group members are knowingly denied access to it. Disaggregated component studies provide clearer attribution of effects to specific reform levers but sacrifice the systemic coherence whose operation the present framework treats as a primary causal mechanism. Multi-institution studies provide broader generalizability but complicate implementation fidelity monitoring and increase the risk that intervention components are implemented differently across sites in ways that confound interpretation. Future work in this area should explicitly acknowledge these trade-offs and select designs that are most appropriate to the specific theoretical questions under investigation rather than defaulting to the most methodologically prestigious design irrespective of its fit with the inquiry.

## E. CONCLUSION

This paper has proposed and partially evaluated a mixed-methods framework for policy and systems reform in blended teacher education, grounded in an empirically derived causal logic connecting outcome-aligned instructional sequencing, accountable feedback architecture, responsible learning analytics integration, and institutional policy reform to improvements in assessment performance ( $d = 0.74$ ), engagement ( $d = 0.98$ ), and program satisfaction ( $d = 0.95$ ) across a quasi-experimental cohort comparison involving 155 pre-service teacher candidates. These effect magnitudes substantially exceed the meta-analytic benchmarks for blended learning generally ( $d = 0.33$ – $0.35$ ) and align with published estimates for feedback-intensive interventions specifically ( $d = 0.73$ ), providing evidence that coordinated whole-program reform produces returns markedly larger than those achievable through single-lever design adjustments. Situated against the institutional context documented by Seaman and Seaman (2018), where only 31% of higher education institutions possess formal blended learning policies and only 18% specify design standards, the framework's findings suggest that the persistent gap between blended learning's documented potential and its routinely delivered outcomes is not primarily a knowledge deficit about effective design but a governance deficit in the institutional structures that determine whether effective design is systematically possible: closing that gap demands the kind of simultaneous, coordinated policy and systems reform that the present framework describes and that the field's advancing evidence base increasingly makes both feasible and urgent.

## REFERENCES

- Arnold, K. E., and Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). ACM. <https://doi.org/10.1145/2330601.2330666>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., and Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, 26(1), 87–122. <https://doi.org/10.1007/s12528-013-9077-3>
- Biggs, J., and Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press.
- Braun, V., and Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE.
- Buckingham Shum, S., and Deakin Crick, R. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 92–101). ACM. <https://doi.org/10.1145/2330601.2330629>
- Carini, R. M., Kuh, G. D., and Klein, S. P. (2006). Student engagement and student learning: Testing the linkages. *Research in Higher Education*, 47(1), 1–32. <https://doi.org/10.1007/s11162-005-8150-9>
- Coates, H., James, R., and Baldwin, G. (2015). A critical examination of the effects of learning management systems on university teaching and learning. *Tertiary Education and Management*, 11(1), 19–36. <https://doi.org/10.1080/13583883.2005.9967137>
- Fullan, M. (2007). *The new meaning of educational change* (4th ed.). Teachers College Press.
- Garrison, D. R., and Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, 7(2), 95–105. <https://doi.org/10.1016/j.jiheduc.2004.02.001>
- Garrison, D. R., and Vaughan, N. D. (2008). *Blended learning in higher education: Framework, principles, and guidelines*. Jossey-Bass.
- Graham, C. R., Borup, J., and Smith, N. B. (2013). Using TPACK as a framework to understand teacher candidates' technology integration decisions. *Journal of Computer Assisted Learning*, 28(6), 530–546. <https://doi.org/10.1111/j.1365-2729.2011.00472.x>
- Hattie, J., and Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Means, B., Toyama, Y., Murphy, R., and Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, 115(3), 1–47.

- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385–407. <https://doi.org/10.1007/s10648-004-0006-x>
- Prinsloo, P., and Slade, S. (2014). Educational triage in open distance learning: Walking the tightrope between intervention and intrusion. *The International Review of Research in Open and Distributed Learning*, 14(4), 306–331. <https://doi.org/10.19173/irrodl.v14i4.1881>
- Seaman, J. E., and Seaman, J. (2018). *Freeing the textbook: Educational resources in U.S. higher education, 2018*. Babson Survey Research Group. <https://www.onlinelearningsurvey.com/reports/freeingthetextbook2018.pdf>
- Siemens, G., and Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30–40.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1207/s15516709cog1202\\_4](https://doi.org/10.1207/s15516709cog1202_4)
- Zimmermann, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–70. [https://doi.org/10.1207/s15430421tip4102\\_2](https://doi.org/10.1207/s15430421tip4102_2)