.

Learning Analytics Dashboards in Online Stem Education: A Mixed-Methods Framework for Improving Conceptual Understanding in Teacher Preparation

Emily Carter
University of Auckland New Zealand
Corresponding author: emily.carter@auckland.edu.nz

Abstract
Conceptual understanding in STEM disciplines represents one of the most persistent and consequential challenges in teacher education, where surface-level procedural knowledge without genuine conceptual depth produces teachers inadequately equipped to develop scientific and mathematical reasoning in their own students. Online STEM teacher education programs have proliferated rapidly across tertiary systems, driven by flexibility demands and widening participation agendas, yet their capacity to foster deep conceptual understanding remains unevenly documented and theoretically underdeveloped. This mixed-methods paper proposes a framework for improving conceptual understanding in online STEM teacher education through the purposeful integration of learning analytics dashboards within inquiry-based, interdisciplinary course designs. Drawing on conceptual change theory, cognitive load research, self-regulated learning models, and the emerging empirical literature on learning analytics in STEM contexts, the framework identifies four interdependent design levers: inquiry-structured interdisciplinary task sequences, analytics-informed formative feedback cycles, metacognitive scaffolding for self-monitoring, and institutional conditions governing equitable analytics implementation. A quasi-experimental study design involving 155 pre-service STEM teachers across two online cohorts provides the empirical architecture within which these levers are evaluated. Three quantitative tables present outcome data on conceptual understanding, engagement, and satisfaction alongside published benchmarks drawn from the STEM education and learning analytics literature. The paper argues that improving conceptual depth in online STEM teacher preparation requires simultaneous reform of instructional design, data-informed feedback practice, and the institutional governance structures that determine whether analytics tools serve learner development or reproduce the performance anxieties and equity disparities that poorly governed data systems characteristically generate. Practical implications address program designers, STEM educators, learning technology specialists, and institutional leaders.

Keywords: *STEM Teacher Education; Learning Analytics; Online Learning; Conceptual Understanding; Inquiry-Based Learning.*

A. INTRODUCTION

Science, technology, engineering, and mathematics education occupies a position of exceptional strategic importance in contemporary knowledge economies, where the capacity to reason quantitatively, to construct and evaluate evidence-based arguments, and to apply disciplinary thinking to novel problems determines not only individual occupational outcomes but the collective capacity of societies to address complex challenges in health, climate, infrastructure, and governance. Across this broad disciplinary terrain, teacher quality is the most powerful school-level determinant of whether students develop the genuine conceptual understanding that those capacities require or acquire instead the procedural fluency and surface-level familiarity with disciplinary language that can satisfy assessment demands without producing transferable intellectual competence (Prinsloo & Slade, 2014). The distinction between surface and deep STEM learning is not merely a rhetorical preference for richer educational experiences: it maps onto measurable differences in learners' ability to apply knowledge to unfamiliar problems, to recognize the structural similarities between problems in different surface representations, and to persist productively through the conceptual difficulty that authentic STEM inquiry inevitably involves. Pre-service STEM teachers who lack deep conceptual understanding of the disciplines they will teach are structurally constrained in their capacity to foster that understanding in their students, propagating a deficit that compounds across educational levels and across the careers of every student those teachers instruct (Tempelaar et al., 2015).

The scale of this challenge has given urgency to efforts to expand and improve STEM teacher education, and online delivery has emerged as a primary vehicle for doing so. The arguments for online

STEM teacher education are genuine and substantial. Online formats extend access to qualified teacher preparation to candidates in geographically remote areas who cannot relocate for residential programs, to mid-career professionals seeking to transition into teaching while maintaining employment commitments, and to members of communities whose cultural circumstances make full-time residential study impractical. New Zealand's tertiary system provides a particularly illuminating context for this expansion: the geographic distribution of the national population across two main islands, numerous smaller islands, and a pattern of low-density regional settlement means that online teacher education is not merely a convenience option but a structural necessity for achieving the geographic equity of preparation access that the national teaching workforce requires. Participation data from the New Zealand Teaching Council indicate consistent growth in online and blended initial teacher education enrolments over the preceding decade, with STEM specializations representing a disproportionate share of growth relative to their representation in the broader teaching population, reflecting both targeted recruitment incentives and the professional flexibility that online delivery provides to candidates from technical and scientific professional backgrounds considering a career transition into teaching (Macfadyen, L. P., & Dawson, S. (2010).

The pedagogical challenge this expansion creates is not trivial. Online STEM education confronts a set of design problems that are more demanding than those facing online education in many other disciplines. The development of conceptual understanding in STEM disciplines characteristically requires learners to confront and reorganize existing mental models that are often deeply entrenched, internally coherent, and phenomenologically compelling even when they are scientifically incorrect. Vosniadou's (1994) framework theory of conceptual change establishes that naïve STEM conceptions are not simply missing pieces of knowledge that direct instruction can fill but structured theoretical commitments whose revision requires sustained, targeted cognitive engagement with carefully designed contradictory evidence. In face-to-face STEM education, instructors can monitor conceptual confusion in real time, ask probing questions that expose the specific nature of a candidate's misconception, and adjust instructional pacing and approach accordingly. Online asynchronous formats make this real-time diagnostic responsiveness structurally more difficult, requiring deliberate design substitutes that can perform the diagnostic and adaptive functions that skilled face-to-face STEM instruction provides through immediate interpersonal responsiveness.

Learning analytics represents one promising structural substitute for the real-time diagnostic function that online format constraints preclude. By collecting, aggregating, and visualizing the behavioral data generated through normal online course participation, learning analytics systems can make patterns in candidate engagement, performance trajectories, and conceptual struggle visible to both instructors and learners themselves in ways that support timely, targeted intervention. The promise of this technology is grounded in a plausible causal mechanism: candidates who can see their own conceptual development trajectories, who can identify the specific topics where their engagement or performance data suggest understanding is thin, and who receive feedback that is calibrated to their individual developmental needs rather than pitched at a hypothetical class average should be better positioned to direct their cognitive resources toward the specific conceptual work their development requires. Instructors who can monitor dashboard representations of class-wide engagement and performance patterns can identify the topics generating the most widespread conceptual difficulty and concentrate instructional attention accordingly, converting aggregate behavioral data into targeted pedagogical decisions that isolated individual assessments cannot efficiently produce.

Yet the translation of this promise into reliably beneficial practice is contingent on a set of design and governance conditions that the literature has begun to specify with increasing precision, and which institutional practice in online STEM education has been considerably slower to implement. Learning analytics tools that present engagement metrics without adequate contextual scaffolding for interpretation can activate performance anxiety rather than productive self-regulation, leading candidates to fixate on their standing relative to cohort benchmarks rather than attending to the substantive quality of their conceptual engagement. Analytics systems that track platform interactions as proxies for learning engagement systematically misrepresent the learning activities of candidates who study from printed notes, engage in physical laboratory work, or discuss course concepts through informal peer channels that generate no platform data. In STEM disciplines where hands-on laboratory experience and collaborative problem-solving are epistemologically central to authentic disciplinary practice, the gap between what platform analytics can measure and what disciplinary learning actually requires is particularly significant and particularly consequential for the validity of analytics-informed feedback (Margulieux et al., 2016).

Equity dimensions compound these design challenges in ways that the analytics literature has only recently begun to address systematically. Online STEM teacher education draws candidates from demographic groups with historically lower rates of STEM degree attainment, including women in physics and engineering specializations, Pacific and Māori candidates in the New Zealand context, and candidates from lower socioeconomic backgrounds who represent the first in their families to pursue tertiary STEM study. These are precisely the candidates whose participation in STEM education is most educationally consequential from an equity standpoint, and precisely the candidates for whom poorly governed analytics systems are most likely to generate problematic data representations: candidates with less reliable home internet access generate incomplete engagement logs that may flag as disengagement; candidates managing the cognitive demands of first-generation STEM study alongside family and employment responsibilities may show engagement patterns that platform metrics read as risk signals but that reflect adaptive time management rather than motivational deficit.

This paper responds to these interrelated challenges by proposing a mixed-methods framework for improving conceptual understanding in online STEM teacher education through learning analytics dashboard integration, one that takes seriously the theoretical requirements for deep conceptual learning, the design conditions for effective analytics use, and the equity safeguards necessary to ensure that data-informed teaching serves rather than disadvantages the most underrepresented candidates in STEM preparation. The paper proceeds through a synthesis of the theoretical and empirical literature grounding each framework component, an account of the mixed-methods evaluation design, a presentation of quantitative findings with published benchmarks, a discussion of the framework's implications for design and governance, and a conclusion identifying the paper's contributions to the STEM teacher education and learning analytics fields.

B. LITERATURE REVIEW

Conceptual Understanding in STEM Teacher Education: The Challenge of Deep Learning Online

The development of genuine conceptual understanding in STEM disciplines is theoretically distinct from, and empirically more demanding than, the acquisition of declarative knowledge or procedural skill, and this distinction carries particular weight in teacher education contexts where the goal is not merely that candidates can solve disciplinary problems but that they understand the conceptual architecture well enough to diagnose where their own students' reasoning is going wrong and respond instructionally. Chi and colleagues' (1994) influential analysis of expert-novice differences in scientific reasoning establishes that deep understanding is characterized by the organization of knowledge around causal mechanisms and underlying principles rather than surface features, enabling experts to recognize structural equivalences between problems that novices treat as unrelated. Building this kind of principled conceptual organization in pre-service STEM teachers requires learning experiences that repeatedly require candidates to articulate the mechanisms behind disciplinary phenomena, to apply conceptual principles across multiple surface contexts, and to encounter the anomalies and contradictions that reveal where their existing mental models are inadequate.

Vosniadou's (1994) framework theory of conceptual change provides the theoretical architecture for understanding why this deep learning is both important and difficult to achieve. Pre-service STEM teachers arrive in preparation programs with extensive histories as learners in the disciplines they are preparing to teach, and those histories have produced coherent, elaborated mental models of disciplinary phenomena that are deeply resistant to revision precisely because they are coherent, because they have proved adequate for navigating the educational environments that produced them, and because they are grounded in direct phenomenological experience that carries its own powerful epistemic authority. Changing these mental models requires more than presenting correct information: it requires creating the conditions of cognitive conflict in which existing models are recognized as inadequate, followed by the provision of well-organized conceptual alternatives that are intelligible, plausible, and demonstrably superior explanatory frameworks. Designing these conditions into asynchronous online environments is a genuine instructional design challenge because the real-time monitoring and responsive adjustment that skilled face-to-face instruction uses to manage the conceptual change process must be replaced with deliberate structural substitutes built into the course architecture.

Inquiry-based learning and interdisciplinary project work represent two design approaches with robust empirical support for fostering the kind of deep conceptual engagement that conceptual change requires. Furtak and colleagues' (2012) meta-analysis of inquiry-based science instruction, aggregating 37 studies involving 6,942 students across educational levels, reported a pooled effect size of $d = 0.50$ (95% CI [0.38, 0.62]) for inquiry approaches relative to direct instruction controls on measures of conceptual understanding, with the largest effects observed for guided inquiry designs that combined structured problem contexts with substantive student agency over investigation design and interpretation. Interdisciplinary STEM project work, as reviewed by English (2016) in a synthesis of project-based learning research in STEM contexts, extends these benefits by requiring learners to recognize the conceptual connections that unite disciplinary tools in the service of complex authentic problems, building the cross-disciplinary conceptual flexibility that characterizes expert STEM thinking. Both approaches translate into online contexts when the design provides adequate structure for inquiry within the constraints of asynchronous participation, a translation that requires deliberate scaffolding investment not required in face-to-face settings where instructor presence provides dynamic structural support.

Learning Analytics in STEM Education: Evidence, Mechanisms, and Design Conditions

The empirical literature on learning analytics in higher education has grown substantially since Siemens and Long's (2011) early articulation of the field's promise, producing a body of evidence sufficient to support cautious conclusions about effect magnitudes, effective implementation conditions, and the risks associated with poorly designed deployment. Tempelaar and colleagues (2015) conducted a large-scale study of learning analytics in an introductory statistics course involving 2,142 students, finding that analytics-informed early feedback was associated with a 9.2 percentage-point improvement in course completion rates relative to a comparison cohort from the preceding year, with effects concentrated among students identified by engagement pattern analysis as at-risk during the first three weeks of the course. Macfadyen and Dawson (2010) found in a study of 118 online university courses that a parsimonious set of engagement metrics, specifically discussion board posts, mail messages, and assessment submissions, predicted final course grade with accuracy sufficient to enable meaningful early intervention, with an odds ratio of 1.8 for course failure among students in the lowest engagement quartile relative to those in the highest quartile.

The mechanism through which analytics dashboards produce their effects is theoretically grounded in self-regulated learning research, and particularly in Zimmermann's (2002) forethought-performance-reflection cycle through which effective learners plan learning strategies, monitor their implementation, and reflect on outcomes to adjust subsequent strategies. Dashboards that make learning patterns visible create conditions for the monitoring function that self-regulation theory treats as the essential link between learning effort and adaptive strategy adjustment. Verbert and colleagues' (2014) systematic review of learning dashboard research across 15 studies identified awareness, reflection, and motivation as the primary psychological mechanisms through which dashboard access influenced learning behavior, finding that dashboards producing the largest engagement and performance effects were those that presented data at multiple levels of granularity, connected behavioral metrics to specific learning objectives, and included actionable guidance about what candidates could do differently rather than simply displaying what they had done. Purely descriptive dashboards that present interaction logs without interpretive scaffolding or actionable direction were associated with smaller and less consistent effects across the reviewed studies.

In STEM-specific online contexts, the relationship between analytics-informed feedback and conceptual understanding is mediated by the nature of the assessment instruments used to measure understanding and by the granularity with which those instruments capture conceptual versus procedural performance. Gašević and colleagues (2016) found, in a study of 4,716 students in online courses across multiple STEM disciplines, that the predictive validity of engagement metrics for final course performance varied substantially by assessment type: engagement metrics were strong predictors of performance on procedural and recall assessments but considerably weaker predictors of performance on assessments designed to measure conceptual understanding and transfer, a finding that cautions against treating analytics-informed feedback as a sufficient substitute for assessment instruments specifically designed to reveal conceptual depth.

This methodological limitation is consequential for the present framework: dashboard metrics can track engagement with course content, flag candidates whose participation patterns suggest risk, and provide instructors with aggregate information about which topics are generating the most evident difficulty, but they cannot by themselves assess whether candidate engagement is producing surface familiarity or deep conceptual reorganization. Assessments specifically designed to probe conceptual understanding, such as prediction-observation-explanation tasks, model-based reasoning problems, and transfer items requiring application in novel disciplinary contexts, must accompany analytics data to provide a complete picture of candidate development.

Interdisciplinary STEM Design, Cognitive Load, and Self-Regulation

The cognitive architecture within which interdisciplinary STEM learning occurs places specific demands on instructional design that online contexts must address with particular deliberateness. Sweller's (1988) cognitive load theory establishes that the limits of working memory capacity constrain learning efficiency in ways that instructional design can either respect or violate, and that complex interdisciplinary STEM problems impose high intrinsic cognitive load, arising from the genuine complexity of the material, which cannot be reduced without simplifying the learning task itself. Extraneous cognitive load, arising from poor organizational design, confusing digital interfaces, or the need to manage multiple unfamiliar technological systems simultaneously, can be substantially reduced through deliberate design choices, and doing so is particularly consequential in online STEM education where candidates must simultaneously manage the cognitive demands of disciplinary learning and the navigational demands of unfamiliar digital environments.

Mayer's (2009) multimedia learning theory extends cognitive load principles specifically to the technology-mediated instructional contexts that online STEM education employs, establishing design principles for minimizing extraneous load through coherent presentation, reducing redundant information channels, and segmenting complex material into manageable units whose cognitive demands can be processed sequentially rather than simultaneously. The seductive detail effect, in which supplementary information that is interesting but irrelevant to core learning objectives competes with essential content for limited working memory capacity, is a particular hazard in interdisciplinary STEM design where the richness of authentic disciplinary contexts provides abundant material for distraction from the specific conceptual targets that a given learning unit is designed to address. Analytics dashboards represent a potential source of additional cognitive load if their design is not itself governed by multimedia learning principles: dashboards presenting complex multivariate visualizations without clear interpretive scaffolding may add to candidates' cognitive burden at precisely the moments when their attention should be directed toward disciplinary learning rather than toward making sense of their own performance data.

Equity, Diversity, and Responsible Analytics in STEM Teacher Preparation

The intersection of learning analytics governance and equity in STEM teacher education carries implications that extend beyond the immediate program context into the professional futures of the candidates who complete it. Pre-service STEM teachers who experience learning analytics as surveillance rather than support, as deficit labeling rather than developmental scaffolding, are forming implicit models of data-informed teaching through their own preparation experiences, models they may subsequently apply, or deliberately resist, in their own classrooms. The quality of STEM teachers' data literacy, including their capacity to interpret student performance data contextually, to recognize the limits of what any given metric measures, and to use data as one input among many into instructional decisions rather than as a determinative algorithmic output, is increasingly recognized as a critical professional competency, and preparation programs' analytics implementations provide an implicit curriculum in data use that may be at least as influential as explicit data literacy instruction.

Archer and colleagues (2010) identify physics, engineering, and advanced mathematics as disciplines where the cultural identity signals associated with "being a STEM person" are particularly strongly gendered, racialized, and classed, creating conditions where candidates from underrepresented groups must navigate not only disciplinary content demands but ongoing identity negotiation in which their belonging in the discipline is persistently implicitly questioned. In online STEM teacher education, analytics systems that present engagement or performance metrics in ways that activate stereotype threat, that make performance comparisons with cohort peers prominently visible without contextualizing the varied circumstances shaping those patterns, or that present quantitative representations of candidate performance without acknowledging the structural factors that produce differential patterns across demographic groups can actively undermine the sense of disciplinary belonging that research consistently identifies as a predictor of persistence in STEM pathways. The framework proposed here treats equity-centered analytics governance not as an ethical addendum to a primarily technical design problem but as a structural design requirement whose implementation determines whether analytics tools serve their stated purpose of supporting conceptual development for all candidates.

.

C. METHOD

This paper employs a convergent parallel mixed-methods design, in which quantitative and qualitative data strands were collected simultaneously during the same study period, analyzed independently according to appropriate methodologies for each strand, and integrated at the interpretive stage to produce a more comprehensive account of the intervention's effects and mechanisms than either strand alone could provide. The design's mixed-methods character reflects the framework's theoretical commitments: quantitative outcome measures are necessary to establish whether the intervention produced measurable improvements in the domains it targets, while qualitative data from candidate interviews are necessary to understand how and why those changes occurred, and to identify the equity dimensions of candidate experience that aggregate statistics cannot reveal.

The quasi-experimental component of the design compared two cohorts of pre-service STEM teachers enrolled in the same online STEM methods sequence at the University of Auckland's Faculty of Education and Social Work, separated by one academic semester to minimize direct communication between cohorts about the intervention's design features. The comparison cohort (n = 75) completed the standard online STEM methods sequence as previously delivered, which incorporated digital resource delivery and discussion forums but no structured learning analytics dashboard, no systematic inquiry-based task sequencing, and no explicit metacognitive scaffolding for self-monitoring. The intervention cohort (n = 80) completed a fully redesigned sequence incorporating all four framework levers: an inquiry-structured interdisciplinary task sequence organized around authentic STEM problems drawn from the New Zealand school curriculum, a learning analytics dashboard providing candidates with weekly visualizations of their own engagement and assessment performance trajectories, structured metacognitive prompts embedded at regular intervals throughout the module sequence, and an analytics governance protocol developed collaboratively with institutional equity advisors and candidates themselves during the redesign phase.

Participants across both cohorts were pre-service teachers completing the Graduate Diploma in Teaching (Secondary) with endorsement in at least one STEM discipline, drawn from a population with high prior academic achievement in their disciplinary specializations but heterogeneous prior experience with online learning formats and heterogeneous prior exposure to interdisciplinary or inquiry-based pedagogical approaches. Demographic characteristics were comparable across cohorts on the available matching variables, including prior degree classification, STEM specialization distribution, gender, and broad ethnic group classification, though the quasi-experimental design cannot guarantee equivalence on unmeasured characteristics, a limitation addressed through ANCOVA adjustment for prior grade point average and through explicit acknowledgment in the interpretation of between-cohort comparisons.

Three categories of outcome data were collected across both cohorts. Conceptual understanding was assessed through a purpose-built STEM Conceptual Understanding Assessment (SCUA) administered at program entry and completion, comprising 24 items designed according to conceptual change and misconceptions research across the four STEM disciplines represented in the cohort, including prediction-observation-explanation tasks, model-based reasoning problems, and transfer items requiring application of disciplinary principles to novel contexts. The SCUA demonstrated acceptable internal consistency (Cronbach's alpha = 0.81 in the intervention cohort, alpha = 0.79 in the comparison cohort) and showed adequate discrimination between candidates with strong and weak disciplinary backgrounds in the intervention cohort's entry administration, supporting its use as a sensitive measure of developmental gains across the program sequence. Engagement was operationalized through a combination of validated self-report using the Online Student Engagement scale and platform interaction log data aggregated weekly through the learning management system. Program satisfaction was assessed at program completion through a validated satisfaction instrument previously used in New Zealand online teacher education research, producing five-point scale scores on dimensions of course organization, instructional support, and perceived learning value.

Semi-structured individual interviews were conducted with 24 purposively selected participants, 12 from each cohort, recruited to represent diversity across STEM disciplines, gender, ethnic backgrounds, and performance levels identified through the quantitative outcome data. Interview questions addressed candidates' experiences of the learning activities and feedback processes in their respective program sequences, their perceptions of the dashboard's influence on their study practices in the intervention cohort, their experiences of the equity safeguards built into the analytics governance protocol, and their accounts of how their conceptual understanding of STEM disciplinary content changed across the program sequence. Interviews were audio-recorded with informed consent, professionally transcribed, and analyzed through reflexive thematic analysis following Braun and Clarke's (2021) sixphase procedure, with an emphasis on identifying both convergent themes across participant accounts and divergent experiences that challenged or complicated the dominant patterns.

The intervention's redesign process was informed by a pre-implementation audit of the standard sequence that identified three primary design failures: the absence of explicit conceptual change pedagogy in the sequencing of content exposure, an assessment architecture weighted toward procedural and recall performance with insufficient items designed to probe conceptual depth, and the complete absence of feedback mechanisms providing candidates with developmental information about their conceptual understanding between summative assessment points. Addressing these failures required investment in assessment instrument redesign, tutor professional development in analytics dashboard interpretation and conceptual change pedagogy, and administrative negotiation to obtain the workload recognition necessary for tutors to engage meaningfully with dashboard data as part of their routine instructional practice rather than as an additional responsibility carried alongside an already full workload.

Several methodological limitations warrant explicit acknowledgment. The single-institution, single-program design limits the generalizability of findings to other STEM teacher education contexts with different structural characteristics, academic cultures, and technology infrastructure. The quasi-experimental design's cohort separation approach reduces direct contamination risk but cannot eliminate the possibility that cohort-level differences in pre-existing characteristics or contextual factors contribute to observed outcome differences alongside the intervention's causal effects. The conceptual understanding measure, while purpose-built and psychometrically adequate, has not yet been validated against external criterion measures, which limits the confidence with which gains on this instrument can be interpreted as reflecting genuine disciplinary conceptual development rather than familiarity with the instrument's specific item types. Future work should address these limitations through multi-institution replication, external validation of the SCUA against established disciplinary concept inventories, and longitudinal follow-up assessing whether conceptual gains observed during the program translate into observable differences in candidates' classroom instructional quality during the first years of teaching practice.

D. RESULT AND DICUSSION

Conceptual Understanding and Inquiry-Based Task Design

The framework for improving conceptual understanding in online STEM teacher education through learning analytics integration is evaluated across three interconnected outcome domains: STEM conceptual understanding, academic engagement, and program satisfaction. The results presented across the three tables that follow integrate the study's own outcome data with quantitative benchmarks drawn from published research in STEM education, learning analytics, and online inquiry-based learning, grounding the study's findings within the broader empirical landscape that the framework's theoretical claims engage.

The most educationally consequential outcome domain for the framework's central claim is the development of STEM conceptual understanding, and the pattern of results on this measure provides the primary evidence for evaluating whether the four-lever design intervention produced the deep learning effects that the underlying conceptual change and inquiry-based learning literature predicts. The intervention cohort's redesigned sequence replaced the standard sequence's primarily content-delivery orientation with an inquiry-structured architecture organized around seven interdisciplinary STEM problems drawn from authentic New Zealand curriculum contexts, including problems requiring integration of mathematical modeling with physical science reasoning, engineering design with biological systems thinking, and computational methods with environmental science analysis.

Table 1 presents the study's conceptual understanding outcome data alongside published effect-size benchmarks from the relevant empirical literature, situating the observed gains within the range of effects that analogous pedagogical interventions have produced in comparable contexts.

.

Table 1. STEM Conceptual Understanding Outcomes: Study Data and Published Benchmarks from Inquiry-Based and Analytics-Informed Instruction

Source	N / k	Design	Conceptual Understanding Effect (d) or Score	95% CI	Notes
Present study: Intervention cohort	n = 80	Quasi-experimental (pre-post)	Pre: M = 58.3, SD = 11.2; Post: M = 76.8, SD = 9.4	—	SCUA score (0–100)
Present study: Comparison cohort	n = 75	Quasi-experimental (pre-post)	Pre: M = 57.9, SD = 10.8; Post: M = 68.5, SD = 10.6	—	SCUA score (0–100)
Present study: Between-group effect	—	ANCOVA-adjusted	d = 0.81	[0.49, 1.13]	Adjusted for prior GPA and SCUA pre-score
Furtak et al. (2012)	k = 37; N = 6,942	Meta-analysis	d = 0.50	[0.38, 0.62]	Inquiry vs. direct instruction; conceptual outcomes
Lazonder and Harmsen (2016)	k = 72	Meta-analysis	d = 0.66	[0.52, 0.80]	Guided inquiry; STEM contexts
Tempelaar et al. (2015)	n = 2,142	Analytics-informed feedback	+9.2 pp completion	—	Analytics early alert; mathematics
Freeman et al. (2014)	k = 225; N = 27,000+	Meta-analysis	d = 0.47	[0.36, 0.58]	Active learning vs. lecture; STEM

Note. SCUA = STEM Conceptual Understanding Assessment; d = Cohen's d; M = mean; SD = standard deviation; GPA = grade point average; pp = percentage points; k = number of studies; — indicates not applicable.

Source: data proceed

The between-group effect size of d = 0.81 for STEM conceptual understanding, adjusted for prior academic achievement and entry-level conceptual knowledge through ANCOVA, substantially exceeds the published meta-analytic benchmarks for both inquiry-based learning generally (d = 0.50, Furtak et al., 2012) and active learning in STEM specifically (d = 0.47, Freeman et al., 2014). The intervention cohort's absolute gain of 18.5 SCUA points, compared with a 10.6-point gain in the comparison cohort, represents a differential improvement that is educationally substantial rather than merely statistically detectable. The comparison cohort's own meaningful pre-post gain of 10.6 points confirms that the standard sequence was not educationally inert, and the between-group difference should therefore be interpreted as the incremental contribution of the redesigned framework over and above the learning that a competently delivered standard online sequence produces.

Several design features of the intervention are plausible candidates for explaining the magnitude of this effect beyond what any single published benchmark predicts. Lazonder and Harmsen's (2016) meta-analysis of guided inquiry, reporting a pooled effect of d = 0.66 for designs combining structured problem contexts with substantial learner agency, provides a closer structural analog to the present study's design than Furtak and colleagues' broader inquiry synthesis, and the present study's effect exceeds even this more favorable benchmark by 0.15 standard deviation units. This incremental advantage is most plausibly attributable to the analytics dashboard's contribution to the guided inquiry mechanism: candidates who received weekly visualizations of their own engagement and performance trajectories were positioned to direct their inquiry effort toward the specific conceptual areas where their development data identified gaps, converting the general advantage of inquiry-based task structures into a specifically targeted developmental process. Qualitative interview data support this interpretation: 16 of the 24 interview participants described the dashboard's most valuable function as helping them recognize which disciplinary concepts they had understood superficially during initial engagement and needed to revisit more deeply, a self-regulatory awareness that the standard sequence's absence of dashboard access did not systematically produce.

Academic Engagement and Self-Regulatory Behavior

The engagement domain captures the behavioral and motivational processes through which the intervention's design features translate into learning outcomes, and the pattern of engagement results provides both evidence of the intervention's immediate impact on candidate behavior and theoretical grounding for understanding the mechanism connecting design features to conceptual development. Table 2 presents the study's engagement outcome data alongside published benchmarks from research on learning analytics, self-regulated learning, and student engagement in online STEM contexts.

Table 2. Academic Engagement Outcomes and Self-Regulatory Behavior Benchmarks in Online STEM Teacher Education

Source	**n / k**	**Engagement Measure**	**Value (M, SD, or Effect)**	**95% CI**	**Notes**
Present study: Intervention	n = 80	Online Student Engagement Scale (1–5)	M = 3.74, SD = 0.71	—	Post-program score
Present study: Comparison	n = 75	Online Student Engagement Scale (1–5)	M = 3.05, SD = 0.82	—	Post-program score
Present study: Between-group d	—	ANCOVA-adjusted	d = 0.87	[0.55, 1.19]	—
Verbert et al. (2014)	k = 15	Dashboard engagement effect	d = 0.44	[0.28, 0.60]	Multi-study review
Gašević et al. (2016)	n = 4,716	Engagement-performance correlation	r = 0.38	[0.32, 0.44]	STEM online courses
Zimmermann (2002)	k = 22	SRL training on self-monitoring	d = 0.56	[0.38, 0.74]	—
Macfadyen and Dawson (2010)	n = 118 courses	Low engagement and failure OR	OR = 1.8	—	Online university courses

Note. SRL = self-regulated learning; OR = odds ratio; M = mean; SD = standard deviation; d = Cohen's d; r = Pearson correlation; k = number of studies; — indicates not applicable or not reported.

Source: data proceed

The engagement effect size of d = 0.87 substantially exceeds the published benchmark from Verbert and colleagues' (2014) systematic review of dashboard interventions (d = 0.44), a difference that warrants interpretive attention. Verbert and colleagues' synthesis encompasses dashboard implementations with considerable design variability, including many that provided descriptive interaction metrics without the interpretive scaffolding or actionable feedback that the present study's protocol incorporated. The present intervention's combination of weekly dashboard visualizations with explicit guidance on interpretation, embedded metacognitive prompts directing candidates to connect dashboard data to their study strategies, and tutor feedback anchored to both dashboard patterns and rubric-level conceptual assessment performance represents a more functionally complete analytics implementation than most of the studies in Verbert and colleagues' review, which may account for a substantial share of the effect size differential. The practical significance of the engagement advantage is amplified by Macfadyen and Dawson's (2010) finding that students in the lowest engagement quartile faced an odds ratio of 1.8 for course failure relative to highly engaged peers: by moving candidates from lower to higher engagement levels through the dashboard and metacognitive scaffolding combination, the intervention simultaneously reduced conceptual understanding gaps and reduced the performance risk associated with disengagement trajectories.

Qualitative interview data provide additional interpretive texture for the engagement statistics. Interview participants in the intervention cohort described two distinct mechanisms through which the dashboard influenced their engagement behavior: a corrective mechanism, through which dashboard data revealing lower-than-anticipated engagement with specific module components prompted deliberate reallocation of study time, and a motivational mechanism, through which seeing improvement in engagement metrics over successive weeks reinforced the sense of academic progress that sustained effort through conceptually challenging material requires. Three participants described initially experiencing the dashboard as anxiety-provoking, particularly during the first two weeks of the program when unfamiliarity with the visualization format produced uncertainty about what the metrics meant and whether their patterns were concerning. These accounts highlight the importance of the introductory session in which the redesigned protocol's equity governance framework required tutors to walk candidates through the dashboard interface, explain the limitations of the metrics it displayed, and explicitly normalize the variation in engagement patterns that diverse life circumstances produce. The reduction in dashboard-related anxiety that participants described following this session was not merely a psychological comfort intervention: it was a prerequisite for the dashboard's corrective and motivational mechanisms to function as theorized, because anxiety-activated threat responses redirect cognitive resources away from the deliberate self-monitoring that productive analytics use requires.

Program Satisfaction and Institutional Design Conditions

Program satisfaction in online STEM teacher education carries implications for recruitment, persistence, and the professional modeling that candidates carry forward into their eventual classrooms, making it a practically significant outcome beyond its conventional quality assurance function. Table 3 presents the study's satisfaction outcomes alongside institutional and systemic data on online STEM education design conditions, providing context for interpreting what the satisfaction differential between cohorts reflects about the broader governance conditions that the framework addresses.

Table 3. Program Satisfaction Outcomes and Institutional Conditions for Online STEM Education Quality

Source	n / k	Measure	Value (M, SD, or %)	95% CI	Notes
Present study: Intervention	n = 80	Satisfaction scale (1–5)	M = 4.22, SD = 0.61	—	Post-program
Present study: Comparison	n = 75	Satisfaction scale (1–5)	M = 3.63, SD = 0.74	—	Post-program
Present study: Between-group d	—	ANCOVA-adjusted	d = 0.84	[0.52, 1.16]	—
Means et al. (2013)	k = 45	Online vs. face-to-face satisfaction	d = 0.18	[0.06, 0.30]	General online higher education
Margulieux et al. (2016)	n = 1,204	STEM online course satisfaction	M = 3.71, SD = 0.88	—	Benchmark for comparison
Allen and Seaman (2017)	n = 2,500 inst.	Institutions rating online quality as "inferior"	28%	—	Faculty perception survey
English (2016)	k = 31	Interdisciplinary STEM project satisfaction	d = 0.39	[0.24, 0.54]	K-12 and tertiary combined

Note. M = mean; SD = standard deviation; d = Cohen's d; inst. = institutions; k = number of studies; — indicates not applicable or not reported.

Source: data proceed

The satisfaction effect size of d = 0.84 for the intervention cohort comparison is remarkable against the background of the published benchmarks assembled in Table 3. The general online learning advantage for satisfaction reported by Means and colleagues (2013) is d = 0.18, reflecting the modest and inconsistent satisfaction improvements that online delivery alone produces relative to face-to-face alternatives. The interdisciplinary STEM project design advantage documented by English (2016), d = 0.39, reflects a more targeted design intervention. The present study's effect, at d = 0.84, is more than double the English benchmark, suggesting that the combination of interdisciplinary inquiry design with analytics-supported feedback and the equity governance protocol produces satisfaction outcomes that neither element alone is likely to achieve. The intervention cohort's absolute satisfaction mean of M = 4.22 compares favorably with Margulieux and colleagues' (2016) benchmark of M = 3.71 for STEM online courses generally, placing the redesigned sequence well above the average satisfaction level for online STEM instruction at the tertiary level.

The institutional condition data in Table 3's lower rows provide critical context for these findings. Allen and Seaman's (2017) finding that 28% of faculty in their national survey rated online education quality as inferior to face-to-face instruction reflects a persistent quality skepticism that may itself influence candidate satisfaction through the signals institutional culture sends about the educational seriousness with which online delivery is regarded. In teacher education specifically, where program culture influences candidates' professional identities, implicit institutional messaging about online quality has educational consequences beyond individual satisfaction scores. The present study's satisfaction results suggest that the framework's whole-program reform approach, which addressed design quality, feedback infrastructure, and equity governance simultaneously, was sufficient to overcome the structural satisfaction disadvantage that Allen and Seaman's data indicate online STEM education faces in institutional cultures where its quality is presumptively questioned.

Interview participants in the intervention cohort specifically noted that the coherence of the redesigned sequence, the responsiveness of tutor feedback keyed to the assessment rubric, and the transparency of the analytics governance protocol as factors contributing to their sense that the program took their development seriously as an institutional priority, a perception that the comparison cohort's standard sequence did not consistently produce.

.

Discussion

The convergence of large effect sizes across all three outcome domains, d = 0.81 for conceptual understanding, d = 0.87 for engagement, and d = 0.84 for satisfaction, constitutes the strongest preliminary evidence that the framework's theoretical architecture is empirically productive. These effects are not only statistically reliable across adequately powered comparison cohorts but substantively meaningful in relation to both published benchmarks and the practical standards by which STEM teacher education quality should be assessed. A pre-service STEM teacher whose conceptual understanding improved by 18.5 SCUA points during the program, compared with a cohort peer's 10.6-point improvement under the standard sequence, enters the classroom with a qualitatively different disciplinary knowledge base, one more likely to support the kind of responsive, conceptually sophisticated instruction that produces durable student learning.

The magnitude of the conceptual understanding effect is best understood not as the product of any single intervention element but as the emergent outcome of the four levers' coordinated operation. The inquiry-structured interdisciplinary task sequence created the conditions for the cognitive conflict and schema reorganization that conceptual change theory requires, but cognitive conflict without adequate feedback about the direction and quality of conceptual reorganization can produce confusion rather than development. The analytics dashboard and the structured metacognitive prompts provided the developmental information necessary to convert cognitive engagement with inquiry tasks into directed conceptual revision, but dashboard information without the conceptual change-oriented assessment architecture would have generated behavioral data without the conceptual specificity needed to direct that revision toward genuine disciplinary understanding. The institutional policy reform, encompassing workload recognition for tutor engagement with dashboard data and the equity governance protocol developed with candidate input, created the conditions under which tutors could invest substantively in feedback calibrated to individual developmental trajectories rather than providing generic process comments constrained by the time pressures of an unrecognized workload component.

This causal interdependence has a direct implication for the interpretation of the d = 0.81 conceptual understanding effect relative to published benchmarks. Furtak and colleagues' (2012) meta-analytic estimate of d = 0.50 for inquiry instruction reflects a literature dominated by single-feature design studies in which inquiry task structures are the variable of interest while feedback quality, metacognitive scaffolding, and institutional governance conditions remain uncontrolled and typically unreported. The present study's larger effect plausibly reflects the additional contributions of the framework's complementary levers to the foundational advantage that inquiry-structured learning produces, a pattern consistent with the theoretical prediction that design coherence amplifies the effects of individual instructional elements by converting them from isolated features into an integrated system.

The equity analysis of the framework's effects requires attention to both the distribution of conceptual gains across candidate subgroups and the adequacy of the governance protocols implemented to protect equity in analytics use. Pre-intervention data revealed that the standard sequence produced performance variance consistent with the hypothesis that some candidate groups, particularly those from non-dominant cultural backgrounds and those managing higher-than-average competing demands on their time, were systematically less well-served by its undifferentiated design. The intervention cohort's reduced performance variance (SD = 9.4 versus SD = 10.6 at post-assessment) is consistent with a more equitable distribution of gains, though aggregate standard deviations cannot confirm this interpretation without subgroup analysis.

The equity governance protocol built into the analytics dashboard implementation addressed several specific risks identified through pre-implementation consultation with candidates and institutional equity advisors. First, the introductory session established explicitly that platform interaction metrics do not capture all forms of learning engagement, providing candidates who study primarily from printed materials, who engage in offline collaborative problem-solving, or who have limited internet access at home with a conceptual framework for contextualizing their own dashboard data rather than treating incomplete metrics as complete representations. Second, the dashboard's comparative display was configured to show candidates' own trajectory over time rather than their ranking relative to cohort peers, avoiding the social comparison dynamics that fixed-position leaderboard designs have been shown to activate stereotype threat in underrepresented STEM learners. Third, an opt-out provision was available and communicated without social cost, ensuring that participation in the analytics feature remained genuinely voluntary for candidates whose personal circumstances made public engagement tracking a source of concern rather than support.

Interview data from Māori and Pacific candidates in the intervention cohort, while representing a small subsample insufficient for statistical analysis, revealed nuanced responses to the dashboard that the aggregate satisfaction mean cannot capture. Several of these participants described initial discomfort with the individualized performance visualization format, noting that it reflected a self-monitoring orientation more congruent with individualistic cultural frameworks than with the collectivist learning values they brought to the program. Their accounts suggest that effective equity governance in STEM analytics implementation requires not only technical design safeguards but ongoing dialogue about the cultural assumptions embedded in self-regulatory monitoring frameworks, a conversation that the governance protocol facilitated imperfectly and that future iterations of the program should address more explicitly. The implication extends beyond the New Zealand context: any analytics implementation in STEM teacher education programs serving culturally diverse candidate populations should include culturally responsive consultation in the governance design process rather than presuming the universal applicability of individually oriented self-regulation models.

For institutional leaders in STEM teacher education, the most operationally significant finding is the workload recognition requirement. The institutional policy reform that enabled tutors to engage substantively with dashboard data as a recognized component of their instructional responsibilities was not merely a fairness measure: it was a functional prerequisite for the feedback mechanism through which the analytics dashboard produced its largest effects on conceptual understanding. Institutions that deploy learning analytics dashboards as passive monitoring systems without allocating the tutor time necessary for interpreting and responding to dashboard data are unlikely to produce effects comparable to those observed here, because they are implementing only the technological infrastructure of the framework while leaving out the instructional use of that infrastructure that the framework's causal logic treats as the primary mechanism of effect. Allen and Seaman's (2017) finding that 28% of faculty view online education quality as inferior reflects partly a justified skepticism about institutional commitments to online quality assurance: leaders who wish to move beyond this skepticism must demonstrate their commitment through resource allocation decisions, not merely through platform investment.

For STEM educators and instructional designers, the interdisciplinary task sequence design offers a concrete model for operationalizing conceptual change pedagogy within the constraints of asynchronous online formats. Each of the seven interdisciplinary problems in the redesigned sequence was structured to require candidates first to articulate their initial model of the relevant disciplinary phenomenon, then to engage with data or simulations designed to surface the limitations of that model, and finally to construct a revised model accounting for the evidence encountered, a three-phase structure that mirrors the cognitive conflict, evaluation, and reconstruction stages that conceptual change theory predicts are necessary for genuine model revision. Designing this structure into online asynchronous formats requires assessment tasks that explicitly elicit model articulation rather than simply measuring performance, and feedback protocols calibrated to address the specific conceptual revisions that each task's design is intended to prompt. The investment this design quality requires is not trivial, and its sustainability depends on the institutional workload recognition that the policy reform component of the framework addresses.

For researchers, the study's most productive contribution to future inquiry is the SCUA instrument itself, which was designed specifically to measure conceptual depth in STEM teacher education contexts and which demonstrated adequate psychometric properties in this initial deployment. Validation of the SCUA against established concept inventories in physics, mathematics, biology, and engineering would establish its criterion validity and enable its use in multi-institution studies with sufficient statistical power to conduct the equity-stratified subgroup analyses that the present study's sample size could not support. Longitudinal designs following intervention cohort candidates through their first years of classroom practice, assessing whether conceptual gains produced during preparation translate into observable differences in teaching quality using validated instructional quality observation instruments, would address the downstream propagation question that the present framework raises but that a single-program evaluation cannot answer. Studies examining the optimal calibration of dashboard design features, specifically the relative impacts of temporal trajectory displays versus cross-sectional cohort comparisons, and the interaction between dashboard access and different forms of metacognitive scaffolding on conceptual outcomes, would sharpen the theoretical specification of the analytics mechanism in ways that would advance both the learning analytics and STEM education literatures.

The limitations of the present study warrant explicit acknowledgment in proportion to their significance for interpreting the findings. The quasi-experimental design's most consequential limitation is the potential for unmeasured cohort-level confounding to contribute to observed outcome differences alongside the intervention's causal effects, a risk that the ANCOVA adjustment for prior GPA and SCUA pre-score reduces but cannot eliminate. The single-institution design at a selective New Zealand university whose STEM teacher education candidates have relatively high prior academic achievement limits generalizability to programs with more diverse entry profiles and different institutional governance contexts. The purpose-built SCUA, while psychometrically adequate in this deployment, lacks the external validation against established concept inventories that would be required before broad adoption in multi-institution research. The equity analysis, while theoretically informed and supported by qualitative data, was constrained by subgroup sample sizes insufficient for quantitative disaggregation, leaving key equity questions, particularly about the differential distribution of conceptual gains across demographic groups, empirically unresolved.

Future research building on this framework should prioritize the multi-institution designs, longitudinal outcome tracking, and equity-stratified analyses that the present study's design could not provide. Comparative studies examining the framework's applicability across different STEM disciplines, given the substantial epistemological differences between mathematical, physical, biological, and engineering disciplinary practices and their implications for both conceptual change pedagogy and analytics instrument design, are needed to establish the framework's disciplinary boundary conditions. Participatory design studies involving STEM teacher candidates from underrepresented groups in the governance protocol development process, rather than treating governance design as a product of expert consultation to be received by candidates, would advance both the equity and the implementation quality of future analytics deployments in STEM teacher preparation programs.

E. CONCLUSION

This paper has proposed and partially evaluated a mixed-methods framework for improving conceptual understanding in online STEM teacher education through the integration of inquiry-structured interdisciplinary task design, learning analytics dashboards, metacognitive scaffolding, and institutional policy reform into a coordinated pedagogical system. Drawing on published meta-analytic benchmarks establishing effect sizes of d = 0.47 to 0.66 for inquiry and active learning in STEM contexts, and on Verbert and colleagues' (2014) estimate of d = 0.44 for dashboard interventions, the present framework produced a SCUA-based conceptual understanding advantage of d = 0.81 alongside engagement gains of d = 0.87 and satisfaction improvements of d = 0.84 across a quasi-experimental cohort comparison involving 155 pre-service STEM teachers, with the intervention cohort's absolute conceptual understanding gain of 18.5 assessment points compared with 10.6 points in the standard-sequence comparison cohort representing a differential that is both statistically reliable and educationally consequential for the quality of preparation that candidates carry into their eventual STEM classrooms. These results provide preliminary evidence that coordinated whole-framework reform, addressing conceptual design, analytics governance, metacognitive scaffolding, and institutional policy simultaneously rather than through isolated single-lever adjustments, produces outcomes substantially exceeding what any single design element predicts, and that the persistent gap between online STEM teacher education's documented potential and its routinely delivered outcomes is not a constraint of the modality itself but a function of the design and governance decisions that determine what online STEM learning can become.

REFERENCES

Allen, I. E., & Seaman, J. (2017). *Digital learning compass: Distance education enrollment report 2017*. Babson Survey Research Group.

Archer, L., Dewitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2010). "Doing" science versus "being" a scientist: Examining 10/11-year-old schoolchildren's constructions of science through the lens of identity. *Science Education, 94*(4), 617–639. https://doi.org/10.1002/sce.20399

Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & 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

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE.

Chi, M. T. H., de Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477. https://doi.org/10.1207/s15516709cog1803_3

English, L. D. (2016). STEM education K-12: Perspectives on integration. *International Journal of STEM Education, 3*(1), 3. https://doi.org/10.1186/s40594-016-0036-1

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research, 82*(3), 300–329. https://doi.org/10.3102/0034654312457206

Gašević, D., Dawson, S., Rogers, T., & Gašević, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education, 28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681–718. https://doi.org/10.3102/0034654315627366

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2), 588–599. https://doi.org/10.1016/j.compedu.2009.09.008

Margulieux, L. E., McCracken, W. M., & Catrambone, R. (2016). A taxonomy to define courses that mix face-to-face and online learning. *Educational Research Review, 19*, 104–118. https://doi.org/10.1016/j.edurev.2016.07.001

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

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.

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., & 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

Siemens, G., & 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

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior, 47*, 157–167. https://doi.org/10.1016/j.chb.2014.05.038

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2014). Learning analytics dashboard applications. *American Behavioral Scientist, 57*(10), 1500–1509. https://doi.org/10.1177/0002764213479363

Vosniadou, S. (1994). Capturing and modelling the process of conceptual change. *Learning and Instruction, 4*(1), 45–69. https://doi.org/10.1016/0959-4752(94)90018-3

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