Human-Centred Learning Analytics for Sustained STEM Inquiry:

Designing and Validating Feedback Systems That Work for Students and Teachers

PhD Proposal
Richard Pedley, 2026

Contents

1. Abstract

Sustained inquiry is among the most powerful forms of STEM learning, but it is also among the most demanding to teach and to assess well. Students working on real problems over weeks and months must learn to connect claims to evidence, compare systematically across conditions, revise their thinking when new data challenges earlier interpretations, and express appropriate uncertainty about what their observations can and cannot support. These are sophisticated epistemic reasoning practices, and they do not develop reliably without sustained scaffolding and timely feedback. At the same time, teachers managing inquiry classrooms face heavy orchestration and feedback workloads with few tools to help them identify where their attention will matter most. The result is a persistent tension: the learning environments that demand the most from students also demand the most from teachers, and both sides often lack the support they need.

This thesis investigates whether human-centred learning analytics can validly and sustainably support epistemic reasoning in sustained school inquiry. It asks three linked questions: whether trace-based indicators can validly represent the quality of epistemic reasoning during sustained inquiry, how narrative feedback and teacher orchestration dashboards are interpreted and acted upon over time, and what socio-technical conditions enable analytics-enhanced inquiry to be sustained under the real constraints of classroom practice.

Situated within the Kōkiri Lab inquiry programme at Whanganui Intermediate School, the research employs a design-based methodology structured across three interlinked studies. Two contrasting community science contexts, an aquaponics system (AwaKai) and an automated ecological monitoring system (MaramaTrap), function as epistemic substrates generating sustained, uncertainty-laden inquiry data. These contexts are not treated as the research contribution. They are authentic environments within which epistemic practices can be examined.

Study 1 co-designs task-aligned documentation and analytics with teachers and students to articulate observable indicators of epistemic reasoning. Study 2 develops and validates trace-based indicators through convergent and discriminant validity strategies, triangulating coded artefacts, platform traces, and qualitative reasoning evidence across contrasting epistemic contexts. Study 3 investigates feedback uptake, teacher orchestration, and governance conditions shaping adoption, including bounded exploration of AI-mediated micro-scaffolds introduced in Year 3.

The design architecture is informed by a national stakeholder survey of 189 educators, structured contextual analyses, and an ecosystem of community and institutional partnerships. These function as participatory constraint-mapping and institutional viability modelling that shape every design decision the analytics system embodies.

The thesis contributes a validated framework for identifying epistemic reasoning in sustained inquiry with transparent reporting of detection limits and failure modes, evidence regarding how narrative feedback and teacher orchestration mediate analytic impact over time, and empirically grounded design principles for implementing learning analytics as sustainable socio-technical systems in school contexts. Because analytics indicators can create seductive but misleading proxies, this thesis treats weak validity as an informative outcome. Indicator limits, confounds, and failure modes are reported with the same prominence as successful indicators. The thesis is designed to produce publishable findings even under weak indicator validity, because mapping detection limits and failure modes in sustained K to 12 inquiry is itself a missing empirical contribution.

2. Introduction

2.1 The Classroom Problem

Science education faces a persistent challenge that shapes both classroom practice and student outcomes. Authentic scientific practice involves genuine uncertainty, contested interpretations, and the ongoing need to justify claims with imperfect evidence. Yet the practical realities of schooling, including time constraints, assessment pressures, curriculum coverage demands, and variable teacher confidence in science, consistently push instruction toward activities with predictable outcomes and clear right answers (Chinn and Malhotra, 2002; Crawford, 2014). Students complete investigations where results are known in advance, analyse data cleaned of ambiguity, and draw conclusions their teachers have already determined. The epistemic work that defines actual science, evaluating competing interpretations, reasoning under uncertainty, revising claims in response to anomalous evidence, rarely features in school science as it is commonly practised (Manz, 2015).

This matters because epistemic practices are not decorative features of scientific work. They are the substance of it. When students engage only with simplified, certain knowledge, they develop misconceptions about how knowledge is actually produced and miss opportunities to develop reasoning capabilities that transfer across contexts and disciplines (Sandoval, 2005). The challenge is particularly acute at the intermediate level, where learners are developmentally ready for more sophisticated reasoning but seldom encounter learning environments that genuinely demand it. Teachers at this level often want to run more ambitious inquiry but lack the time, confidence, or tools to manage the feedback and assessment complexity that authentic inquiry creates.

A related and equally pressing problem concerns feedback itself. In a class of thirty 12-year-olds working with living systems, ambiguity is not an abstract concept. It is a daily management reality. When students are generating observations, claims, comparisons, and reflections across weeks and months, teachers cannot realistically read every entry in detail, monitor every group simultaneously, or provide individualised feedback on reasoning quality at the pace that learning requires. Teachers need tools that help them decide quickly where their attention will matter most, which students are struggling with evidence use, which groups have stalled in their reasoning, and where a well-timed question or redirection could shift the trajectory of an inquiry. At the same time, students need feedback that supports them to notice gaps in their own reasoning and take informed action, not feedback that merely reports metrics, assigns scores, or tells them what to think (Winstone and Carless, 2019; Carless and Boud, 2018).

AI tools are increasingly available in educational settings, but their integration into sustained inquiry raises questions that extend beyond technical capability. When an AI system provides feedback on student reasoning, it becomes a mediator of epistemic practice. It shapes what students attend to, what they consider important, and how they evaluate the quality of their own claims. If that mediation is poorly designed, it risks encouraging epistemic outsourcing, where students defer to AI-generated suggestions rather than exercising their own interpretive judgement, or flattening the very complexity that makes authentic inquiry valuable. The design of feedback in inquiry contexts, whether human-generated or AI-supported, is therefore not a technical problem to be solved by better algorithms. It is a pedagogical and epistemological design problem with direct implications for how learners develop the capacity to reason with evidence (Knight, Buckingham Shum, and Littleton, 2014).

These are not abstract problems drawn from the literature. They are daily realities of intermediate classrooms where inquiry is attempted seriously. They are the problems that motivated this research.

2.2 Origins of the Research Questions

My previous research in ecology examined whether biological indicators, specifically spider assemblages, could serve as reliable signals of underlying soil health conditions under different pastoral grazing regimes. That experience shaped my approach to the present study in ways I did not initially expect. In both ecological and educational systems, indicators promise to simplify complex, dynamic processes into measurable signals that support decision-making. In both cases, the central question is not whether measurement is possible but whether the resulting signals genuinely represent the underlying phenomena they claim to capture, or whether they create misleading proxies that flatten important complexity. A soil health index that responds primarily to moisture rather than biological condition may look useful but guide poor decisions. A learning analytics indicator that responds primarily to word count rather than reasoning quality may look informative but misrepresent what students are actually doing. Ecological training makes you wary of seductive indices. The methodological discipline required to distinguish genuine signal from artefact is similar in both domains, and my training in ecological indicator validation directly informed the construct validity framework adopted in this thesis.

There is a further parallel worth noting. Ecological indicators drift. Sensitivity changes with context, season, and sampling effort. Unintended consequences emerge when indicators become targets, when management decisions optimise for the index rather than the underlying condition the index was supposed to represent. These dynamics map directly onto the consequential validity concerns and adoption challenges that this thesis investigates. An analytics indicator that students learn to game, or that teachers use as a sorting mechanism rather than a teaching tool, has failed in the same way that a soil health index fails when farmers optimise for the metric rather than the soil.

Working as an intermediate teacher at Whanganui Intermediate School, I began exploring whether analytics tools could reduce feedback burden during sustained inquiry while maintaining the integrity of student reasoning. I was running environmental inquiry projects where students monitored living systems over weeks, collected biodiversity data, and attempted to build interpretations from accumulated evidence. The inquiry was authentic and engaging, but the feedback demands were substantial. I could not provide timely, individualised feedback on reasoning quality to every student, and the tools available to me, learning management systems designed for content delivery and compliance tracking, captured almost nothing about whether students were actually linking claims to evidence, comparing systematically, or revising their thinking in response to new data.

As part of the STEM inquiry programme's development, I have planned two structured pilot substrates, aquaponics monitoring and light-trap biodiversity monitoring, to be used as the sustained data-generating contexts for this research. These pilots are scheduled to begin following a Science Teaching Leadership Programme placement in 2026 and are not yet completed at the time of writing. The rationale for selecting them is that they exert different epistemic pressures and therefore stress-test different aspects of the analytics design.

Aquaponics foregrounds delayed systems reasoning and cumulative interpretation across time: students must track a living system where changes unfold over days and weeks, where causation is ambiguous because multiple variables interact simultaneously, and where interpretation must be built from accumulated observation rather than single experiments.

Light-trap biodiversity monitoring foregrounds probabilistic classification reasoning and evidence evaluation under uncertainty: students must classify specimens using both traditional methods and AI-assisted identification tools where classification confidence varies, data quality depends on conditions students cannot fully control, and the need to distinguish what the data can support from what it cannot is constant. The biodiversity context is particularly relevant to questions about human-AI epistemic interaction, because students must evaluate when automated classification tools are reliable and when they are not.

Formative observations from programme design work and earlier classroom inquiry activities suggest several areas where inquiry support breaks down in real classroom time: students often record observations without linking them to specific data, comparisons are made loosely rather than systematically, and uncertainty tends to be expressed as either blanket confidence or blanket hedging. These observations inform hypotheses that the planned pilots will test explicitly. The most pressing practical constraint was feedback. Even with strong student engagement, the time required to notice reasoning issues early and respond in ways that move learning forward does not scale well in an intermediate classroom. Those constraints are what pushed me toward a learning analytics approach that is explicitly designed to be interpretable, bounded, and usable during day-to-day teaching.

The more I investigated the learning analytics literature, the more I encountered two persistent problems that defined the scope of this thesis. The first was a measurement problem. Learning analytics has developed sophisticated capacity for modelling engagement and predicting performance, but it has weak validated indicators for the quality of epistemic reasoning in sustained school inquiry. Most indicators focus on behavioural proxies: time on task, click patterns, resource access, submission timing (Gasevic, Dawson, and Siemens, 2015).

Very few systems attempt to capture whether claims are grounded in specific evidence, whether comparisons are systematic, whether revisions reflect genuine reconsideration rather than surface correction, or whether students calibrate their certainty appropriately. Where inquiry-focused indicators have been proposed, they are rarely validated against human judgement of artefact quality, and almost never in sustained school contexts where learners return to the same questions over weeks and months.

The second was an adoption and mediation problem. Even well-designed analytics systems are frequently abandoned when they add work without clearly reducing burden or improving outcomes (Tsai, Whitelock-Wainwright, and Gasevic, 2020). Feedback literacy research has established that feedback only influences practice when learners notice it, interpret what it means, evaluate its relevance, decide whether and how to act, take action, and evaluate the result (Carless and Boud, 2018). Each step in that chain can fail, and in busy classrooms it often does.

When AI tools are introduced as additional mediators of feedback, the interaction becomes more complex still. Most feedback analytics research examines single assessment events or course-level dashboards in higher education. Very limited work investigates how analytics-driven feedback functions during sustained inquiry where understanding builds cumulatively, where the feedback concerns reasoning processes rather than grades, and where students are 11 to 13 years old. On the teacher side, the conditions under which analytics-enhanced inquiry can be sustained, including orchestration routines, workload impact, trust in system outputs, and institutional support, remain undertheorised and underinvestigated.

These two problems, measurement and adoption, define the niche for this thesis. They emerged from classroom practice rather than from a systematic gap analysis, though they are well supported by the literature reviewed in Section 3.

2.3 What This Project Will Do

This doctoral project will attempt to design and validate human-centred learning analytics indicators for epistemic reasoning quality in sustained school inquiry, and study how narrative feedback and teacher orchestration dashboards are interpreted, taken up, and sustained in practice. The research centres on a digital inquiry platform deployed within a school-based STEM inquiry programme at Whanganui Intermediate School in Aotearoa New Zealand.

The two inquiry substrates are deliberately selected as epistemic stress tests for the analytics system. Aquaponics challenges the system to detect revision depth and longitudinal evidence linkage in contexts where causation is delayed and ambiguous. Biodiversity monitoring challenges the system to detect calibrated uncertainty and comparative reasoning under variable data quality with tool-mediated classification. Together they test whether analytics indicators generalise across contexts or require context-specific calibration, a question with direct implications for whether school-based analytics can be practically useful beyond the specific environment in which they were developed.

The project proceeds through three linked studies structured around a three-layer architecture.

The first layer concerns construct articulation and validation. Study 1 co-designs the analytics system with students, teachers, and community stakeholders, producing a design model, a constraints taxonomy spanning transparency, governance, explainability, workload, and cultural responsiveness, and governance mechanisms that function as testable adoption levers. Study 2 develops and validates trace-based indicators against human-coded artefacts across both inquiry contexts, testing what the indicators detect, what they miss, what requires context-specific calibration, and what remains dependent on human judgement.

The second layer concerns mediation and uptake. Study 3 examines the full socio-technical feedback cycle: how narrative feedback is interpreted and acted upon by learners over time, how teachers appropriate orchestration dashboards as decision-support tools under real classroom constraints, and how feedback literacy develops across the school year. In later iterations, optional AI micro-scaffolds are introduced as a bounded exploratory strand, examining how AI-supported tools influence reasoning practices and whether students treat AI outputs as provisional tools or authoritative answers.

The third layer concerns adoption and sustainability. Study 3 also investigates what conditions of workload, trust, governance, and institutional support enable or constrain sustained adoption, treating governance not only as an ethical requirement but as an empirically testable condition of adoption.

Cultural responsiveness functions throughout as a design constraint and an adoption mechanism. In Aotearoa New Zealand, where bicultural obligations under Te Tiriti o Waitangi shape educational practice and institutional expectations, analytics systems must be co-designed with the communities they serve, transparent in what they capture and compute, and governed in ways consistent with data sovereignty expectations (Macfarlane, Macfarlane, and Gillon, 2017; Smith, 1999). The research examines whether transparent, explainable analytics are associated with higher trust and more sustained use by both students and teachers.

2.4 Intellectual Scope and Boundaries

This proposal integrates epistemic cognition, learning analytics, feedback literacy, orchestration research, governance analysis, and bounded AI-mediated interaction. The aim is coherence across levels of a socio-technical system. Analytics in schools operate within nested constraints: the epistemic demands of specific tasks, the feedback literacy of individual learners, the orchestration capacity of teachers, the institutional conditions of schools, and the policy, cultural, and economic forces that shape what is possible. Studying analytics without attending to these layers produces findings that work in laboratories but fail in classrooms.

The thesis will definitively contribute three things: validated indicators of epistemic reasoning quality with transparent reporting of their limits, longitudinal evidence on how narrative feedback and teacher dashboards are interpreted and used, and an adoption conditions framework grounded in triangulated evidence from a real classroom over multiple years.

It will partially explore two further areas: how AI micro-scaffolds and optional translational support tools shape epistemic practices when introduced in Year 3, and how governance mechanisms function empirically as adoption levers. These are bounded exploratory strands with genuine research value, but they are analytically secondary to the core contributions.

It will not claim to have built a generalisable AI-in-education system, established causal relationships between analytics and learning outcomes, or resolved the challenge of computationally detecting reasoning quality. The thesis explores boundaries rather than erasing them.

2.5 Research Context and Programme Ecosystem

The research is situated at Whanganui Intermediate School, where I am employed as a teacher. The STEM inquiry programme is embedded within my teaching role and operates within normal curriculum time. Working within the constraints of a real intermediate classroom, where time is limited, devices are shared, internet connectivity is variable, and feedback must compete with assessment deadlines, behaviour management, and the dozen other demands that structure a school day, shaped the design of this study. These constraints are not noise to be controlled away. They are the conditions under which any analytics system must function if it is to have lasting value.

The programme operates within an ecosystem of partnerships that strengthen feasibility, authenticity, and sustainability. A Science Teaching Leadership Programme placement with Bushy Park Tarapuruhi and UCOL in early 2026 provides the scientific grounding and field methodology that ensures the inquiry contexts are authentic and credible, not simplified classroom simulations. The protected time of the STLP placement offers an unusual opportunity to examine carefully how authentic ecological monitoring practices might be translated into classroom inquiry in ways that remain epistemically honest, preserving the genuine uncertainty and interpretive challenge that make these contexts valuable for analytics research.

Bushy Park Tarapuruhi provides ecological site access and conservation science context, directly supporting the biodiversity monitoring substrate. Horizons Regional Council has expressed support for the programme's alignment with regional biodiversity reconnection and environmental stewardship goals, confirming that the community science contexts are regionally legitimate rather than artificially fabricated for classroom purposes. Skills Group provides advisory input regarding alignment of programme outputs with microcredentials and skills standards. EdTechNZ provides network access and sector positioning. Regenpreneur Ltd provides administrative support and assists with equipment procurement and programme infrastructure.

These partners strengthen the research in three ways. They provide site access and infrastructure that make the inquiry contexts authentic. They provide advisory perspectives that ground the analytics design in real-world educational and environmental expectations. And they create conditions under which adoption can be observed under realistic constraints rather than artificial pilot conditions. The boundary is clear: partners inform context and feasibility; they do not set research questions, control analysis, or determine publication. Academic control of the PhD research, including research design, data analysis, interpretation, and publication decisions, remains with the candidate under university supervision.

This section addresses the programme ecosystem because adoption conditions, studied empirically in Study 3, are shaped by the partnerships, infrastructure realities, and community relationships within which the analytics system must operate. A platform designed without reference to these forces would be designing for a school that does not exist.

2.6 Supervisory Alignment

This section should be customised to name the specific supervisory team at whichever institution the proposal is submitted to. The intellectual alignment points below apply across both potential supervisory configurations.

The research draws on and extends existing work across three complementary areas of expertise within the supervisory team.

The first area concerns epistemic cognition and task-centred analytics. The thesis is grounded in research on how learners evaluate evidence, navigate uncertainty, and construct justified claims, and how digital tools and analytics mediate those epistemic practices. The construct definition, centring on observable epistemic practices including evidence-claim linkage, comparative reasoning, revision depth, and uncertainty handling, draws directly on the epistemic cognition traditions that inform the supervisory team's work. The deliberate emphasis on task-centrality in the analytics design, building indicators around the epistemic demands of specific inquiry contexts rather than around generic behavioural patterns, reflects the argument that analytics must be grounded in what tasks actually demand of learners (Knight, Buckingham Shum, and Littleton, 2014; Chinn, Buckland, and Samarapungavan, 2011).

The second area concerns feedback analytics, orchestration, and socio-technical adoption. The thesis extends research on feedback analytics, the sensemaking of data-based feedback, feedback literacy as a developmental capacity, narrative and storytelling approaches to analytics design, and the socio-technical conditions that enable or constrain sustained adoption. The feedback uptake analysis in Study 3 operationalises the notice, interpret, evaluate, act cycle central to feedback literacy research. Teacher dashboard interaction is analysed using an orchestration framework that traces how analytics cues shape instructional decisions under classroom constraints and how orchestration patterns evolve across the school year (Tsai and Gasevic, 2017; Martinez-Maldonado et al., 2020; Carless and Boud, 2018).

The third area concerns sustainable learning design and narrative practice. The thesis is informed by research on how educational innovations can be designed for institutional viability beyond initial pilot phases, attending to workload, infrastructure constraints, and the professional learning conditions that enable teachers to adopt and sustain new approaches. The narrative feedback design draws on principles of dialogic and relational communication rather than treating feedback as metric delivery. The three-part feedback structure, identifying salient features, interpreting them in relation to epistemic practices, and proposing feasible next actions, is designed to function as a relational prompt that supports student agency rather than directing compliance.

The theoretical framework also draws on scholarship in epistemic cognition, particularly Chinn, Buckland, and Samarapungavan (2011) on dimensions of epistemic cognition, and Sandoval (2005, 2014) on practical epistemology and conjecture mapping.

2.7 Contribution

This thesis is designed to make four contributions.

First, it tests whether trace-based indicators can validly represent aspects of epistemic reasoning quality in sustained school inquiry, and reports transparently on what they detect, what they miss, what requires context-specific calibration, and what remains dependent on human judgement. Where indicators succeed, the thesis provides validated tools. Where they fail, the diagnostic analysis of failure modes advances understanding of the boundaries of computational detection of reasoning quality.

Second, it examines how narrative analytics feedback is interpreted, trusted, and acted upon by school-age learners during sustained inquiry, extending feedback analytics research beyond higher education and single-assessment contexts.

Third, it produces a human-centred analytics design model grounded in co-design with students, teachers, and culturally diverse communities, including a constraints taxonomy, governance mechanisms designed as both ethical safeguards and adoption levers, and transferable design patterns.

Fourth, it identifies the teacher, tool, and institutional conditions under which analytics-enhanced inquiry can be sustained, based on orchestration analysis, temporal dashboard interaction, and triangulated workload evidence rather than self-report surveys.

3. Literature Review

3.1 Learning Analytics: Measurement and Its Limits

Learning analytics, defined as the measurement, collection, analysis, and reporting of data about learners and their contexts for purposes of understanding and optimising learning (Siemens, 2013; Siemens and Long, 2011), has expanded rapidly. The promise lies in making learning processes visible in ways that support timely, evidence-informed decisions by both learners and teachers.

However, reviews consistently identify a gap between technical sophistication and pedagogical impact. Gasevic, Dawson, and Siemens (2015) argue that analytics research must remain centred on learning rather than drifting toward prediction or engagement monitoring as ends in themselves. Providing data does not automatically improve learning. The critical questions are how analytics are designed, presented, and integrated into practice (Tsai and Gasevic, 2017; Wise and Shaffer, 2015).

Most established indicators focus on behavioural proxies: time on task, click frequency, resource access, submission timing. These are computationally tractable and have demonstrated some predictive value for course outcomes in higher education. However, they are poor proxies for the quality of reasoning, evidence use, or conceptual development that characterise meaningful inquiry learning. A student who logs many entries may be documenting without linking to claims. A student who revises frequently may be correcting surface errors rather than reconsidering interpretations. The practices that matter most, evaluating evidence, constructing warranted claims, handling uncertainty, comparing systematically, revising in light of new data, are not easily captured by behavioural traces alone (Knight, Wise, and Chen, 2017).

The validation challenge is significant. Trace-based indicators must be assessed not only for reliability but for construct validity: the degree to which they represent the constructs they claim to measure (Messick, 1995). This requires comparison with human judgement, examination of volume confounding, and cross-context testing. Such work is rare and almost absent in sustained K to 12 settings. The centrality of the task itself in shaping what analytics can meaningfully capture is underappreciated. Analytics designed around generic engagement metrics may miss entirely the epistemic demands that specific inquiry tasks place on learners. This thesis takes task-centrality as a design principle, building indicators around the epistemic demands of specific inquiry contexts rather than around generic behavioural patterns.

It remains an open question how far trace data can meaningfully reach into the quality of reasoning without flattening its complexity. Part of this study is to explore that boundary honestly.

3.2 Epistemic Cognition: The Construct Being Measured

Epistemic cognition refers to how individuals understand knowledge and knowing, including beliefs about certainty, simplicity, sources, and justification (Hofer and Pintrich, 1997). In educational contexts these translate into observable epistemic practices: what students do when evaluating evidence, constructing arguments, handling anomalous data, and revising understanding (Kelly, 2008; Sandoval, 2005).

A distinction within this tradition has direct consequences for analytics design. One strand examines epistemic beliefs, what individuals believe about knowledge. Another examines epistemic practices, what individuals do when reasoning with evidence under uncertainty. This thesis deliberately aligns with the practices tradition. The analytics indicators are designed to detect what students do in their artefacts, not what they believe about knowledge in the abstract. This choice is theoretically motivated, because practices are more directly observable and amenable to trace-based detection, and practically motivated, because feedback based on observable practices is more actionable for 11 to 13-year-olds than feedback based on inferred beliefs.

Chinn, Buckland, and Samarapungavan (2011) articulate dimensions of epistemic cognition relevant to science learning: evaluating evidence reliability, coordinating claims with data, understanding the role of uncertainty and disagreement in knowledge construction, and recognising that claims require justification through evidence and reasoning. These dimensions develop through sustained engagement with environments that demand their use. Students do not learn to handle uncertainty by being told about it. They learn through repeated encounters with situations where outcomes are genuinely unknown and interpretations genuinely matter (Manz, 2015). Sandoval (2005) distinguishes formal epistemology from practical epistemology, arguing that learners develop epistemic understanding through the practice of constructing and evaluating knowledge claims in authentic contexts.

The relationship between epistemic cognition and technological mediation is particularly relevant. When students interact with AI-supported tools during inquiry, the technology becomes part of the epistemic environment. An AI classification tool that provides confidence ratings shapes how students think about uncertainty. A feedback prompt that highlights evidence gaps shapes what students attend to in their own reasoning. Understanding how technological mediation influences epistemic practices requires studying the quality of reasoning as it unfolds within a socio-technical system that includes both human and computational actors.

3.3 Human-AI Interaction and Epistemic Mediation

The increasing presence of AI in educational settings raises questions that extend beyond implementation logistics to the epistemological relationship between human reasoning and machine-generated outputs. When AI systems provide feedback, suggest comparisons, or assist with classification tasks within inquiry, they do not simply deliver information. They mediate the epistemic practices of learners by shaping what is attended to, what counts as relevant evidence, and how uncertainty is represented and managed.

Research on human-AI interaction in education has examined trust calibration, the degree to which learners appropriately adjust their reliance on AI outputs based on task characteristics and output quality (Holmes and Tuomi, 2022). However, most of this work has been conducted with adult learners in higher education. Little is known about how younger learners, particularly those aged 11 to 13, evaluate the epistemic status of AI-generated suggestions. Do they treat AI outputs as authoritative answers or as provisional claims to be tested against their own evidence and reasoning?

Epistemic outsourcing, where learners defer judgement to an external source rather than exercising their own reasoning, is a well-documented concern in the epistemology literature. AI tools may exacerbate this tendency if they present outputs as definitive rather than uncertain, or if learners lack the epistemic skills to evaluate AI suggestions critically. Conversely, AI tools designed to foreground uncertainty, to present outputs explicitly as claims requiring verification, and to prompt comparison and evaluation rather than acceptance may function as productive epistemic partners that extend rather than replace human reasoning.

The biodiversity monitoring context in this thesis provides a rich environment for studying these dynamics. Students using AI-assisted classification tools encounter outputs with variable confidence levels, classifications that sometimes conflict with traditional identification methods, and situations where the data quality is insufficient for reliable automated processing. These are authentic conditions for studying how AI mediation shapes epistemic practice. The AI interaction investigation is introduced in Year 3 as a bounded strand within Study 3 and is analytically secondary to the core feedback uptake and orchestration work.

This thesis does not claim to resolve the challenge of AI-mediated epistemic interaction. It contributes bounded empirical evidence from one sustained context about how younger learners respond to AI outputs during authentic inquiry, and whether design features, specifically framing outputs as claims to check rather than answers to accept, influence the quality of that response.

3.4 Feedback Literacy and Feedback Analytics

Feedback is widely recognised as one of the most powerful influences on learning, yet its effects are highly variable (Hattie and Timperley, 2007). Feedback literacy, as articulated by Carless and Boud (2018), reconceptualises feedback from an event to a process requiring learners to appreciate its purpose, judge their own work quality, manage the emotional dimensions of receiving evaluative information, and take informed action. This reconceptualisation has significant implications for analytics. If feedback is a process requiring active engagement at multiple stages, systems must support the notice, interpret, evaluate, act cycle that characterises productive feedback engagement (Winstone and Carless, 2019).

The feedback literacy framework explicitly includes affect and emotional regulation as part of productive engagement. The present study does not deploy a formal affect instrument, but captures affective response indirectly through interview data and observation of how students respond emotionally to analytics feedback. This acknowledges that reasoning about feedback is never purely cognitive, particularly for younger learners encountering evaluative information about their thinking. The language and relational qualities of feedback prompts matter substantially for this age group, and the design of narrative feedback in this thesis attends to those qualities deliberately.

Feedback analytics investigates how data-driven feedback can support learning rather than merely report metrics. Tsai and Gasevic (2017) have investigated how learners engage with data-based feedback over time: whether they access, revisit, revise, and change strategies. This work traces feedback engagement as a temporal sequence with identifiable patterns.

Narrative analytics structures feedback as a short, evidence-grounded storyline identifying salient features in the learner's work, interpreting them in relation to learning goals, and proposing a feasible next action (Echeverria et al., 2018). For younger learners with limited experience interpreting data displays, narrative framing may be particularly important for bridging the gap between computation and comprehension. The dialogic quality of narrative feedback, its capacity to position the learner as an active interlocutor rather than a passive recipient, distinguishes it from metric-based feedback that positions the learner as a data point.

Most feedback analytics research has been conducted in higher education on completed assignments. Little is known about how feedback literacy develops among younger learners engaging with analytics embedded in ongoing inquiry. Longitudinal studies tracking how interpretation, trust, and response quality change as learners become more familiar with systems and as inquiry deepens are rare. This thesis addresses these gaps.

3.5 Teacher Orchestration and Dashboard Design

Teachers in inquiry classrooms face orchestration demands qualitatively different from traditional instruction. Multiple groups pursuing different investigations require simultaneous monitoring, rapid decision-making about intervention, and real-time balancing of competing attention demands (Dillenbourg, 2013). These decisions are consequential.

Dashboards have been proposed as support tools, but design research reveals persistent challenges. Dashboards fail when they display too much information, lack actionability, ignore workflow, or are not integrated into routines (Martinez-Maldonado et al., 2020). The critical question is not what data to show but how to present it for rapid sensemaking under realistic conditions, where a teacher may have thirty seconds before a student approaches.

The analytic unit in orchestration research is the teacher decision episode: notice cue, interpret meaning within lesson context, make pedagogical move, observe consequence. This reconceptualises dashboards from information displays into decision-support tools situated within classroom ecology. The dashboard does not teach. The teacher teaches, informed by the dashboard. Sensemaking theory (Weick, 1995) complements this: teachers interpret analytics under time pressure and competing demands, so dashboard information must support rapid interpretation rather than careful study.

A question that has received less attention is whether analytics-supported dashboards change the nature of teacher orchestration over time, whether they become integrated into professional routines in ways that enhance rather than constrain pedagogical judgement, or whether they introduce new dependencies and new forms of labour. This study extends orchestration research into sustained intermediate school inquiry, analysing dashboard interaction as temporally situated episodes across the school year.

3.6 Community Science, Project-Based Learning, and Curriculum Integration

Community science offers a promising approach to epistemic authenticity. When students contribute to genuine scientific efforts, they encounter the same epistemic challenges as professional practice (Phillips et al., 2019; Bonney et al., 2014). Classification is uncertain. Data are incomplete. Patterns require interpretation. The need to justify claims extends beyond satisfying a teacher to contributing credibly to a shared knowledge base.

However, most educational implementations remain peripheral to core instruction, and learning analytics has rarely been applied to community science despite the rich process data generated. Project-Based Learning positions learners as active investigators over extended periods (Blumenfeld et al., 1991). Its effectiveness depends on scaffolding and facilitation quality, precisely where analytics might contribute (Thomas, 2000). Systems thinking develops capacity to recognise feedback loops, tolerate delayed effects, and reason about causation in complex systems (Meadows, 2008; Hmelo-Silver and Azevedo, 2006). In this study, systems thinking appears as a domain-specific expression of epistemic practice within the aquaponics context.

The programme operates within the New Zealand Curriculum's Nature of Science strand (Ministry of Education, 2007) and is designed with principles consistent with international inquiry frameworks, supporting transferability. The challenge of designing analytics systems that strengthen epistemic reasoning while remaining sustainable under classroom constraints is shared across educational systems internationally. Rather than depending on specific content standards, the programme generates portable evidence of reasoning practice mappable to whatever credential frameworks apply.

3.7 Ethics, Equity, and Culturally Responsive Analytics Design

Learning analytics in schools raises ethical questions about privacy, algorithmic bias, equity of access, and the risk that analytics are experienced as surveillance rather than support (Drachsler and Greller, 2016; Tsai and Gasevic, 2017). In Aotearoa, culturally responsive pedagogy requires that technologies respect community values (Macfarlane, Macfarlane, and Gillon, 2017; Smith, 1999). Māori data sovereignty requires attention to control, use, and benefit of data (Te Mana Raraunga, 2018). These are design constraints from the earliest stages, not add-ons.

Systematic reviews show relatively few AI-in-education studies engage critically with ethics or connect ethical design to pedagogy (Zawacki-Richter et al., 2019; Bates et al., 2020). The integration of practical ethics within system architecture, rather than as a post-hoc compliance exercise, is central to this thesis. Ethics-by-design means that transparency mechanisms, governance protocols, and consent processes are embedded within the platform and the research design from the outset. Concerns about surveillance and erosion of professional autonomy find parallels in schools (Williamson, 2018). In this study, governance and transparency are both ethical obligations and testable adoption mechanisms.

3.8 Socio-Technical and Policy Context for Analytics Adoption

Analytics in schools operate within forces that most analytics research ignores but that determine whether systems are adopted, sustained, or abandoned. Understanding those forces was essential for making design decisions that would survive contact with a real school. The contextual analysis described here functions as socio-technical constraint-mapping that shaped the analytics architecture and adoption analysis. Each factor is linked to a specific design constraint or methodological decision.

Political and policy factors. The New Zealand curriculum refresh context, where Te Mātaiaho's emphasis on science capability, inquiry, and local curriculum design supports the inquiry framing but increases variability across schools, means GrowHub must be adaptable rather than tied to one scheme. School board expectations around safety, data use, and reputational risk increase the need for governance, audit trails, and transparent analytics. Te Tiriti obligations mean co-design and shared governance are not optional ethics extras; they are adoption conditions. Design constraint: governance co-designed with community; platform adaptable to local curriculum.

Economic factors. Budget constraints and fragile device and sensor resourcing demand offline-first, low-cost workflows designed for shared devices and intermittent connectivity. Teacher time is the scarcest resource; dashboards must save time, not create new interpretive labour. Study 3 explicitly measures workload impact. Tools often die when researcher support ends; the adoption conditions framework must address sustainability beyond the pilot phase. Design constraint: lightweight infrastructure; workload measurement as core research activity; routines over features.

Sociocultural factors. Mixed teacher confidence in science and analytics means the design must reduce fear: bounded indicators, explainable feedback, claims-to-check framing rather than authoritative scoring. Student affect and surveillance sensitivity demand non-deficit language, student visibility of tracking, and explanations of purpose. Whānau trust and local legitimacy require governance hui, clear data controls, and culturally responsive consent. Design constraint: non-deficit framing; student-visible analytics; community-governed data.

Technological factors. Connectivity variability during field sessions and within school Wi-Fi demands offline capture with asynchronous sync, avoiding heavy real-time dependence. Many New Zealand schools run Google Workspace, so GrowHub is designed for interoperability with commonly adopted school ecosystems without implying partnership or endorsement. AI tools shift rapidly; all AI features are bounded, optional, rollbackable, and logged for epistemic outsourcing risk analysis. Design constraint: offline-first architecture; Google ecosystem compatibility; AI features auditable and removable.

Environmental factors. The inquiry contexts are environmental, so technology choices should align with environmental responsibility: lightweight compute, minimal data retention, sensible sampling. Fieldwork constraints including weather, seasonality, and species variation affect trace and artefact regularity; these are methodological realities to plan for rather than noise to eliminate. Design constraint: resource-efficient platform; data collection plans accommodate seasonal variability.

Legal and ethical factors. New Zealand Privacy Act obligations, school policies, and platform consent requirements demand role-based access, data minimisation, pseudonymisation, and deletion pathways. Māori data sovereignty expectations require governance mechanisms, local control, and clear boundaries on data sharing. Child safeguarding requires clarity that analytics are decision-support, not automated judgement, with no ranking and no punitive use. Design constraint: privacy-by-design; governance protocols co-designed with community; no automated judgement of students.

These forces imply that Study 1 must co-design governance and workflow rather than imposing them, Study 3 must measure workload and trust as core adoption variables, and AI features must be optional and auditable throughout.

3.9 Identified Research Gap

The literature identifies several converging gaps. Validated indicators for epistemic reasoning quality in sustained K to 12 inquiry are rare. Feedback literacy research has not examined how younger learners engage with analytics-based feedback during sustained inquiry. Teacher orchestration studies have not tracked dashboard use evolution under real workload constraints over extended periods. Human-AI interaction research in school settings has not examined how younger learners evaluate the epistemic status of AI-generated suggestions within sustained inquiry over time. Learning analytics has rarely been applied to community science contexts. Culturally responsive analytics design with integrated ethics-by-design remains largely aspirational.

This thesis takes up these gaps by attempting to design and validate human-centred analytics for epistemic reasoning quality and studying how feedback, dashboards, and bounded AI mediation are interpreted, acted upon, and sustained in a school-based programme in Aotearoa New Zealand. The contribution is not to solve all gaps comprehensively but to provide carefully bounded empirical evidence from one sustained, real-world context that advances understanding of what analytics can and cannot do for inquiry learning in schools.

4. Stakeholder-Informed Foundations

4.1 Purpose and Rationale

Before committing to the analytics architecture, I conducted a national educator survey to ensure the design would respond to real constraints and priorities rather than assumptions about what teachers and students need. The survey engaged 189 educators, school leaders, and youth development practitioners across Aotearoa. Its function was to surface what stakeholders considered essential for engagement, equity, and meaningful outcomes, and to identify barriers the analytics design would need to accommodate (Barab and Squire, 2004; Selwyn et al., 2020).

The survey established an empirical basis for design constraints and provided a baseline reference against which perspectives can be tracked as AI tools evolve. The sample was recruited through professional networks using purposive sampling, likely biased toward enthusiastic early adopters and therefore not nationally representative, but useful as a constraints signal from engaged practitioners (Cohen, Manion, and Morrison, 2018).

4.2 Method

The survey adopted a mixed-methods design. The quantitative strand used a 0 to 100 slider scale for twelve programme components, analysed descriptively. The qualitative strand used open-ended questions analysed through reflexive thematic analysis (Braun and Clarke, 2006, 2021), with codes generated inductively from participant language.

4.3 Participant Profile

The dataset reflected professional breadth: school-based educators across primary, intermediate, and secondary settings; leadership roles; specialist positions; and community-based practitioners. Geographic distribution was balanced across urban (48%), peri-urban (27%), and rural or remote (25%) contexts. Self-reported AI familiarity ranged from beginner (31%) through intermediate (49%) to advanced (20%). Over 90% expressed strong enthusiasm for structured student engagement with inquiry and digital tools.

4.4 Quantitative Findings

All twelve components received mean ratings above the midpoint. The highest-rated were Introduction to AI (mean 93.16) and Ethical Use of AI (mean 90.33), demonstrating consensus that early engagement must be grounded in literacy and ethical awareness (UNESCO, 2023; Long and Magerko, 2020). Other highly rated elements included Teamwork (84.65), 3D Design and Digital Fabrication (83.72), and NZQA Microcredentials (82.30). Mid-ranked components such as Design Thinking (69.28, range 5 to 100) showed wider variation, indicating diversity in familiarity and confidence.

4.5 Qualitative Findings

Six themes emerged from 118 open-ended responses. Equity of access (66%) concerned cost, transport, device availability, and time. Tangible outcomes and future pathways (53%) stressed recognised outputs. Cultural relevance and tikanga Māori (50%) highlighted te ao Māori values and Māori data sovereignty. Ethics and responsible technology use (46%) called for critical interrogation of AI outputs. Tool fluency and everyday application (41%) recommended familiar tools. Inclusive design for diverse learners (26%) emphasised cultural and emotional safety, neurodiversity support, and flexible pathways.

4.6 Survey-to-Design Traceability

The stakeholder findings are not merely contextual background. They directly shaped specific design decisions and methodological choices. This traceability from survey evidence to design constraint to research method is the mechanism through which participatory constraint-mapping translates into system architecture.

Stakeholders consistently identified offline access and equity constraints as barriers to participation. This produces the design constraint that GrowHub must operate with offline capture, asynchronous sync, minimal device requirements, and shared-device workflows. Methodologically, adoption analysis in Study 3 treats infrastructure failures as part of the socio-technical conditions to be examined, not as noise to be excluded.

The high priority placed on AI literacy and ethical use produces the design constraint that AI outputs within GrowHub are framed as claims to check rather than authoritative answers, with built-in reflection prompts, optionality, and explicit explanations of why the model might be wrong. This framing directly addresses the epistemic outsourcing risk identified in the literature. Methodologically, the uptake analysis in Study 3 codes for evaluation behaviour: whether students verify AI suggestions, cite evidence, and revise in response, or whether they accept AI outputs uncritically.

The desire for tangible outcomes and pathways produces the design constraint that evidence of reasoning practice is exportable in portfolio-compatible formats and linked to artefact-based evidence trails. This is also where the AI translational features find their warrant: supporting students in converting inquiry reasoning into executable artefacts, prototypes, and designs. Methodologically, productive uptake is defined in ways that include building, testing, and making, not only writing more text.

Stakeholders did not ask for more data about students. They asked for tools that help students do meaningful work and help teachers understand what students are learning. This aligns with the feedback literacy literature's emphasis on analytics for learner agency rather than surveillance, and provides empirical warrant for the human-centred design approach in Study 1.

4.7 Prototype- and Pilot-Informed Design Signals (Planned 2026)

As part of the programme's 2026 rollout, the two inquiry substrates, aquaponics monitoring and light-trap biodiversity monitoring, will be used as structured pilots to surface workflow constraints and inform the GrowHub interface and feedback design. These pilots are scheduled to begin after the STLP placement (from approximately Term 2 to 3, 2026) and are not yet completed at the time of writing.

The pilots will explicitly examine non-technical constraints that frequently determine classroom viability: offline-capable capture, low-lift routines fitting within standard lesson time, student comprehension of prompts without feeling surveilled, and teacher views that highlight actionable exceptions rather than comprehensive displays. They will also test whether students engage more productively when AI outputs are framed as claims to be checked rather than answers to be accepted. These are treated as empirical questions to be investigated through observation, trace data, and participant feedback during Studies 1 and 3, rather than as findings already established.

5. Theoretical Framework

5.1 Inquiry Process Quality as the Central Construct

The central construct is inquiry process quality: the observable epistemic practices learners engage in when they evaluate evidence, link claims to data, compare across conditions, revise reasoning in response to new information, and handle uncertainty. This construct is defined in terms of practices rather than dispositions or beliefs, because practices can be observed in artefacts, traced through platform logs, and coded with measurable reliability.

The construct draws on two overlapping traditions. Epistemic cognition specifies what reasoning practices matter: evaluating evidence reliability, coordinating claims with data, tolerating uncertainty, and revising interpretations (Chinn, Buckland, and Samarapungavan, 2011; Sandoval, 2005). Self-regulated learning specifies how learners plan, monitor, and adapt inquiry processes, including engaging with and responding to feedback (Winne and Hadwin, 1998; Zimmerman, 2002). In sustained inquiry these are not separate. When a student re-tests a measurement because yesterday's result was unexpected, that involves both epistemic judgement and self-regulatory monitoring. The thesis uses epistemic cognition to define rubric dimensions and self-regulated learning to explain the feedback engagement process.

Four observable practices are examined.

Evidence-claim linkage captures whether students ground assertions in specific, identified evidence. At a basic level this might appear as "the fish are not eating well." At a more sophisticated level: "the fish ate only two of five pellets on Tuesday and refused food on Wednesday, which suggests appetite has declined since the pH dropped below 6.5 last week."

Comparative reasoning captures whether students examine patterns systematically across conditions, time periods, or data sources rather than relying on single-observation impressions.

Revision depth captures whether revisions address surface content or underlying reasoning.

Uncertainty handling captures whether students calibrate confidence appropriately, distinguishing well-supported claims from tentative ones and areas of insufficient data. This is distinct from blanket hedging and blanket certainty.

These are defined as coding rules closer to observable behaviours than to inferences about internal states, making them tractable for analytics, transparent for validation, and defensible as feedback foundations.

Two design tensions are acknowledged. Structured templates may shape writing in ways that artificially inflate indicator detection. If templates prompt evidence references, indicators may partly measure compliance with structure rather than spontaneous epistemic reasoning. Revision detection may over-represent students comfortable with text editing and under-represent careful thinkers who write less. These are monitored through validation and reported as boundary conditions rather than treated as oversights. Where template-induced artefacts are detected, they become findings about the limits of trace-based measurement.

5.2 AI as Bounded Epistemic Mediator

A distinctive theoretical commitment of this thesis is the treatment of AI tools not as neutral productivity aids but as epistemic mediators that shape the reasoning practices of the learners who interact with them. This framing draws on the epistemic cognition tradition's recognition that reasoning does not occur in a cognitive vacuum but is shaped by the tools, tasks, and social environments within which it is practised.

When an AI classification tool provides a confidence-rated species identification, it shapes the epistemic environment by making certain information salient, by representing uncertainty in a particular way, and by creating a default interpretation that the learner must either accept or evaluate. When a narrative feedback prompt highlights an evidence gap, it directs attention to specific aspects of the learner's reasoning and frames what counts as an appropriate next step.

The thesis therefore examines AI features not primarily as technical components but as epistemic interventions whose effects on reasoning practices must be empirically studied. The key questions are whether AI mediation supports the development of the four epistemic practices defined above, and whether patterns of interaction with AI tools shift over time as learners develop more sophisticated strategies for evaluating AI outputs.

However, AI mediation is bounded within this thesis. It is introduced only in Year 3 as an exploratory strand within Study 3. It does not bear the weight of the thesis. The core contributions concern construct validation, feedback uptake, and adoption conditions. AI mediation is studied where it intersects with those contributions, not as a standalone investigation.

5.3 Feedback as Socio-Technical Process

Analytics feedback is treated as a socio-technical process. Multiple layers of mediation determine whether data-driven feedback influences practice. Feedback supports learning only when learners notice, interpret, evaluate, decide, act, and evaluate the result (Carless and Boud, 2018; Winne, 2017). Each step can fail.

Teacher mediation adds a critical layer. Teachers interpret dashboards through professional judgement, student knowledge, and inquiry context, then translate interpretation into instructional moves. This mediation is not noise. It is a central mechanism to be studied, representing the pathway through which analytics become pedagogical action. The relational dynamics of teacher-student interaction shape how feedback is received and whether it is experienced as supportive or evaluative.

Institutional conditions add a further layer. Sustained use depends on workload, trust, professional alignment, leadership support, and fit with routines. Socio-technical frameworks (Orlikowski, 2000) are more productive than simple technology acceptance models because they examine interactions among tools, practices, roles, and structures.

5.4 Narrative Feedback as Dialogic Design

Narrative feedback performs three functions: identifying salient features in the learner's artefact, interpreting those features in relation to the inquiry-quality construct, and proposing a feasible next action (Echeverria et al., 2018). For example: "Your last three entries described what happened but did not connect observations to specific measurements. Which data points support what you observed? Try linking your claim about plant growth to the nitrate readings from Tuesday and Wednesday."

The feedback prompt is structured not as a verdict but as an invitation to re-examine. Its language is deliberate: it names what was observed in the student's work, offers an interpretation that the student can evaluate, and proposes a specific, feasible action. For 11 to 13-year-olds encountering evaluative information about their reasoning, the tone, framing, and perceived intent of feedback shape whether it is experienced as helpful or threatening, and therefore whether it is engaged with productively.

5.5 Conjecture Mapping

Sandoval's (2014) conjecture mapping framework traces whether design mechanisms operate in practice. Four initial conjectures structure the investigation.

First, evidence-linked feedback prompts will be associated with reasoning-level revision rather than surface edits. Second, dashboards presenting actionable exceptions will be used during lessons rather than only between them. Third, transparent, explainable analytics will be associated with higher trust and lower abandonment, such that governance features function as adoption mechanisms. Fourth, optional AI features will be used selectively under higher uncertainty, with uptake varying by task type and context.

These are deliberately modest. They specify expected mediating processes, not guaranteed outcomes.

5.6 Construct Validity Framework

Trace-based indicators must be validated against the constructs they claim to represent (Messick, 1995).

Convergent validity examines agreement between analytics indicators and human coding. Discriminant validity models indicator scores against volume proxies (word count, entry frequency, time on task) using multilevel regression models appropriate for the nested data structure (students within groups within cohorts) and tests whether indicators retain explanatory power after controlling for volume. Indicators primarily detecting quantity are reclassified. Consequential validity is examined through Study 3 by investigating whether indicator use supports or harms inquiry practices in actual use.

6. Research Questions

RQ1 (Construct Articulation and Co-Design): How can epistemic reasoning in sustained K to 12 inquiry be operationalised into observable, task-aligned indicators through participatory co-design with teachers, students, and community stakeholders, and what transparency, governance, and explainability constraints emerge from that process?

Data: qualitative co-design data including workshop transcripts, design artefacts, usability recordings, and interview data.

RQ2 (Indicator Validation): To what extent do trace-based indicators of epistemic reasoning demonstrate convergent and discriminant validity when triangulated against human-coded artefacts and qualitative reasoning evidence across contrasting inquiry contexts?

Data: platform logs, human-coded artefacts (approximately 120 to 150 per context per year, generated by approximately 60 students per cohort across two cohort years, yielding an estimated 4 to 5 coded artefacts per student per context per year), validation statistics including weighted kappa, intraclass correlation, and discriminant validity modelling using multilevel regression.

RQ3 (Uptake, Orchestration, and Sustainability): How are epistemic analytics interpreted, taken up, and orchestrated by teachers and students in sustained inquiry, and what socio-technical conditions shape their sustainable adoption?

Data: feedback uptake sequences with typology coding, orchestration episode observations, workload logs from multiple triangulated sources, student and teacher interviews, platform analytics, and AI interaction traces.

The design is quasi-experimental. The research traces mechanisms and associations rather than establishing causation. Causal claims are not inferred.

7. Methodology

7.1 Overall Design

This research adopts a design-based research methodology (Brown, 1992; Collins, 1992; Design-Based Research Collective, 2003). Design-based research supports the iterative development of educational interventions in authentic settings, produces both practical tools and theoretical understanding, and treats the complexity of real classrooms as the context to be understood rather than as noise to be eliminated (Barab and Squire, 2004; McKenney and Reeves, 2019; Anderson and Shattuck, 2012).

The choice reflects the nature of the research problem. The analytics system being studied does not exist as a finished product to be evaluated. It is being co-designed with its users, developed iteratively in response to what is learned through each deployment cycle, and refined as evidence accumulates about what works, what fails, and why. In my experience as a teacher working with complex learning environments, the interventions that succeed are those that evolve in response to what actually happens, not those locked into a fixed design before they encounter the realities of a classroom. Design-based research formalises that responsive process within a rigorous framework.

The research proceeds through three linked studies, each addressing one research question and building on the findings of the studies before it. Study 1 produces the design model and constraints taxonomy that shape the platform. Study 2 validates the analytics indicators the platform computes. Study 3 examines how the feedback and dashboards generated by those indicators are interpreted, used, and sustained in practice. The studies are not independent investigations. They form a coherent chain in which each depends on the outputs of the previous one and informs the interpretation of the next.

The overall data architecture combines qualitative and quantitative methods in a convergent mixed-methods design (Creswell and Plano Clark, 2017). Qualitative data, including co-design transcripts, interviews, observation field notes, and artefact analysis, provide depth and explanatory power. Quantitative data, including platform trace logs, validation statistics, and uptake frequency counts, provide breadth and systematic coverage. Integration occurs at the interpretation stage, where qualitative evidence explains quantitative patterns and quantitative evidence tests the generalisability of qualitative themes.

7.2 Methodological Hierarchy

It is worth being explicit about the hierarchy governing the research design, because it directly addresses a legitimate concern about scope.

The intellectual centre of this thesis is human-centred learning analytics: how indicators of epistemic reasoning quality are designed, whether they validly measure what they claim to measure, how feedback based on those indicators is interpreted and acted upon, and what conditions sustain analytics use in schools. Every methodological decision is evaluated against that centre.

The environmental inquiry activities, aquaponics monitoring and biodiversity monitoring, are the contexts in which analytics are validated and feedback is studied. They generate the sustained, authentic, uncertainty-laden inquiry data that the analytics require. They are essential infrastructure, but they are not the intellectual contribution. Methodological decisions about how these activities are structured serve the analytics research, not the other way around. The inquiry substrates are not evaluated as environmental education interventions within this thesis. Their function is to generate authentic, sustained, and epistemically contrasting inquiry traces against which the analytics system is designed and tested.

The co-design process is both a research method that produces findings, specifically the design model and constraints taxonomy reported in Study 1, and a generative process that produces the analytics system studied in Studies 2 and 3. The thesis reports it as such.

The platform, GrowHub, is the analytics system being studied. It is not a product being promoted. Decisions about platform features serve the research questions. Where features work, they provide evidence about what analytics can do. Where they fail or are abandoned, they provide evidence about what analytics cannot do, or about the conditions that would be needed. Both outcomes are informative.

The contextual analyses, including the stakeholder survey, SWOT, and socio-technical constraint-mapping, function as participatory constraint-mapping and institutional viability modelling. They feed directly into Study 1 design decisions and Study 3 adoption analysis.

7.3 Reflexive Safeguards

The dual role of teacher and researcher creates specific risks that require procedural mitigation. I am embedded in the classroom, which provides sustained access and deep contextual knowledge. It also means I have a professional relationship with student participants, a stake in the programme's success, and inevitable prior interpretations of what is happening in my classroom. These conditions can produce confirmation bias, demand characteristics, and selective attention.

The following safeguards are built into the research design. A reflexive journal is maintained throughout the research, documenting interpretive decisions, moments of uncertainty, and instances where teaching priorities and research priorities diverge. This journal is reviewed periodically with supervisors. All rubric coding for Study 2 is conducted by independent coders; the researcher does not code artefacts from their own students for validation purposes.

Thematic analysis in Studies 1 and 3 is subject to supervisory audit at the theme development stage, with supervisors reviewing coding decisions against raw data. Teaching evaluation and assessment data generated through normal professional responsibilities are kept separate from research data; the research uses only data collected through ethically approved research instruments and protocols. Where the researcher's contextual interpretations inform analysis, these are made visible and available for scrutiny rather than embedded invisibly in findings. The aim is not to eliminate the influence of the dual role but to make that influence visible, bounded, and auditable.

7.4 Study 1: Co-Design and Design Model

Purpose. Study 1 addresses RQ1 by co-designing the analytics system with the people who will use it and be affected by it, producing a design model and constraints taxonomy that reflect their values, priorities, and practical realities.

Participants. Three stakeholder groups are involved: students aged 11 to 13 from the inquiry programme (approximately 60 per cohort year), teachers at Whanganui Intermediate School, and community stakeholders including whānau members and representatives from partner organisations. The inclusion of community stakeholders is particularly important in the Aotearoa context, where Māori data sovereignty principles and Te Tiriti obligations require that communities have genuine influence over how data about their young people is collected, computed, and used.

Process. The co-design follows an iterative workshop process across Terms 1 and 2 of Year 1. Three to four workshops are conducted with each stakeholder group, progressing from needs and values elicitation through design specification and usability testing of early prototypes. Workshops are structured around specific design decisions rather than abstract preferences. Rather than asking what kind of feedback students want, the workshops present contrasting feedback formats, such as a numeric score, a colour-coded dashboard, and a narrative prompt, and ask participants to compare them: which would help you understand what to do next, which feels fair, which would you trust, which raises concerns about what is being tracked.

Design decisions are documented as design patterns using a structured format: problem, context, solution, and trade-offs. This preserves the rationale for each decision and ensures the model is transferable rather than tied to the specific configuration of one school.

Governance mechanisms. A specific focus is the development of governance protocols including an indicator explainer describing in accessible language what is tracked and why, data sharing controls specifying who sees what under what conditions, an audit trail recording what feedback was generated and when, and exit policies enabling data deletion on request. These features are designed as both ethical safeguards and testable adoption mechanisms. Study 3 examines whether transparent, explainable analytics are associated with different trust and use patterns through temporal alignment analysis. The integration of practical ethics within the system architecture reflects the ethics-by-design principle that ethical considerations should inform system design from the outset rather than being added as compliance exercises after the fact.

Consent. Rather than relying solely on written forms, the consent process uses visual explanations, platform demonstrations, and hui with whānau. The goal is genuinely informed consent rather than procedural compliance. Ethics approval will be sought from the relevant university human research ethics committee and relevant New Zealand authorities before data collection begins.

Outputs. Study 1 produces: a design model documented as transferable design patterns; a constraints taxonomy spanning transparency, governance, explainability, workload, infrastructure, and cultural responsiveness; governance protocols grounded in community values; and Platform Version A, the initial deployment-ready analytics system.

Analysis. Workshop transcripts, design artefacts, and usability recordings are analysed through reflexive thematic analysis (Braun and Clarke, 2006, 2021), with theme development audited by supervisors as described in Section 7.3.

7.5 Study 2: Indicator Development and Validation

Purpose. Study 2 addresses RQ2 by developing trace-based indicators of inquiry process quality and validating them against human-coded artefacts. The study tests what indicators can and cannot detect, whether they function consistently across contrasting epistemic contexts, and where context-specific calibration is required. Transparent reporting of indicator limits is a central commitment.

Inquiry contexts as epistemic stress tests. The two substrates are selected because they exert different epistemic pressures and therefore stress-test different aspects of the analytics design. Aquaponics monitoring engages students in tracking a living system where parameters change slowly, causal relationships are delayed and ambiguous because multiple variables interact simultaneously, and interpretation must be built from accumulated observation over weeks rather than from single experimental trials. This context challenges the analytics system to detect revision depth and longitudinal evidence linkage where causation is never simple. Biodiversity monitoring using light-trap methods engages students in classifying organisms using both traditional identification and AI-assisted classification tools, where classification confidence varies with specimen quality and environmental conditions, data quality depends on factors students cannot fully control, and students must evaluate when automated tools are reliable and when they are not. This context challenges the system to detect calibrated uncertainty and comparative reasoning under variable data quality.

These contexts will be used as structured pilots beginning in 2026 to surface workflow constraints and generate the sustained artefact data against which indicators are validated. They are epistemic stress tests, not the objects of investigation.

Artefact collection and human coding. Approximately 120 to 150 student artefacts are collected per context per year from regular inquiry entries on the platform. With approximately 60 students per cohort, this yields an estimated 4 to 5 coded artefacts per student per context per year. Across two contexts and two cohort years, the total validation corpus is expected to comprise approximately 500 to 600 coded artefacts. Each artefact is coded by two independent human coders using the inquiry process quality rubric operationalising the four practices defined in Section 5.1. Each practice is coded on a four-level scale from absent through emerging, developing, and proficient, with concrete behavioural anchors at each level.

Coding protocols are developed and piloted during Year 1. Inter-rater reliability is assessed using weighted kappa, with a minimum threshold of kappa greater than or equal to 0.60 required for each rubric dimension before indicator validation proceeds. This threshold represents moderate agreement (Landis and Koch, 1977) and is selected as a pragmatic minimum given the interpretive complexity of coding epistemic practices in the artefacts of 11 to 13-year-olds, where wording is often informal and reasoning is frequently implicit. If reliability exceeds 0.70 on specific dimensions, this is reported as stronger evidence. Where reliability falls below threshold, rubric definitions are refined and coders retrained before resuming.

AI-assisted coding workflow. AI tools may be used in a limited way to support the coding workflow, for example by suggesting candidate segments where evidence references or comparisons may be present. However, rubric level assignment remains a human judgement task. Any AI-assisted step is auditable and is tested for its effect on inter-rater reliability and potential bias. The primary validation signal remains agreement between trace-derived indicators and independent human coding. This approach allows the study to examine whether AI assistance changes reliability or introduces systematic distortion, which is itself a useful methodological finding.

Trace-based indicator development. The platform captures process traces as students work: text content, entry timing, revision history, use of data tables and comparison tools, query patterns, and optional AI scaffold interactions. From these traces, computational indicators are derived for each practice. For example, evidence-claim linkage indicators draw on whether entries reference specific data points, measurements, or dated observations. Comparative reasoning indicators draw on whether students use comparison tools, reference multiple conditions, or explicitly contrast observations.

Indicator development is iterative: initial indicators are computed, compared with human coding, refined where discrepancies suggest miscalibration, and retested. This iterative refinement is standard in construct validation work and consistent with design-based research methodology.

A specific risk is acknowledged. The structured templates through which students enter their inquiry work may shape writing in ways that artificially inflate indicator detection. If templates prompt students to reference evidence, the indicator may partly measure compliance with template structure rather than spontaneous epistemic reasoning. Revision indicators may over-represent students comfortable with text editing and under-represent students who reason carefully but write less. These template-induced artefacts are monitored through validation and reported as findings about the boundary conditions of trace-based measurement.

Validation analysis. Convergent validity is assessed through agreement between analytics indicators and human rubric codes for the same artefacts. Agreement is quantified using weighted kappa and intraclass correlation coefficients. The target is moderate to substantial agreement (kappa greater than or equal to 0.60) with indicator-by-indicator reporting of agreement levels.

Discriminant validity is assessed using multilevel regression models appropriate for the nested data structure (artefacts within students within groups within cohort years), testing whether indicators retain explanatory power for rubric levels after controlling for volume proxies including word count, entry frequency, and time on task. If indicators retain explanatory power, they are functioning as reasoning-quality indicators rather than engagement proxies. If they do not, they are reclassified as volume-confounded and reported as such. Without this test, there is no way to distinguish a reasoning-quality indicator from an engagement proxy.

Cross-context comparison examines whether indicators function similarly across the aquaponics and biodiversity contexts, or whether context-specific features produce systematic differences in indicator behaviour. This comparison is important for the task-centrality principle. If indicators require substantial recalibration between contexts, that finding has direct implications for the transferability of analytics systems designed around epistemic practices. Where context-specific calibration is required, this is documented as a finding about boundary conditions rather than treated as a failure.

Failure reporting. Where indicators fail validation, diagnostic analysis examines why, distinguishing among construct under-specification where the rubric dimension is too broad for a single indicator to capture, trace sparsity where the platform data does not contain sufficient signal, context dependence where an indicator works in one epistemic environment but not the other, and volume confounding where the indicator primarily reflects quantity rather than quality. This diagnostic analysis is itself a contribution. It advances understanding of what aspects of epistemic reasoning quality are and are not computationally detectable in authentic school settings. The thesis is designed to produce publishable findings even under weak indicator validity, because mapping detection limits and failure modes in sustained K to 12 inquiry is a missing empirical contribution.

Output. Study 2 produces: a set of validated indicators with transparent reporting of agreement levels, detection limits, and context sensitivity; diagnostic analysis of indicators that fail with classified failure modes; and cross-context comparison examining generalisability and calibration requirements.

7.6 Study 3: Feedback Uptake, Teacher Orchestration, and Adoption

Purpose. Study 3 addresses RQ3 by examining the full socio-technical feedback cycle: how narrative analytics feedback is interpreted and acted upon by learners over time, how teachers appropriate orchestration dashboards as decision-support tools, and what conditions enable or constrain sustained adoption under real classroom constraints. AI-mediated epistemic interaction is examined as a bounded exploratory strand when AI features are introduced in Year 3. This is the largest and most complex study, running across Years 2 and 3. The study is structured around a primary focus on feedback uptake and teacher orchestration.

Narrative feedback design. The system generates short, evidence-grounded prompts based on validated indicators from Study 2. Each prompt follows the three-part dialogic structure defined in Section 5.4: it identifies a salient feature in the student's recent work, interprets that feature in relation to the inquiry-quality practice it represents, and proposes a specific, feasible next action. For example: "Your last two entries described what you observed but did not connect those observations to specific measurements.

Your nitrate data from Monday and Wednesday showed a drop. Try linking your claim about fish behaviour to those readings and explaining what the connection might be."

Feedback is visible to both student and teacher and designed to be comprehensible without teacher mediation, though mediation is expected and studied as a mechanism. Feedback does not assign scores, grades, or rankings and does not compare students to each other.

Feedback interpretation calibration. Early in Study 3, a calibration phase explicitly examines whether students interpret feedback prompts accurately. Students are asked in brief interviews to paraphrase what a feedback prompt is telling them and what it is suggesting they do. This directly operationalises the interpret stage of the feedback literacy cycle rather than inferring interpretation solely from subsequent behaviour. Misinterpretation patterns identified during calibration inform iterative refinement of feedback language. This strengthens alignment with the feedback literacy framework by ensuring that the interpretation step is empirically examined rather than assumed.

Teacher orchestration dashboards. The dashboard presents class-level and group-level summaries of inquiry activity and indicator patterns. Following the principle that dashboards should present actionable exceptions rather than comprehensive data, teachers see which students or groups have indicators suggesting difficulty with specific practices. The dashboard highlights where teacher attention is likely to matter most rather than displaying everything about everyone.

The dashboard supports use during lessons, where a teacher may have thirty seconds to scan before a student approaches, and between lessons, where more reflective planning is possible. Both modes are studied as distinct orchestration patterns.

Learner feedback uptake tracking. Student engagement with narrative feedback is tracked through platform traces and analysed as temporal uptake sequences. For each feedback prompt, the research examines whether the student accessed the feedback, returned to it, what action they took in subsequent entries, and whether observable practice changes are evident in subsequent artefacts. This creates an uptake sequence for each student: accessed, revisited, action visible, practice change observed, with each step coded as present or absent.

Over time, individual sequences are aggregated to identify uptake typologies: patterns of feedback engagement that characterise different students or different phases of inquiry. Some students may consistently access and act on feedback from early in the year. Others may initially ignore feedback but begin engaging after teacher mediation. Others may access feedback but show no observable practice change, suggesting a gap between interpretation and action. These typologies provide a more nuanced picture of feedback literacy development than simple access rates or before-and-after comparisons.

Uptake is examined across three time windows: early engagement in the first term, mid-programme around the end of the second term, and late-programme in the third and fourth terms. Changes across windows provide evidence about whether feedback literacy develops over time or remains static.

Teacher orchestration observation. Teacher interaction with the dashboard is observed during lessons and documented through structured field notes, screen recordings where feasible, and brief post-lesson interviews. Each observation captures orchestration episodes: moments where the teacher consulted the dashboard, interpreted what they saw, made an instructional decision, and what the observable consequence was for student activity. Episodes are coded using a structured scheme capturing the analytic cue noticed, the interpretation made, the instructional move chosen, timing relative to lesson activity, and perceived outcome.

Observations at multiple points across the school year capture how orchestration evolves. Early-year observations examine initial encounter and sensemaking. Mid-year observations examine whether routines have developed and how they function under typical conditions. Late-year observations examine whether use has been sustained, intensified, or abandoned, and what factors explain the trajectory.

Workload documentation. Teacher workload is documented through three triangulated sources. Brief workload diary entries at regular intervals record time spent on analytics-related activities and perceived impact on teaching. Platform usage logs provide behavioural data on when and how long teachers access the dashboard, independently of diary compliance. Teacher event markers embedded in the platform allow teachers to flag moments when the dashboard prompted a specific action or when they chose not to use it and why. If diary compliance degrades over time, as is realistic, the behavioural data from logs and event markers continues to provide evidence about usage patterns and workload integration.

AI micro-scaffolds: bounded exploratory strand. In Year 3, Platform Version C introduces optional AI micro-scaffolds designed to support bounded comparison. These are short prompts suggesting specific comparisons based on available data, for example: "You have nitrate readings from three different weeks. Have you compared how they changed after you adjusted the feeding schedule?" These scaffolds are optional. Students choose whether to use them. They do not replace narrative feedback.

This strand examines uptake patterns, whether scaffolds are used selectively or routinely, whether they support or distort reasoning, and how students evaluate their usefulness and trustworthiness. If scaffolds introduce problems such as encouraging uncritical acceptance rather than reasoned judgement, the feature is rolled back and the rollback documented as evidence about AI integration limits. The study explicitly examines whether AI-supported comparison strengthens reasoning or inadvertently encourages epistemic outsourcing, where students defer to AI-generated suggestions rather than exercising their own interpretive judgement. This connects to the theoretical framing of AI as bounded epistemic mediator developed in Section 5.2.

Optional translational support tools (exploratory). In later iterations of GrowHub, an optional AI translational layer may be introduced to assist students in converting inquiry ideas into executable artefacts. This layer does not evaluate inquiry quality, assign scores, or replace validated indicators. It supports translation tasks that frequently create friction in sustained STEM inquiry.

In environmental monitoring contexts, students often develop plausible explanations but struggle to operationalise them. A student may hypothesise that feeding schedules influence nitrate levels but lack the coding confidence to model this relationship or generate structured comparisons across time periods. Similarly, students designing sensor prototypes or physical models may have conceptual clarity but difficulty translating that clarity into code blocks, data structures, or printable formats.

These tools function as provisional supports rather than authoritative instructors. They may suggest code structures based on described logic, propose structured comparisons from available datasets, convert design descriptions into modifiable templates, generate alternative interpretations to prompt epistemic reflection, or surface potential variables not yet considered. All outputs are explicitly labelled as suggestions requiring student verification. Interaction traces are logged to examine whether suggestions increase revision depth, improve comparative reasoning, distort uncertainty calibration, or lead students to treat outputs as authoritative answers rather than provisional tools to be tested.

Teacher mediation remains central. Teachers observe AI usage patterns through dashboard indicators and intervene where epistemic outsourcing or over-reliance is observed. If AI integration introduces epistemic distortion, reduces student reasoning agency, or conflicts with governance constraints, the feature is withdrawn and the withdrawal documented as evidence about the limits of AI augmentation in sustained inquiry. This exploratory extension asks whether students treat AI as a tool or as a collaborator, and how that distinction influences the quality of their reasoning. It is bounded and does not bear the weight of the thesis.

Adoption conditions analysis. The final strand synthesises evidence across all data sources to identify conditions under which analytics-enhanced inquiry is sustained, adapted, or abandoned. Drawing on socio-technical frameworks (Orlikowski, 2000), the analysis examines conditions at three levels. At the learner level: what feedback literacy capabilities are required, and how do they develop? At the teacher level: what orchestration routines, professional confidence, and mediation skills support sustained use? At the institutional level: what leadership support, timetabling flexibility, infrastructure reliability, and professional culture conditions enable or constrain adoption?

Governance features are examined as a specific adoption variable. The research tracks trust markers before and after governance features are introduced, examining whether trust in analytics outputs changes after students and teachers gain access to the indicator explainer and audit trail. This temporal alignment analysis allows adoption shifts to be observed without claiming causal attribution; the design traces temporal associations and seeks mechanism-level explanations through qualitative data.

The adoption analysis triangulates self-report from interviews with behavioural evidence from platform logs, observational evidence from orchestration coding, and workload evidence from diary and event-marker data. This addresses a persistent weakness in adoption research where studies rely on post-implementation surveys that may not reflect actual use patterns. The sustainability focus is central. The question is not whether the system can be used under favourable conditions during a brief pilot but whether it can be sustained under the routine pressures and competing demands that characterise everyday teaching.

Output. Study 3 produces: a learner feedback uptake typology with temporal development evidence; a teacher orchestration episode analysis with temporal evolution across the school year; a workload impact analysis based on triangulated evidence; an analysis of AI micro-scaffold and translational tool uptake and effects as a bounded subsection; and an empirically grounded adoption conditions framework identifying learner, teacher, and institutional conditions for sustained analytics use.

8. The Digital Inquiry Platform

8.1 Purpose and Position

GrowHub is the working name for the inquiry interface and analytics system that functions as the research instrument in this thesis. It is not a commercial product or a pre-existing tool being evaluated. It is a purpose-built research instrument, co-designed through Study 1, that captures inquiry process traces, computes indicators of reasoning quality, generates narrative feedback for students, and provides orchestration dashboards for teachers. GrowHub has been iterated through early programme design work and will continue to evolve through the co-design cycles in Study 1.

The thesis does not depend on GrowHub becoming a polished product. It depends on GrowHub remaining a transparent, inspectable environment where inquiry traces can be captured, indicators can be tested against human judgement, and feedback designs can be studied under the constraints of real classroom practice.

The platform is designed around the constraints identified by stakeholders and confirmed through the socio-technical analysis: it must operate on standard school devices with variable connectivity, function with offline capability, be transparent about what it tracks and computes, and be governed according to data sovereignty expectations. These are minimum viability conditions for any analytics system intended for use in New Zealand intermediate schools. GrowHub is designed for interoperability with commonly adopted school ecosystems such as Google Workspace for Education to support feasibility and minimise workflow disruption; this does not imply any existing commercial partnership or endorsement.

8.2 Core Functions

Inquiry logging. Students record observations, measurements, claims, comparisons, and reflections through structured entry templates. The templates provide scaffolding that supports entry without constraining content, prompting students to identify what they observed, what data they collected, what they think it means, and how confident they are. The templates serve a dual purpose: they support inquiry practice and generate the structured trace data the analytics require. Entry content, timing, revision history, use of data tables and comparison tools, and query patterns are captured as process traces.

Indicator computation. The platform computes indicators for each inquiry-quality practice using algorithms developed and validated in Study 2. Indicators are computed after each session and stored with artefact data. Computation logic is documented and accessible to teachers and, in age-appropriate form, to students. No hidden analytics operate within the platform.

Narrative feedback. Based on computed indicators, the platform generates narrative feedback prompts using the three-part dialogic structure described in Section 5.4. In Version B, feedback follows rule-based logic drawing on indicator patterns to select appropriate templates populated with specific references to the student's recent work. In Version C, AI micro-scaffolds and translational tools supplement this with optional comparison suggestions and design translation support. Feedback is visible to both students and teachers and timestamped in an audit trail.

Teacher dashboard. The dashboard displays class-level and group-level summaries highlighting where teacher attention is likely to have the greatest impact. It follows the actionable-exception principle: surfacing cases where indicators suggest difficulty or where patterns have changed notably, rather than displaying comprehensive data. The dashboard supports both during-lesson scanning and between-lesson reflection, and both modes are studied in Study 3.

8.3 Platform Versioning

Version A (Year 1) provides inquiry logging with structured templates, basic data visualisation, and initial indicator computation for pilot testing. No narrative feedback or teacher dashboards. Supports baseline data collection and coding protocol development.

Version B (Year 2) adds validated indicator computation, narrative feedback generation, and teacher orchestration dashboards. This is the primary research version for Studies 2 and 3.

Version C (Year 3) adds optional AI micro-scaffolds for bounded comparison and translational support tools. If AI features prove problematic, Version B continues as primary deployment and the rollback is documented as a finding. The versioning structure ensures the core analytics and feedback system is established and validated before AI features are introduced.

8.4 Design Evolution Through Programme Development

The current platform design reflects lessons from programme planning and early classroom observation. Interface decisions have been driven by anticipated and observed friction rather than feature ambition. Templates have been simplified to reduce confusion. Offline data capture has been prioritised given unreliable connectivity during field-based sessions. Dashboard displays have been designed to avoid information overload during lesson transitions. Feedback prompts have been kept short and structured after observing that longer narrative blocks are likely to be skimmed rather than read by this age group. These design decisions will be formally tested and refined through the co-design process in Study 1 and the structured pilots in 2026. The platform that enters the formal research phase carries the imprint of authentic programme development but treats all design decisions as hypotheses to be examined rather than conclusions already reached.

9. Quality and Rigour

9.1 Approach to Quality

Design-based research does not lend itself to the quality criteria of controlled experimentation. The research operates in authentic settings where variables cannot be fully controlled, participants are not randomly assigned, and the intervention itself evolves across iterations. Quality is established through transparency, systematic evidence collection, analytical discipline, honest reporting of what the research can and cannot support, and the reflexive safeguards described in Section 7.3.

9.2 Construct Validity

The analytics indicators are validated against a theoretically grounded construct through convergent, discriminant, and consequential validity analysis as detailed in Section 5.6. Rubric dimensions are defined at the level of observable practices with concrete behavioural anchors. Inter-rater reliability must meet minimum thresholds before validation proceeds. Because analytics indicators can create seductive but misleading proxies, this thesis treats weak validity as an informative outcome. Indicator limits, confounds, and failure modes are reported with the same analytical attention as successful indicators.

9.3 Transparency and Replicability

All coding protocols, rubric definitions, indicator computation logic, validation procedures, and failure analyses are documented in sufficient detail for other researchers to scrutinise, critique, and replicate. Where indicators fail, failure modes are classified and reported. Where design decisions involve trade-offs, the trade-offs are documented. The research claims to have developed and tested one analytics system transparently and reported what was found.

9.4 Triangulation

Key findings are supported by multiple data sources. Feedback uptake is examined through platform traces, student interviews, and artefact analysis. Orchestration is examined through dashboard logs, observation field notes, and post-lesson interviews. Workload impact is examined through diary entries, platform usage data, and event markers. Adoption conditions are identified through convergence across evidence streams. Where streams diverge, the divergence is reported and interpreted rather than suppressed.

9.5 Ecological Validity

The research is conducted in authentic classroom conditions with real students, real teachers, real time constraints, and real competing demands. This ecological validity is a deliberate strength. It means findings about indicator performance, feedback uptake, and adoption conditions are grounded in realistic conditions rather than idealised laboratory settings. It also means some data will be messy, some participation uneven, and some features used in ways the design did not anticipate. Design-based research treats these realities as data about how interventions function under authentic conditions, not as methodological contamination.

9.6 Analytical Honesty

The design cannot establish causal relationships. Maturation, teacher learning, curriculum effects, and other factors co-occur with the intervention and cannot be separated from it. The research addresses this through mechanism tracing rather than causal claims, examining whether specific feedback types are associated with specific practice changes, whether timing aligns with feature introduction, and whether the mediating processes specified in the conjecture map are observed. Claims are bounded to what the evidence supports. Where findings are ambiguous, the ambiguity is reported.

10. Limitations

Single school context. Findings are grounded in one school with a specific demographic, cultural, and institutional profile. The design model and constraints taxonomy are intended to be transferable as adaptable design patterns, but they are not tested across multiple sites within this thesis.

Quasi-experimental design. No random assignment or control conditions. Maturation effects cannot be fully separated from intervention effects. The study traces mechanisms and associations, not causal relationships.

Researcher dual role. I am both the teacher in the classroom and the researcher conducting the study. This creates potential observer effects, conflicts of interest, and demand characteristics. These are managed through the reflexive safeguards detailed in Section 7.3, university supervision of all research decisions, independent coding for validation, and honest reporting of the dual role's potential influence on findings. The dual role also provides genuine advantages: sustained access, deep contextual knowledge, and the ability to observe analytics functioning across an entire school year rather than during a brief research visit. Both the risks and advantages are acknowledged.

Small teacher sample. Fewer teachers produce fewer but deeper cases. Claims about teacher-level patterns are bounded accordingly. The study identifies conditions and mechanisms rather than generalisable teacher profiles.

Indicator ceiling. Some aspects of inquiry quality may not be computationally detectable from trace data. The quality of an argument, the originality of an interpretation, and the depth of engagement with uncertainty may require human judgement that no algorithm can replicate. The study explores this boundary honestly. Where trace-based indicators reach their limits, that is a finding, not a failure.

Template-induced measurement artefacts. Structured entry templates may scaffold the very practices the indicators claim to detect, creating the risk that indicators partly measure template compliance rather than spontaneous reasoning. This is monitored through validation and reported as a boundary condition of the measurement approach.

Infrastructure dependence. Despite lightweight design, the platform depends on functional devices and at least intermittent connectivity. If sustained infrastructure failure occurs, paper-based scaffolds preserve inquiry activities and human coding, but platform-based trace data would be unavailable for the affected period.

Cultural specificity. Governance mechanisms are developed in a specific bicultural context shaped by Te Tiriti obligations and the values of the Whanganui community. The model may not transfer directly to other cultural settings, though the co-design process and the principle of community-governed analytics are intended to be adaptable.

Affective dimensions. Feedback literacy includes emotional regulation, which this study captures indirectly through interviews and observation rather than through a dedicated affect instrument. The affective pathway through which analytics feedback influences learner engagement is acknowledged but not fully modelled.

Pilot contexts not yet tested. The aquaponics and light-trap biodiversity substrates are planned for 2026 deployment and have not yet generated research data at the time of writing. Design decisions informed by formative programme development observations are treated as hypotheses to be tested, not as established findings.

Human-AI interaction scope. The AI investigation is bounded and exploratory. Findings about AI mediation are contextually bounded and represent an opening investigation rather than a comprehensive model of human-AI epistemic interaction.

11. Scope, Sequence, and Timeline

11.1 Scope

The scope is defined by its focus on the design, validation, and study of human-centred learning analytics for sustained inquiry in an intermediate school context. The research produces three linked outputs: a design model and constraints taxonomy, validated analytics indicators with transparent reporting of detection limits, and longitudinal evidence on feedback uptake, teacher orchestration, and adoption conditions. AI-mediated epistemic interaction is examined as a bounded exploratory strand within the third output.

The research context is the STEM inquiry programme at Whanganui Intermediate, with two environmental inquiry substrates providing contrasting epistemic stress tests. GrowHub is the analytics system being designed and evaluated. The programme may include additional activities as part of normal teaching, but only the two substrates serve as research contexts.

11.2 Sequence

Phase 1 (Year 1) involves co-design, foundations, and pilot. The STLP placement provides protected time for scientific grounding, field methodology development, and careful examination of how authentic ecological monitoring practices translate into classroom inquiry. Following return to teaching, co-design workshops are conducted, ethics approval is finalised, and Platform Version A is deployed. Inquiry substrates are piloted in Terms 3 and 4. Baseline data are collected. Study 1 is completed. Candidature confirmation occurs by year end.

Phase 2 (Years 2 to 3) involves full programme delivery, indicator validation, and feedback and orchestration analysis. Version B is deployed with validated indicator computation, narrative feedback, and teacher dashboards. Study 2 proceeds with artefact collection, human coding, and indicator validation. Study 3 begins with feedback uptake tracking, orchestration observation, and workload documentation. In Year 3, Version C introduces AI features as a bounded exploratory strand. Study 3 continues across the full year. Papers are drafted and submitted.

Phase 3 (Year 4) involves synthesis, writing, and dissemination. The thesis integrates findings across studies. Final publications are completed. Practitioner resources are created. Community reports are shared with programme partners.

11.3 Timeline and Milestones

The PhD is undertaken on a full-time equivalent basis across four years, with research embedded within ongoing teaching and programme delivery. This provides continuous access to participants, authentic inquiry environments, and sustained iterative design opportunities. It also means the research is constantly tested against real classroom constraints.

Year 1 (2026). The STLP placement in early 2026 provides protected time for scientific grounding, system construction, and field methodology. Co-design workshops are conducted. Ethics is finalised. Version A is deployed. Substrates are piloted in Terms 3 and 4. Baseline artefacts are collected. Coding protocols are tested. Study 1 is completed. Candidature confirmation is supported by research framing, literature review, co-design outputs, and pilot evidence.

Year 2 (2027). Full programme delivery begins with both substrates operating across terms. Version B is deployed. Study 2 proceeds: artefacts are collected, human coding is conducted, indicator validation begins. Study 3 begins: feedback uptake is tracked, orchestration is observed, workload is documented. Progress assessment is supported by validation findings, draft chapters, and conference presentations.

Year 3 (2028). Version C introduces AI features as bounded exploratory strand. Study 2 is extended for cross-version analysis. Study 3 continues with deepened orchestration analysis, AI interaction investigation, and adoption conditions analysis. Journal articles are submitted. Thesis readiness review is supported by completed findings chapters and case studies.

Year 4 (2029). The thesis integrates design principles, validation findings, and uptake and adoption evidence. Final publications are submitted. Practitioner resources including design guidelines and exemplar dashboards are created. Community reports are shared. Thesis is submitted and examined.

12. Risk Mitigation

The research is designed to remain viable under realistic disruptions. The mitigations below reflect both anticipated risks and lessons from programme development work to date.

One inquiry context proves infeasible. The research continues with a single context. Cross-context analysis is reframed as a robustness check. The thesis remains viable because indicator validation, feedback uptake, and orchestration study can proceed within a single sustained inquiry environment.

Platform development is delayed. Paper-based scaffolds using structured templates preserve Studies 1 and 2. Human coding and co-design proceed independently of platform functionality. Study 3 trace-based analysis would be limited but interview and observation data continue.

AI features show low uptake or introduce problems. Low uptake is a finding about the conditions under which AI mediation is accepted or resisted by younger learners. If AI features introduce misconceptions, encourage epistemic outsourcing, or distort inquiry, the feature is rolled back and the rollback documented as evidence about AI integration limits. Core Studies 1 and 2 operate independently of AI features.

Indicator validation produces weak results. Indicators that fail are reported transparently with diagnostic analysis of failure modes. The contribution shifts from validated indicators to mapping detection limits and failure modes. This remains a meaningful and publishable contribution. The thesis is designed so that mapping what trace-based indicators cannot detect in sustained K to 12 inquiry advances the field regardless of whether indicators succeed.

Limited teacher participation. Fewer teachers produce deeper cases. Small numbers suit case study logic. Claims are bounded accordingly.

Low dashboard use during lessons. Non-use is a valid pathway for analysis: why teachers do not use dashboards during lessons, when they use them in other contexts, and what alternatives they employ. The dashboard includes preparation and reflection modes supporting between-lesson use.

Maturation effects. The quasi-experimental design cannot fully control for maturation, teacher learning, or curriculum effects. The research addresses this through mechanism tracing, examining whether specific feedback types associate with specific practice changes and whether timing aligns with feature introduction.

Workload diary degradation. Triangulation with platform usage logs and teacher event markers provides behavioural data independent of diary compliance.

Researcher dual role. Managed through the reflexive safeguards described in Section 7.3, university supervision, and honest reporting. Academic control of all research decisions remains with the candidate under supervision.

Template-induced artefacts. Monitored through discriminant validity analysis and qualitative examination of whether indicator scores correlate with template compliance rather than genuine reasoning variation. Reported as a boundary condition finding.

Pilot contexts untested at time of writing. The aquaponics and light-trap substrates have not yet been deployed as research contexts. If early pilots reveal that one substrate does not generate sufficient epistemic variation or sustained engagement to support analytics validation, the research adapts by deepening analysis within the functioning context.

Some cohorts will be messier than ideal. This is expected. Design-based research treats messiness as data about how interventions function under realistic conditions, not as methodological contamination.

13. Ethics and Governance

13.1 Ethical Approval

Ethics approval will be sought from the relevant university human research ethics committee and relevant New Zealand authorities prior to data collection.

13.2 Data Security

Data will be stored on university-controlled or school-approved secure cloud infrastructure with encryption at rest and in transit. Access roles are defined: students see their own data and feedback; teachers see class and group summaries; researchers see pseudonymised data only. Raw identifying data is stored separately. Retention follows ethics conditions. Deletion is available on request.

13.3 Data Governance as Adoption Mechanism

Decisions about what the platform captures, who sees it, and how it is shared are made collaboratively during co-design. Data ownership remains with students and whānau. Cross-site sharing requires opt-in with teacher moderation.

Governance protocols include an indicator explainer documenting what is tracked and why, data sharing controls specifying who sees what, and an audit trail recording what feedback was generated and when. These are designed as both ethical obligations and testable adoption mechanisms. Study 3 examines whether access to explainable indicators and auditable feedback trails is associated with different trust and use patterns. Governance is reviewed annually through consultation hui with whānau and community.

13.4 Consent

Consent processes use visual explanations, platform demonstrations, and hui rather than relying solely on written forms. Co-design in Study 1 identifies culturally appropriate approaches respecting both university requirements and community expectations. The goal is genuinely informed consent rather than procedural compliance.

13.5 Transparency

No hidden analytics operate within the platform. Students can see what is tracked and how. All AI features are clearly labelled. AI outputs are explicitly presented as suggestions requiring verification, not authoritative assessments. Indicator computation logic is documented and accessible.

13.6 Cultural Responsiveness

Specific mechanisms emerge from co-design and are documented as Study 1 findings. The thesis contributes the process and framework rather than claiming to have resolved all challenges. Structural safeguards include annual governance hui, defined access policies, and exit policies including deletion on request. The study acknowledges that it operates within a single school context and that the governance model may need adaptation in other communities.

13.7 Researcher Position

The candidate is employed as a teacher at Whanganui Intermediate School and is a director of Regenpreneur Ltd. Academic control of the PhD research remains with the candidate under university supervision. Neither the school nor any partner organisation determines research questions, controls analysis, or holds authority over publication.

13.8 Third-Party Involvement

Whanganui Intermediate School provides the educational setting and student access. UCOL provides scientific and technical advisory support. Bushy Park Tarapuruhi provides ecological site access. Skills Group provides advisory input on credential alignment. EdTechNZ provides network access and sector positioning. Regenpreneur Ltd provides administrative support. Horizons Regional Council supports alignment with regional biodiversity and stewardship goals. All organisations are involved in advisory or site-support capacities only. Partners inform context and feasibility; they do not set research questions, control analysis, or determine publication. There is potential for collaborative publication of substrate feasibility findings where appropriate, but all PhD research decisions remain under university supervision.

14. Publication Plan

Three core papers are aligned to the three studies.

Paper A (Study 1): Human-Centred Analytics Design for Sustained School Inquiry. Reports the co-design model, constraints taxonomy, ethics-by-design integration, governance as adoption mechanism, and design patterns for narrative feedback and orchestration dashboards. Target venues: LAK Conference, Journal of Learning Analytics, British Journal of Educational Technology.

Paper B (Study 2): Validating Trace-Based Indicators of Epistemic Reasoning Quality Across Contrasting Inquiry Contexts. Reports indicator validation, cross-context comparison, discriminant validity analysis, and diagnostic failure analysis for indicators that do not meet validation thresholds. Target venues: Journal of Learning Analytics, Computers and Education, LAK Conference.

Paper C (Study 3): From Indicator to Action: Feedback Uptake, Teacher Orchestration, and Sustained Adoption in Analytics-Supported Inquiry. Reports the learner feedback uptake typology, teacher orchestration episode analysis, workload impact, AI feature uptake as a bounded subsection, and adoption conditions framework. Target venues: Computers and Education, British Journal of Educational Technology, Journal of Computer Assisted Learning.

15. Expected Contributions

Validated learning analytics indicators for epistemic reasoning quality in sustained school inquiry. Including transparent reporting of what indicators detect, what they miss, what requires calibration, and what remains dependent on human judgement. Where indicators fail, diagnostic analysis of failure modes maps the boundaries of computational detection of epistemic practices in ways the field currently lacks.

Evidence on feedback literacy development during sustained inquiry among school-age learners. Extending feedback analytics beyond higher education into longitudinal school-based inquiry with 11 to 13-year-olds. Examining how narrative feedback is interpreted, trusted, mediated by teachers, and acted upon, and how patterns change over time.

A human-centred analytics design model for school inquiry. Grounded in co-design with students, teachers, and culturally diverse communities, including a constraints taxonomy, governance mechanisms designed as both ethical safeguards and adoption levers, ethics-by-design integration, and transferable design patterns documented for adaptation rather than replication.

An empirically grounded adoption conditions framework. Identifying learner, teacher, and institutional conditions for sustained analytics use, based on orchestration analysis, temporal dashboard interaction, and triangulated workload evidence. Moving beyond self-report to examine how analytics become or fail to become part of professional practice.

16. SWOT Analysis

This analysis shaped design decisions and is included because the forces it identifies are studied empirically in Study 3.

Strengths. Real classroom embedding with sustained access across the full school year provides rare ecological validity for learning analytics research. The national stakeholder survey pre-anchors design constraints in practitioner perspectives rather than researcher assumptions. Two epistemically contrasting inquiry contexts provide analytic leverage for cross-context indicator comparison. The ethics-by-design approach addresses a gap in existing analytics research. Structured contextual analyses ground the design in empirically identified forces. The researcher's dual role as teacher provides deep contextual knowledge and continuous access to participants.

Weaknesses. The single-school context limits generalisability of findings. The teacher-researcher dual role introduces potential observer effects and conflicts of interest, managed through reflexive safeguards and university supervision. Dependence on platform reliability and classroom routines means technical failures could disrupt data collection. Potential template-induced artefacts and volume confounds may limit what indicators can claim to measure. The pilot substrates have not yet been deployed as research contexts. The AI interaction investigation is bounded and exploratory.

Opportunities. GrowHub as an interoperable inquiry layer compatible with Google Workspace ecosystems supports feasibility without requiring vendor partnership. The diagnostic analysis of indicator failure modes is publishable regardless of whether indicators succeed. The co-design process and governance framework have potential for adaptation to other school contexts and cultural settings internationally. The AI interaction strand addresses a contemporary gap in human-AI interaction research in K to 12 settings. The ethics-by-design integration contributes where existing research is largely aspirational.

Threats. Shifting AI ethics and regulatory expectations may reduce school willingness to adopt analytics. Rapid changes in AI tool availability and cost may threaten reliance on specific third-party services; mitigation through bounded, optional, rollbackable AI features. Misinterpretation of dashboards as surveillance threatens trust; mitigation through transparent indicator explainer, student visibility, audit trail, and no ranking or automated judgement. Programme sustainability is threatened if workload increases rather than decreases; mitigation through explicit workload measurement and dashboard design focused on reducing rather than adding decision load. One or both pilot substrates may not generate sufficient epistemic variation; mitigation through single-context viability of the thesis design.

17. Conclusion

Sustained inquiry is where epistemic reasoning quality matters most and where the classroom feedback burden is heaviest. Learning analytics has largely avoided measuring the quality of reasoning in school inquiry, and the field has limited evidence about how feedback is interpreted and sustained in schools over time. This thesis takes up those gaps directly.

The research defines inquiry process quality as four observable epistemic and self-regulated practices. It develops transparent trace-based indicators and tests what they can and cannot detect across two epistemically contrasting contexts. It studies how narrative analytics feedback and orchestration dashboards are taken up through learner feedback literacy, teacher mediation, and institutional conditions over time. In a bounded exploratory strand, it examines how AI-mediated epistemic scaffolds influence reasoning and whether students treat AI as a provisional tool or an authoritative source.

The three studies form a coherent chain from co-design through validation through sustained use. If the indicators work well, the thesis provides validated tools others can build on. If they work partially, it maps the boundaries of what trace data can capture and where human judgement remains essential. If adoption proves fragile, it identifies the conditions that would need to be in place. Each of these outcomes contributes to the field.

The proposal integrates stakeholder-informed foundations, socio-technical constraint-mapping, and contextual analysis alongside the core analytics research because analytics systems in schools operate within nested constraints that determine whether they survive contact with real classrooms. Studying the analytics without studying the forces that shape whether anyone actually uses them would produce elegant research that changes nothing.

This thesis does not claim to resolve the challenges of measuring reasoning or sustaining analytics in classrooms. It seeks to contribute careful empirical evidence from one sustained, real-world context to a broader conversation about what analytics can and cannot do for inquiry learning in schools. The findings will be trustworthy precisely because they are honest about what was found and what was not.

This is an appropriate and achievable scope for a doctoral thesis, and the design is structured to deliver it.

18. References

Anderson, T., and Shattuck, J. (2012). Design-based research: A decade of progress in education research? Educational Researcher, 41(1), 16 to 25.

Barab, S. A., and Squire, K. D. (2004). Design-based research: Putting a stake in the ground. The Journal of the Learning Sciences, 13(1), 1 to 14.

Barron, B., and Darling-Hammond, L. (2008). Teaching for meaningful learning: A review of research on inquiry-based and cooperative learning. Edutopia/George Lucas Educational Foundation.

Bates, T., Cobo, C., Marino, O., and Wheeler, S. (2020). Can artificial intelligence transform higher education? International Journal of Educational Technology in Higher Education, 17(1), 1 to 12.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., and Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. Educational Psychologist, 26(3 to 4), 369 to 398.

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., and Oxley, V. (2024). A meta-systematic review of AI in higher education: A call for increased ethics, collaboration, and rigour. International Journal of Educational Technology in Higher Education, 21(4), Article 4.

Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J., and Parrish, J. K. (2014). Next steps for citizen science. Science, 343(6178), 1436 to 1437.

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77 to 101.

Braun, V., and Clarke, V. (2021). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qualitative Research in Sport, Exercise and Health, 13(2), 201 to 216.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. The Journal of the Learning Sciences, 2(2), 141 to 178.

Carless, D., and Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. Assessment and Evaluation in Higher Education, 43(8), 1315 to 1325.

Chinn, C. A., Buckland, L. A., and Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. Educational Psychologist, 46(3), 141 to 167.

Chinn, C. A., and Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. Science Education, 86(2), 175 to 218.

Cohen, L., Manion, L., and Morrison, K. (2018). Research methods in education (8th ed.). Routledge.

Collins, A. (1992). Toward a design science of education. In E. Scanlon and T. O'Shea (Eds.), New directions in educational technology (pp. 15 to 22). Springer.

Crawford, B. A. (2014). From inquiry to scientific practices in the science classroom. In N. G. Lederman and S. K. Abell (Eds.), Handbook of research on science education (Vol. 2, pp. 515 to 541). Routledge.

Creswell, J. W., and Plano Clark, V. L. (2017). Designing and conducting mixed methods research (3rd ed.). SAGE.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. Educational Researcher, 32(1), 5 to 8.

Dillenbourg, P. (2013). Design for classroom orchestration. Computers and Education, 69, 485 to 492.

Drachsler, H., and Greller, W. (2016). Privacy and analytics: It's a DELICATE issue. In Proceedings of the 6th International Conference on Learning Analytics and Knowledge (pp. 89 to 98). ACM.

Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., and Buckingham Shum, S. (2018). Driving data storytelling from learning design. In Proceedings of the 8th International Conference on Learning Analytics and Knowledge (pp. 131 to 140). ACM.

Floridi, L., and Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1).

Gasevic, D., Dawson, S., and Siemens, G. (2015). Let's not forget: Learning analytics are about learning. TechTrends, 59(1), 64 to 71.

Hattie, J., and Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81 to 112.

Hmelo-Silver, C. E., and Azevedo, R. (2006). Understanding complex systems: Some core challenges. Journal of the Learning Sciences, 15(1), 53 to 61.

Hmelo-Silver, C. E., Duncan, R. G., and Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). Educational Psychologist, 42(2), 99 to 107.

Hofer, B. K., and Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. Review of Educational Research, 67(1), 88 to 140.

Holmes, W., and Tuomi, I. (2022). State of the art and practice in AI in education. European Journal of Education, 57(4), 542 to 570.

Jacobson, M. J., and Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. Journal of the Learning Sciences, 15(1), 11 to 34.

Kelly, G. J. (2008). Inquiry, activity, and epistemic practice. In R. Duschl and R. Grandy (Eds.), Teaching scientific inquiry: Recommendations for research and implementation (pp. 99 to 117). Sense Publishers.

Knight, S., Buckingham Shum, S., and Littleton, K. (2014). Epistemology, assessment, pedagogy: Where learning meets analytics in the middle space. Journal of Learning Analytics, 1(2), 23 to 47.

Knight, S., and Littleton, K. (2015). Discourse-centric learning analytics: Mapping the terrain. Journal of Learning Analytics, 2(1), 185 to 209.

Knight, S., Wise, A. F., and Chen, B. (2017). Time for change: Why learning analytics needs temporal analysis. Journal of Learning Analytics, 4(3), 7 to 17.

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159 to 174.

Long, D., and Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1 to 16). ACM.

Lundqvist, K., Persson, A., and Andersson, R. (2023). Bias, equity, and cultural sustainability in AI-mediated education. Computers and Education, 194, 104662.

Macfarlane, A., Macfarlane, S., and Gillon, G. (2017). Culturally responsive evidence-based practice in education. International Journal of Inclusive Education, 21(11), 1071 to 1086.

Manz, E. (2015). Resistance and the development of scientific practice: Designing the mangle into science instruction. Cognition and Instruction, 33(2), 89 to 124.

Martinez-Maldonado, R., Echeverria, V., Fernandez Nieto, G., and Buckingham Shum, S. (2020). From data to insights: A layered storytelling approach for multimodal learning analytics. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1 to 15). ACM.

McKenney, S., and Reeves, T. C. (2019). Conducting educational design research (2nd ed.). Routledge.

Meadows, D. H. (2008). Thinking in systems: A primer. Chelsea Green.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50(9), 741 to 749.

Ministry of Education. (2007). The New Zealand curriculum. Learning Media.

OECD. (2025). What should teachers teach and students learn in a world of powerful AI? OECD Education Spotlights, No. 20. OECD Publishing.

Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. Organization Science, 11(4), 404 to 428.

Phillips, T. B., Ballard, H. L., Lewenstein, B. V., and Bonney, R. (2019). Engagement in science through citizen science: Moving beyond data collection. Science Education, 103(3), 665 to 690.

Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. Science Education, 89(4), 634 to 656.

Sandoval, W. A. (2014). Conjecture mapping: An approach to systematic educational design research. Journal of the Learning Sciences, 23(1), 18 to 36.

Selwyn, N., Hillman, T., Eynon, R., Ferreira, G., Knox, J., Macgilchrist, F., and Sancho-Gil, J. M. (2020). What's next for Ed-Tech? Critical hopes and concerns for the 2020s. Learning, Media and Technology, 45(1), 1 to 6.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. American Behavioral Scientist, 57(10), 1380 to 1400.

Siemens, G., and Long, P. (2011). Penetrating the fog: Analytics in learning and education. EDUCAUSE Review, 46(5), 30 to 32.

Smith, L. T. (1999). Decolonizing methodologies: Research and indigenous peoples. Zed Books.

Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645 to 3650).

Te Mana Raraunga. (2018). Principles of Māori data sovereignty. Te Mana Raraunga.

Thomas, J. W. (2000). A review of research on project-based learning. Autodesk Foundation.

Tsai, Y.-S., and Gasevic, D. (2017). Learning analytics in higher education: Challenges and policies. In Proceedings of the 7th International Learning Analytics and Knowledge Conference (pp. 233 to 242). ACM.

Tsai, Y.-S., Whitelock-Wainwright, A., and Gasevic, D. (2020). The privacy paradox and its implications for learning analytics. In Proceedings of the 10th International Conference on Learning Analytics and Knowledge (pp. 230 to 239). ACM.

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO Publishing.

Weick, K. E. (1995). Sensemaking in organizations. SAGE.

Williamson, B. (2018). The hidden architecture of higher education: Building a big data infrastructure for learning. Learning, Media and Technology, 43(3), 281 to 295.

Winne, P. H. (2017). Leveraging big data to help each learner and accelerate learning science. Teachers College Record, 119(3), 1 to 24.

Winne, P. H., and Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, and A. C. Graesser (Eds.), Metacognition in educational theory and practice (pp. 277 to 304). Erlbaum.

Winstone, N. E., and Carless, D. (2019). Designing effective feedback processes in higher education: A learning-focused approach. Routledge.

Wise, A. F., and Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. Journal of Learning Analytics, 2(2), 5 to 13.

Zawacki-Richter, O., Marin, V. I., Bond, M., and Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? International Journal of Educational Technology in Higher Education, 16(1), Article 39.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory into Practice, 41(2), 64 to 70.