Available online at: https://jurnal-fkip.ut.ac.id/index.php/ijrse/issue/archive
International Journal of Research in STEM Education (IJRSE)
ISSN 2721-2904 (online)
Volume 8 Number 1 (2026): 108 - 128

Assessment-Driven Learning Analytics: A Paradigm Shift for Equitable K–12 STEM Education

Kim, Tai Ki
Yonsei University, Seoul, South Korea
AI Psychometrics Inc., Canada
studygps@yonsei.ac.kr

Abstract

Learning analytics (LA) has transformed educational practice through data-driven personalization, yet its dependence on continuous digital trace data systematically excludes learners in low-technology K–12 environments. This paper proposes Assessment-Driven Learning Analytics (ADLA) a paradigm shift that repositions psychometric assessments as primary analytic signals rather than supplementary data. Grounded in socio-cognitive theory and computational psychometrics, ADLA operationalizes three-dimensional learner modeling (cognitive, affective, behavioral) that functions without digital infrastructure. The framework also advances a theoretical alternative to digital-trace dependency by positioning psychometric assessment as a primary analytics infrastructure. We articulate theoretical foundations distinguishing ADLA from clickstream-based approaches, specify detailed operational frameworks with validated instruments, outline implementation architecture, propose comprehensive research validation agenda, and examine policy implications for educational equity and SDG 4. The framework further generates testable propositions concerning predictive validity, learner profiling, analytics equity, and model interpretability. ADLA demonstrates that meaningful learning analytics need not depend on technological abundance, offering methodological pathways toward inclusive, evidence-based STEM education accessible to all students regardless of infrastructural constraints.

Keywords: *STEM , Learning Analytics, Paradigm shift, Psychometric assessment, K–12*

INTRODUCTION

Learning analytics (LA), defined as "measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning" (Siemens & Gasevic, 2012, p. 1), has become central to data-informed STEM education. Contemporary LA enables identification of misconception patterns, trajectory modeling, and adaptive interventions supporting complex conceptual development (Chen et al., 2023; von Davier, 2023). However, LA systems exhibit fundamental architectural dependency: they require continuous digital interaction data to function (Baker & Inventado, 2023).

This dependency creates systematic exclusion. While 82% of U.S. public schools report device access, actual integration varies profoundly by socioeconomic status and geography, with under-resourced districts experiencing fragmented digital footprints insufficient for analytics (NCES, 2023; West et al., 2022). Globally, the majority of K–12 students attend schools with minimal digital infrastructure (UNESCO, 2023). As LA becomes central to educational improvement, this analytics divide risks exacerbating inequities.

Holstein et al. (2023) document how clickstream-based systems create "analytics deserts" in under-resourced schools, systematically excluding students not because they learn differently but because infrastructure privileges digital interaction as the sole epistemological pathway. Liu et al. (2023) emphasize that substantial pedagogically valuable K–12 STEM activity—hands-on experimentation, collaborative problem-solving, verbal discourse—occurs offline, generating no digital trace yet being central to learning.

Corresponding author:
Kim, Tai Ki, studygps@yonsei.ac.kr,
DOI: https://doi.org/10.33830/ijrse.v8i1.14355

This paper proposes Assessment-Driven Learning Analytics (ADLA) —a paradigm fundamentally reconceptualizing assessment-analytics relationships. Rather than treating psychometric instruments as supplementary, ADLA positions structured assessments—cognitive diagnostics, affective inventories, behavioral observations—as primary, intentional data-generating mechanisms enabling comprehensive learner modeling without digital dependency.

Beyond introducing a new framework, this paper contributes to learning analytics theory in three ways. First, it challenges the dominant assumption that meaningful analytics must originate from continuous digital trace data. Second, it reconceptualizes psychometric assessment as an intentional analytics infrastructure rather than a peripheral evaluative mechanism. Third, it extends equity discourse in learning analytics by proposing a technology-independent pathway for learner modeling. Collectively, these contributions position ADLA not merely as a practical alternative but as a theoretical expansion of what constitutes legitimate evidence within learning analytics ecosystems.

At its core, ADLA advances a fundamental epistemological claim: learner understanding should be inferred primarily from theoretically grounded evidence intentionally designed to represent learning constructs, rather than exclusively from behavioral residues generated through digital interaction. This proposition challenges a prevailing assumption within contemporary learning analytics and invites reconsideration of what constitutes valid evidence for educational decision-making.

LITERATURE REVIEW

The Digital Trace Dependency Problem

Contemporary LA research concentrates in higher education and MOOCs where continuous digital interaction generates abundant behavioral data (Viberg et al., 2018). This architecture—descriptive analytics visualizing engagement, predictive analytics forecasting outcomes, prescriptive analytics recommending actions—fundamentally relies on granular clickstreams, login patterns, and resource access logs (Baker & Inventado, 2023; Sailer et al., 2021).

Holstein et al. (2023) analyzed 127 K–12 LA implementations: 89% required sustained digital platform engagement, excluding schools with intermittent technology. Liu et al. (2023) reviewed 215 K–12 LA studies: fewer than 15% addressed low-technology implementations. West et al. (2022) documented that while device availability increased, meaningful integration enabling continuous LA remains limited in rural and high-poverty districts. Current LA paradigm creates two-tiered systems where students in well-resourced schools benefit from data-informed instruction while those in under-resourced settings lack equivalent support due to infrastructural constraints.

Psychometric Assessment as Alternative Infrastructure

Educational psychology and psychometrics offer robust alternatives. Mislevy et al. (2021) argue well-designed assessments directly operationalize theoretical constructs—conceptual mastery, metacognitive awareness, self-efficacy—with established validity rather than requiring inferential leaps from behaviors. Recent advances in computational psychometrics (Chen et al., 2021; von Davier, 2020, 2023) demonstrate how item response theory (IRT), diagnostic classification models (DCMs), and Bayesian networks support sophisticated learner modeling from sparse, heterogeneous data without continuous streams.

International frameworks demonstrate mature psychometric STEM assessment. TIMSS organizes evaluation around content × cognitive process matrices, achieving high validity across 60+ countries with diverse technological contexts (Mullis & Martin, 2017). PISA measures creative thinking and AI literacy through instruments designed for paper-based/computer-based equivalence (OECD, 2023). Affective assessment has matured with validated instruments measuring self-efficacy, interest, and growth mindset predicting STEM persistence even after controlling for achievement (Duckworth & Yeager, 2023; Wang & Degol, 2024).

Behavioral observation protocols—CLASS (Pianta et al., 2008), COPUS (Smith et al., 2013)—demonstrate structured observation yields quantifiable data about engagement. Lund et al. (2024) show trained teachers achieve inter-rater reliability (ICC > .75) comparable to external researchers.

The Critical Gap

Despite advances, significant gaps persist: (1) disciplinary fragmentation prevents synthesis; (2) fewer than 15% of LA studies focus on K–12 STEM; (3) psychometric research increasingly assumes computer-based delivery, recreating digital dependency; (4) equity treated as constraint rather than design principle. ADLA directly addresses these gaps by synthesizing LA goals with psychometric rigor, developing frameworks for K–12 STEM, designing for technology-agnostic implementation, and centering equity foundationally.

Existing approaches such as Educational Data Mining, Multimodal Learning Analytics, and Hybrid Learning Analytics have expanded the range of evidence available for understanding learning. However, these approaches largely maintain digital interaction as the primary source of analytic inference. ADLA differs by reversing this logic: assessment evidence becomes the primary analytic infrastructure, while digital traces—when available—serve a complementary role.

THEORETICAL FRAMEWORK: ADLA AS PARADIGMATIC ALTERNATIVE

Epistemic Shift: From Behavioral Inference to Direct Measurement

Traditional clickstream-based LA and ADLA represent fundamentally different epistemological approaches. **Table 1** systematizes distinctions :

Table 1. Paradigmatic Comparison: Clickstream LA vs. ADLA

Dimension	Clickstream LA	Assessment-Driven LA (ADLA)
Epistemology	Indirect inference from behavior	Direct construct measurement
Primary Data	Digital traces (clicks, navigation)	Psychometric assessments
Cognitive Access	Inferred from task patterns	Diagnostic instruments (DCMs, IRT)
Affective Access	Rarely captured	Systematic validated inventories
Behavioral Access	Automated digital interactions	Structured observations, rubrics
Theoretical Ground	Often atheoretical patterns	Explicit theory-driven measurement
Technology Dependency	Requires continuous infrastructure	Functions without digital systems
Data Frequency	Continuous/high-frequency	Strategic/lower-frequency
Data Depth	Shallow (surface behaviors)	Deep (theoretically meaningful)
Validity Foundation	Assumes behavior→cognition	Established psychometric validation
Equity Implications	Excludes low-tech contexts	Inclusive across all contexts
Primary Limitation	Digital divide exclusion	Assessment literacy requirement

It is important to emphasize that ADLA is not proposed as a universal replacement for digital-trace learning analytics. Rather, ADLA is intended for contexts where digital evidence is absent, incomplete, unreliable, or insufficient for representing meaningful learning processes. In digitally mature environments, assessment-driven and trace-driven analytics may function synergistically, producing richer learner representations than either approach alone.

This reveals epistemic shift: ADLA replaces indirect behavioral inference with direct measurement of theoretically specified variables. Rather than assuming clickstream patterns proxy for engagement, ADLA measures engagement through validated observation rubrics. Rather than inferring cognitive states from task completion, ADLA employs diagnostic assessments explicitly designed to reveal conceptual understanding and misconceptions.

Sociocognitive Foundations: Integrated Learner Modeling

ADLA's theoretical foundation rests on sociocognitive learning theory (Bandura, 1986; Pintrich, 2003; Zimmerman, 2020), conceptualizing learning as emerging from dynamic interactions among: (1) personal factors (cognitive processes, affective states); (2) behavioral patterns (observable actions); (3) environmental contexts (instructional conditions, social interactions). This triadic reciprocal causation implies comprehensive understanding requires simultaneous measurement across all components.

Traditional LA emphasizes behavioral traces while treating cognitive and affective dimensions as latent variables requiring inference (Gašević et al., 2015). ADLA operationalizes integrated measurement enabling direct access to all three dimensions through established psychometric methods.

Cognitive dimension employs diagnostic assessments revealing specific skill mastery, conceptual understanding, reasoning strategies, and metacognitive processes. Drawing on cognitive diagnostic models (de la Torre & Douglas, 2004; Rupp et al., 2010), ADLA moves beyond undifferentiated achievement scores to fine-grained attribute profiles enabling targeted instruction.

Affective dimension systematically measures motivational and emotional constructs predicting STEM engagement and persistence: self-efficacy (Bandura, 1997), interest (Renninger & Hidi, 2019), achievement goals (Elliot & Hulleman, 2024), growth mindset (Dweck, 2023; Yeager et al., 2023), academic emotions (Pekrun et al., 2023). These variables are not mere correlates but mechanisms of learning—they influence cognitive resource allocation, persistence through difficulty, and help-seeking behaviors directly impacting learning quality (Chen et al., 2023).

Behavioral dimension captures observable engagement manifestations through structured observation: behavioral engagement (participation, task completion, persistence), cognitive engagement (substantive questioning, strategic problem-solving), agentic engagement (proactive learning construction, goal-setting) (Fredricks et al., 2024; Reeve & Tseng, 2023).

Psychometric integration through multidimensional IRT, structural equation modeling, and Bayesian networks simultaneously estimates all dimensions while modeling interdependencies (Liu et al., 2024; von Davier, 2023), enabling identification of learner profiles invisible in single-dimension analyses—e.g., high cognitive skills + low affective engagement + minimal behavioral participation indicates capable but unmotivated students requiring relevance-building rather than remediation.

Equity and Data Justice as Foundational Design Principles

ADLA embeds equity not merely as a peripheral constraint but as a foundational design principle. Drawing upon established data justice frameworks (D'Ignazio & Klein, 2020; Prinsloo & Slade, 2017), the ADLA framework operationalizes four critical commitments to inclusive analytics. First, it ensures **visibility** by representing all learners regardless of their technological context. By functioning independently of digital infrastructure, ADLA prevents students in under-resourced environments from being rendered invisible to data-informed support systems. Second, the framework promotes **transparency** through assessment-based data collection that remains comprehensible to stakeholders. Unlike opaque, "black-box" algorithmic systems, ADLA allows teachers, students, and families to understand precisely what is being measured and for what purpose (Slade & Prinsloo, 2024).

Furthermore, ADLA fosters **participation** by engaging local educators and communities in the instrument development process, thereby ensuring cultural responsiveness and contextual validity (Kim & Park, 2023). Finally, the framework upholds **fairness** through rigorous psychometric protocols. This involves systematic bias detection via Differential Item Functioning (DIF), culturally responsive adaptation procedures, and extensive validation across diverse populations to guarantee equitable assessment quality (Garcia et al., 2025; van de Vijver & Tanzer, 2024). Through these four pillars, ADLA transitions from passive data collection to an active architecture of educational equity.

OPERATIONALIZING ADLA: THREE-DIMENSIONAL ASSESSMENT ARCHITECTURE

Conceptual Architecture and System Integration

ADLA operationalizes three-dimensional learner modeling through systematic integration of cognitive diagnostics, affective inventories, and behavioral observations. Figure 1 illustrates the conceptual architecture:

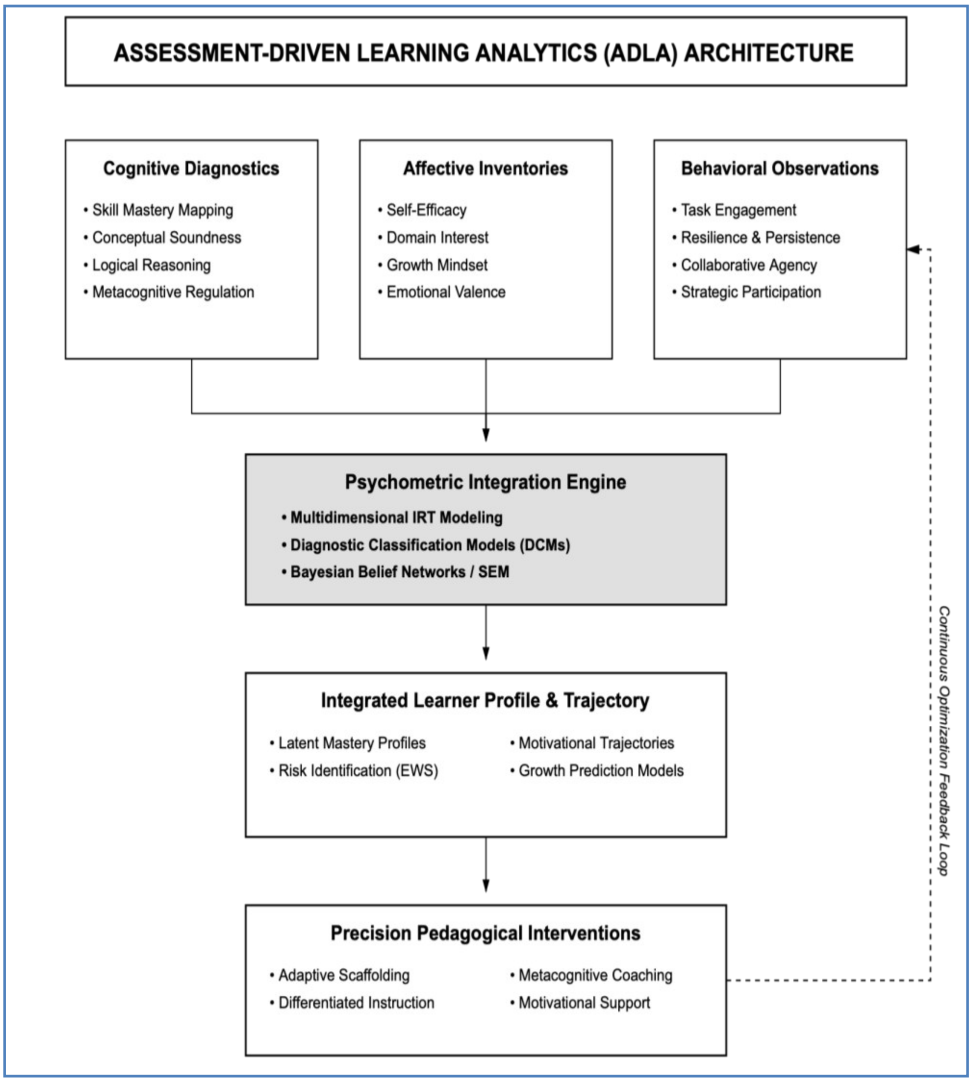

Figure 1. ADLA Three-Dimensional Assessment Architecture

Unlike clickstream LA capturing surface behaviors, ADLA systematically measures theoretically grounded constructs whose relationships are modeled through established psychometric frameworks. The system generates **actionable pedagogical insights** through integration rather than volume, enabling teachers to identify specific learning needs, track multidimensional trajectories, and implement evidence-based instructional responses.

Cognitive Component: Diagnostic Assessment of STEM Competencies

Diagnostic Classification Models as Psychometric Foundation

DCMs provide psychometric infrastructure for cognitive assessment, probabilistically inferring which specific skills each student has mastered based on item response patterns (de la Torre & Douglas, 2004; Rupp et al., 2010). Unlike classical test theory treating total scores, DCMs generate skill mastery profiles showing specific strengths and gaps enabling targeted instruction.

Q-Matrix specification maps items to required attributes. For example, a middle school mathematics fractions assessment might specify six items with varying attribute requirements. Item 1 requires both fraction concepts and procedural fluency but not translation or reasoning skills. Item 2 combines fraction concepts with word problem translation abilities. Item 3 necessitates procedural fluency and strategic reasoning without requiring conceptual knowledge. Item 4 represents a complex task requiring all four attributes simultaneously. Items 5 and 6 target different two-attribute combinations, enabling comprehensive diagnostic coverage across the skill space.

Given student responses and the Q-matrix specification, DCMs estimate the probability that each student has mastered each attribute. For instance, Student A might demonstrate 92% probability of mastering fraction concepts, 88% probability for procedural fluency, but only 35% probability for word problem translation and 41% probability for strategic reasoning. This profile interpretation reveals strong conceptual understanding and computational skill but struggles with application contexts and strategic thinking, indicating that targeted intervention should focus on translation and reasoning skills rather than remedial concept instruction. Recent psychometric advances (Lee et al., 2025; Zhang et al., 2024) demonstrate that DCMs can be estimated from paper-based assessments using freely available software packages including CDM and GDINA for R, making sophisticated diagnostic modeling accessible without requiring advanced technical infrastructure or proprietary systems.

Alignment with International Frameworks

ADLA cognitive assessment aligns with established international frameworks ensuring theoretical rigor and enabling cross-context comparisons. The TIMSS framework (Mullis & Martin, 2017) organizes assessment through comprehensive matrices crossing content domains with cognitive processes. Content domains include Number, Algebra, Geometry, and Data for mathematics, while science encompasses Biology, Chemistry, Physics, and Earth Science. These content areas intersect with three cognitive process levels: Knowing, encompassing factual recall and conceptual understanding; Applying, involving procedural execution and problem-solving; and Reasoning, requiring analysis, synthesis, and evaluation. ADLA cognitive instruments adopt this validated two-dimensional structure, ensuring comprehensive and balanced measurement demonstrated effective across more than 60 countries with diverse educational systems and technological contexts.

The PISA framework (OECD, 2023) emphasizes authentic problem-solving situated in real-world contexts, creative thinking involving generation of diverse ideas and exploration of possibilities, and most recently, AI literacy encompassing understanding of algorithms, data processing, and societal implications of artificial intelligence. ADLA incorporates performance tasks requiring integrated application of knowledge and skills in realistic scenarios, moving beyond decontextualized item formats to capture competencies as they manifest in actual STEM practice.

Computational thinking frameworks developed by Shute et al. (2017) and Wing (2006) operationalize AI-relevant cognitive skills essential for contemporary STEM+AI education. Decomposition involves breaking complex problems into manageable components amenable to systematic solution. Pattern recognition requires identifying regularities in data, situations, or problem structures. Abstraction focuses attention on essential features while deliberately ignoring irrelevant details. Algorithmic design creates step-by-step solution procedures that can be systematically executed. ADLA includes paper-based computational thinking assessment through tasks such as flowchart construction requiring students to design solution algorithms, pattern completion exercises identifying and extending regularities, and algorithm debugging scenarios where students identify and correct errors in provided solution procedures. These paper-based formats enable AI literacy measurement without requiring digital technology infrastructure, ensuring equitable access to assessment of these increasingly essential competencies.

Instrument Design Specifications

A hybrid item format approach balances diagnostic value with practical feasibility constraints. Selected-response items, constituting approximately 60% of assessments, include multiple-choice, true/false, and matching formats that enable efficient content coverage, objective scoring, and straightforward psychometric analysis. These items are particularly appropriate for measuring foundational knowledge, conceptual understanding, and basic application skills. Constructed-response items comprise the remaining 40% of assessments, requiring students to generate rather than select responses through short answers, explanations, diagrams, or extended problem-solving tasks. These items capture authentic reasoning processes, communication skills, and complex problem-solving abilities that remain invisible to selected-response formats, making them critical for measuring higher-order cognitive processes and identifying specific misconceptions in student thinking.

Analytic rubric development for constructed-response items systematically decomposes complex responses into scorable components aligned with specific cognitive attributes. Table 2 illustrates an exemplary rubric for scientific explanation tasks, demonstrating how multidimensional assessment can be operationalized.

Table 2. Analytic Rubric Example: Scientific Explanation Task

Component	Absent (0)	Basic (1)	Proficient (2)	Advanced (3)
Claim Formulation	No clear claim or off-topic	Vague statement without context	Clear, contextualized claim	Precise claim with appropriate qualifications
Evidence Citation	No evidence provided	Anecdotal or irrelevant evidence	Relevant observational data cited	Multiple credible sources integrated
Reasoning Quality	No connection evidence→claim	Superficial or circular link	Logical connection established	Sophisticated causal mechanism articulated
Scientific Accuracy	Major conceptual errors	Minor misconceptions present	Generally accurate understanding	Fully accurate with precise terminology

Each rubric component corresponds to a distinct cognitive attribute. Claim formulation reflects conceptual understanding of the phenomenon being explained. Evidence citation demonstrates data literacy and the ability to identify relevant empirical support. Reasoning quality reveals causal thinking and the capacity to construct logical arguments linking evidence to claims. Scientific accuracy indicates domain knowledge and precision in applying scientific concepts and terminology. This multidimensional structure enables detailed skill profiling from single complex tasks while providing actionable diagnostic feedback on specific reasoning components rather than holistic judgments that obscure areas requiring instructional attention.

The administration protocol for cognitive assessments follows a three-point measurement design enabling trajectory modeling and growth documentation. Pre-assessment administered at unit beginning establishes baseline skill status and identifies prerequisite knowledge gaps requiring attention before new instruction commences. Formative checkpoint assessments at unit midpoint monitor learning progress and enable mid-course instructional adjustments based on emerging patterns in student understanding. Summative assessment at unit conclusion evaluates mastery achievement and informs subsequent instructional planning for review, extension, or advancement to new content. This strategic timing aligns measurement occasions with pedagogically meaningful junctures when instructional decisions naturally occur, avoiding assessment saturation that would undermine valuable instructional time while providing sufficient data points for meaningful trajectory analysis.

Affective Component: Measuring Motivation, Beliefs, and Emotions

Theoretical Justification and Construct Selection

Affective variables exert powerful influence on STEM learning through multiple interconnected pathways (Bandura, 1997; Duckworth & Yeager, 2023; Pekrun et al., 2023). Unlike cognitive skills that represent capability, affective constructs explain motivation to engage that capability, making them essential components of comprehensive learner understanding and critical for early intervention before achievement decline becomes apparent.

Self-efficacy, defined as domain-specific belief in one's capability to succeed in particular tasks, represents a foundational motivational construct (Bandura, 1997). Students demonstrating high STEM self-efficacy attempt more challenging problems, persist longer when experiencing confusion, recover more effectively from failure experiences, and seek help appropriately rather than avoiding assistance due to concerns about appearing incompetent. These behaviors directly promote learning by increasing time on task, encouraging productive struggle, and facilitating timely intervention when misconceptions emerge.

Interest, characterized as intrinsic enjoyment of domain engagement, predicts sustained attention, deeper cognitive processing, and long-term persistence in STEM fields (Renninger & Hidi, 2019). Critically, Wang and Degol (2024) document through 15-year longitudinal studies that elementary STEM interest predicts adult career aspirations even after statistically controlling for achievement levels, suggesting that early interest assessment provides essential information for intervention before trajectories toward or away from STEM solidify during adolescence.

Growth mindset, the belief that intelligence and ability develop through effort rather than being fixed traits, shapes how students interpret struggle and respond to difficulty (Dweck, 2023; Yeager et al., 2023). Students endorsing growth mindsets interpret challenging tasks as opportunities for learning and development rather than as evidence of inadequacy, leading to greater persistence, more adaptive help-seeking behaviors, and ultimately higher achievement. Recent large-scale intervention research demonstrates that mindset beliefs are malleable through targeted interventions and consequential for learning outcomes, particularly for students from underrepresented groups in STEM fields.

Achievement goals, representing the purposes that motivate engagement in learning activities, shape the specific strategies students employ and their patterns of help-seeking behavior (Elliot & Hulleman, 2024). Mastery goals oriented toward understanding and improvement promote deep processing strategies, appropriate help-seeking when genuinely confused, and resilience in the face of difficulty. In contrast, performance-avoidance goals focused on not appearing incompetent predict surface-level learning strategies, help-avoidance even when struggling, and disengagement when challenges arise that might reveal limitations.

Academic emotions including enjoyment, anxiety, boredom, frustration, and confusion during STEM learning activities modulate cognitive engagement and learning efficiency (Pekrun et al., 2023). Moderate levels of confusion when followed by resolution through productive struggle enhance learning by stimulating cognitive conflict and subsequent knowledge restructuring. However, prolonged frustration or pervasive boredom undermines engagement and achievement by depleting cognitive resources and motivating withdrawal from learning activities. Understanding individual students' typical emotion patterns enables teachers to design responsive instructional environments and provide timely support during emotionally challenging learning moments.

Instrument Design and Cultural Adaptation

Effective affective assessment in ADLA leverages validated scales adapted for specific contexts rather than developing new measures, thereby benefiting from extensive prior psychometric work and enabling cross-context comparisons. The PISA student questionnaires provide well-established instruments that have been validated across more than 80 countries for measuring STEM self-efficacy, interest, and epistemic beliefs (OECD, 2019, 2023). These publicly available items demonstrate robust psychometric properties including acceptable internal consistency coefficients exceeding .80, adequate test-retest reliability, and demonstrated predictive validity for STEM achievement and persistence outcomes across diverse international samples. The CASEL social-emotional learning assessments offer complementary validated instruments measuring self-awareness, self-management, and social awareness, representing broader social-emotional competencies that support academic learning engagement (CASEL, 2020). Domain-specific adaptations of these generic scales contextualize individual items to particular STEM domains and current instructional topics, maximizing both relevance for students responding to surveys and actionability for teachers making real-time instructional decisions based on assessment results.

Developmental appropriateness constitutes a critical consideration in instrument design, as cognitive and linguistic capacities vary substantially across elementary, middle, and high school grade levels. Elementary students require substantially simpler vocabulary, visual response formats such as pictorial scales depicting emotion faces, and concrete items referencing familiar experiences from their daily lives rather than abstract hypothetical scenarios. The contrast between secondary and elementary versions of self-efficacy items illustrates this adaptation principle clearly. A secondary-level item appropriate for high school students might state in relatively complex language that "I am confident in my ability to learn advanced mathematical concepts even when they are initially confusing," employing abstract terminology about confidence, advanced concepts, and initial confusion that assumes sophisticated metacognitive awareness. The developmentally appropriate elementary version targeting younger students would simplify this substantially to the straightforward statement "I can do well in math," accompanied by a four-point pictorial response scale progressing from a very sad face through a sad face and a happy face to culminate in a very happy face. These careful adaptations maintain construct fidelity and measurement validity while ensuring adequate comprehensibility for younger students whose less developed abstract reasoning abilities might otherwise produce invalid responses reflecting misunderstanding rather than true self-efficacy levels.

Rigorous cultural adaptation procedures ensure measurement validity across diverse linguistic and cultural populations, following systematic processes outlined by van de Vijver and Tanzer (2024) in their comprehensive framework for cross-cultural test adaptation and validation. The adaptation process begins with forward translation conducted by bilingual educators who possess familiarity with both educational terminology and local dialectical variations, ensuring that translated items maintain semantic equivalence while employing naturally occurring language patterns in the target culture. Back-translation performed by independent translators who have not seen the original items serves to identify instances of semantic drift where meaning has shifted during the translation process and cultural misalignment where direct translation produces awkward or inappropriate phrasing in the target language. Cultural review panels comprising community representatives, local educators, and cultural experts evaluate the appropriateness of specific examples, contextual references, and conceptual frameworks embedded in items for the target population, identifying items that may carry substantially different meanings, connotations, or cultural resonance across different cultural contexts.

Cognitive interviews conducted with representative samples of target population students, typically involving 15 to 20 participants per cultural group, verify whether students interpret items as intended by researchers and identify specific items that students find confusing, ambiguous, or susceptible to multiple conflicting interpretations. Pilot testing with substantially larger samples ranging from 100 to 200 students enables examination of psychometric properties including internal consistency reliability, factor structure consistency with theoretical models, and individual item functioning characteristics. Differential Item Functioning analysis employing modern item response theory methods systematically identifies specific items that function differently across demographic groups despite students possessing equivalent levels of the underlying construct being measured, thereby flagging potential sources of systematic measurement bias requiring either revision or removal from final instruments. Finally, iterative refinement based on empirical evidence accumulated from pilot testing and DIF analysis produces culturally adapted instruments that demonstrate measurement equivalence across diverse populations, ensuring that observed score differences reflect genuine differences in underlying constructs rather than artifacts of cultural bias, while simultaneously maintaining appropriate sensitivity to local contexts, experiences, and meaning systems that shape how students understand and respond to affective assessment items.

An exemplary affective scale measuring STEM self-efficacy demonstrates appropriate item construction. The six-item scale employs five-point Likert response format ranging from Strongly Disagree to Strongly Agree. Items include "I can master the skills taught in my STEM class," "I can do well on STEM tests even when material is challenging," "I can understand difficult STEM concepts if I try hard," "I can succeed at solving complex STEM problems," "I am confident I can learn advanced STEM topics," and "Even when STEM work is hard, I can figure it out." Psychometric validation demonstrates acceptable internal consistency ($\alpha = .87$), adequate four-week test-retest reliability ($r = .76$), and single-factor structure confirmed through confirmatory factor analysis (CFI = .96, RMSEA = .048). Criterion validity evidence includes moderate positive correlations with STEM achievement ($r = .52$), engagement ($r = .64$), and course-taking intentions ($r = .41$), supporting the scale's utility for identifying students requiring confidence-building interventions.

Administration frequency for affective assessments balances comprehensive trajectory tracking with practical feasibility constraints inherent in classroom settings. Quarterly measurement provides sufficient data points for meaningful pattern detection while avoiding excessive assessment burden. Fall baseline administration in September establishes initial motivational profiles at academic year commencement. Winter checkpoint measurement in January tracks mid-year changes and identifies students showing concerning declining trajectories requiring intervention. Spring outcome assessment in May evaluates year-end motivational status and informs summer programming or next-year planning decisions. Additional administration following significant instructional events such as major projects, particularly challenging units, notable failures, or exceptional successes captures reactions to these learning experiences that may substantially influence subsequent motivation and engagement patterns.

Behavioral Component: Structured Observation of Engagement

Engagement Frameworks and Operational Definitions

Observable behaviors during learning activities provide external evidence of internal cognitive and affective processes (Fredricks et al., 2024; Reeve & Tseng, 2023). Three distinct forms of engagement demonstrate particular relevance for STEM learning contexts.

Behavioral engagement encompasses active participation through contributing to classroom discussions and completing assigned work, consistent attendance, adherence to classroom norms and procedures, and demonstrated effort through persistence on challenging activities. While behavioral engagement alone proves insufficient for ensuring deep learning, it creates necessary preconditions for cognitive engagement to occur by establishing physical and social presence in learning environments and commitment to completing learning tasks.

Cognitive engagement manifests through observable indicators of mental effort including asking substantive questions that seek conceptual understanding rather than merely procedural guidance about what to do next, offering explanations that include explicit reasoning connecting evidence to conclusions rather than simply stating answers, making explicit connections between new material and prior knowledge through verbal or written articulation, and employing strategic problem-solving approaches involving planning before executing solutions, monitoring progress during work, and evaluating solution quality after completion. Though cognitive processes themselves remain internal and unobservable, these external behavioral manifestations can be reliably identified and systematically coded through structured observation protocols.

Agentic engagement involves students proactively constructing their learning experiences through suggesting investigation topics, questions, or approaches that extend beyond assigned requirements, requesting clarification when confused or additional challenges when ready to advance, taking initiative in group work by proposing solution approaches, delegating tasks to team members, and ensuring collective progress toward goals, and setting personal learning goals that exceed minimum requirements specified by teachers (Reeve & Tseng, 2023). This form of engagement demonstrates particular importance for developing self-regulated learning capacities essential for success in complex STEM domains requiring substantial student initiative and autonomy.

Operational definitions translate these theoretical engagement constructs into specific, observable actions amenable to systematic coding. Behavioral engagement indicators include arriving to class prepared with necessary materials and completed prerequisite work, remaining on-task during independent work periods for more than 80% of observed intervals, completing assigned activities within allocated timeframes, and persisting on challenging problems by continuing effort for more than two minutes before seeking teacher assistance. Cognitive engagement indicators comprise asking conceptual questions employing "why" or "how" formulations seeking understanding rather than procedural questions asking "what do I do next," offering explanations that include explicit reasoning using "because" or similar connective language rather than stating conclusions without justification, making explicit verbal or written connections to previous learning using phrases such as "this is like when we studied," and using strategic problem-solving approaches observable through planning activities before beginning work, monitoring progress through checking intermediate steps, and evaluating solutions through verification procedures. Agentic engagement indicators involve suggesting topics, questions, or investigation approaches beyond assigned requirements, requesting additional resources, examples, or challenges when demonstrating readiness to advance, taking initiative in group work through proposing approaches, delegating responsibilities, or ensuring team progress, and setting personal learning goals documented through written statements or verbal declarations that exceed minimum teacher-specified requirements.

Observation Methods and Implementation

Systematic interval sampling employs structured teacher observation of focal students for brief two-to-three-minute intervals at regular points throughout class sessions, with observers coding behaviors manifested during each observation window. Rotating through different focal students across days provides comprehensive class coverage without creating overwhelming attention demands that would interfere with instructional responsibilities (Pianta et al., 2008). For example, during a 45-minute STEM lesson, a teacher might observe six different students for three-minute intervals each, distributing 18 minutes of total observation time across the class period while maintaining primary attention on instruction and student support.

Event recording involves teachers tallying each occurrence of specific target behaviors using simple recording systems such as clipboard checklists, seating charts with tally marks, or mobile applications featuring student name lists. This method demonstrates particular effectiveness for discrete, easily identifiable behaviors including questions asked with coding distinguishing conceptual versus procedural inquiries, peer explanations offered to classmates, instances of appropriate help-seeking when genuinely confused, and off-task incidents disrupting learning (Hamre et al., 2023). The cumulative frequency counts generated through event recording provide quantitative indicators of engagement patterns over time.

Holistic rating scales require teachers to provide overall judgments of each student's engagement level based on accumulated observations over class periods or weeks, employing rubrics with clear behavioral anchors at each performance level. While less granular than event recording methods, rating scales prove more feasible for daily implementation as they require minimal in-the-moment attention during instruction (Lund et al., 2024). An exemplary rubric for assessing STEM persistence using weekly ratings on a four-point scale illustrates this approach. Level 4 (Exceptional) describes students who consistently persist through multiple challenging problems across varied contexts, actively seek alternative strategies when initial approaches fail, view difficulty as opportunity for learning, rarely require teacher prompting to maintain productive effort, and recover quickly from setbacks. Level 3 (Proficient) characterizes students who usually persist on challenging problems, try alternative approaches when initial attempts prove unsuccessful, respond positively to occasional brief teacher prompting sufficient to maintain engagement, generally view difficulty positively or neutrally, and persist through at least one setback per session. Level 2 (Developing) indicates students demonstrating inconsistent persistence, sometimes abandoning challenging problems quickly without sustained effort, requiring regular teacher encouragement to continue working, alternating between productive engagement and avoidance behaviors, and showing increasingly frequent negative reactions to difficulty. Level 1 (Beginning) identifies students who rarely persist independently on challenging problems, quickly become frustrated and disengaged when encountering difficulty, require frequent teacher intervention to maintain any engagement, predominantly view difficulty negatively, and actively avoid challenge when possible.

Quality assurance procedures maintain data reliability across observers and over time. Initial training requires teachers to complete practice coding sessions using video examples previously scored by expert raters, receiving immediate corrective feedback until achieving acceptable inter-rater reliability with target thresholds of ICC greater than .70 for continuous ratings and κ greater than .70 for categorical codes. Ongoing calibration through monthly meetings involves jointly coding identical classroom observations or video segments, calculating agreement statistics, discussing discrepancies in interpretation, and recalibrating shared understanding of behavioral indicators to prevent drift from established standards.

Bias awareness training provides explicit instruction about common observation biases including halo effects where global impressions of students influence specific judgments, recency bias emphasizing recently observed behaviors over typical patterns, and stereotype-based expectations shaped by demographic characteristics rather than actual observed behaviors. Mitigation strategies include rotating focal students to prevent oversampling of familiar individuals, employing structured protocols that reduce opportunities for subjective judgment, and maintaining conscious awareness of potential bias sources during observation. Periodic reliability checks conducted every four to six weeks involve two observers independently coding the same classroom sessions; when calculated reliability falls below established thresholds (ICC less than .70), intensive recalibration occurs before continuing data collection to ensure maintained data quality standards.

Psychometric Integration: Building Holistic Learner Profiles

Multidimensional IRT, structural equation modeling (SEM), and Bayesian networks enable simultaneous modeling of cognitive, affective, and behavioral dimensions while estimating relationships and accounting for measurement error (Liu et al., 2024; von Davier, 2023). Cross-dimensional pattern recognition identifies actionable learner profiles invisible in single-dimension analysis:

Table 3. Learner Profile Examples and Pedagogical Responses

Cognitive Level	Affective Pattern	Behavioral Pattern	Interpretation	Targeted Response
High skills across attributes	Low self-efficacy, declining interest	Minimal participation, inconsistent effort	Capable but unmotivated; possibly lacks relevance or challenge	Increase task complexity, emphasize real-world applications, provide autonomy, build confidence through success experiences
Low skills, specific gaps identified	High interest, growth mindset	Strong behavioral engagement, high persistence	Motivated but struggling; lacks prerequisite knowledge or effective strategies	Targeted skill intervention on specific attributes, explicit strategy instruction, maintain motivation through appropriate challenge
Declining across multiple attributes	Declining self-efficacy and interest	Decreasing participation and persistence	Comprehensive disengagement trajectory; early warning signal	Immediate multifaceted intervention: academic support, motivational counseling, environmental assessment, family communication
Uneven profile (high in some, low in others)	Topic-specific interest variation	Context-dependent engagement	Interest-driven selective engagement	Leverage strengths to build confidence, connect weaker areas to interests, differentiate by readiness and interest

Visualization tools support teacher interpretation even in resource-constrained contexts. Simple spreadsheet-based dashboards employing conditional formatting provide substantial functionality without requiring sophisticated technology infrastructure (Garcia & Smith, 2024). Skill mastery heat maps display color-coded or pattern-coded grids showing mastery status for each cognitive attribute across students, with red indicating not yet mastered, yellow approaching mastery, and green representing achieved mastery. Affective trajectory line graphs plot individual student and class-level trends in self-efficacy, interest, and mindset across measurement occasions, enabling identification of improving, stable, or declining patterns. Engagement timeline charts present weekly behavioral ratings revealing temporal patterns and concerning changes requiring attention. Automated alert flags highlight students demonstrating declining trajectories across multiple dimensions or significant discrepancies between capability and engagement levels, prompting timely intervention before difficulties become entrenched.

Theoretical Propositions

The conceptual architecture of ADLA generates several theoretically grounded propositions that can guide future empirical validation. The first proposition concerns predictive validity. Contemporary learning analytics research demonstrates that psychometric factors such as ability, motivation, personality, and learning strategies contribute significantly to the prediction of academic performance and can strengthen learner models beyond achievement indicators alone (Gray, 2014). At the same time, behavioral trace data derived from digital interactions have been successfully used to model learner characteristics and predict educational outcomes through machine learning approaches (Landers et al., 2022). Evidence from web-based educational systems further shows that student activity patterns can achieve high levels of predictive accuracy in identifying learners at risk of poor performance (Casey & Azcona, 2017). Taken together, these findings suggest that assessment-driven learner profiles may possess predictive validity comparable to digital trace-based models, particularly in educational contexts where continuous digital interaction data are unavailable.

A second proposition concerns the value of multidimensional learner modeling. ADLA is built on the premise that meaningful educational decisions require evidence not only about cognitive achievement but also about affective and behavioral dimensions of learning. The integration of multiple forms of evidence aligns with evidence-centered design (ECD), which emphasizes the systematic collection and interpretation of evidence linked to theoretically meaningful constructs (Mislevy et al., 2012). From this perspective, learner profiles that combine cognitive diagnostics, affective indicators, and behavioral observations should provide a more comprehensive representation of learning processes than achievement-only assessment systems. Consequently, ADLA proposes that multidimensional learner profiles may improve the early identification of students who are academically or motivationally at risk before such risks become visible through achievement measures alone.

A third proposition addresses educational equity. Existing learning analytics systems are often dependent on digital infrastructures capable of generating continuous streams of learner interaction data. While such approaches have demonstrated substantial analytical power, they also risk excluding learners in technology-constrained environments where digital traces are sparse or unavailable. By contrast, psychometric assessment data can be collected independently of technological infrastructure while still providing theoretically meaningful evidence about learner characteristics (Gray, 2014). Therefore, ADLA proposes that technology-independent analytics systems may reduce disparities in educational data availability between resource-rich and resource-constrained schools, thereby extending the benefits of data-informed decision making to broaden student populations.

Finally, ADLA advances a proposition regarding interpretability and transparency. Traditional psychometric assessments are intentionally designed to measure specific constructs and therefore provide evidence that can be directly understood and communicated by teachers, students, and families (Gray, 2014). In contrast, models derived from complex behavioral traces often require sophisticated computational procedures to infer underlying learner characteristics, raising questions about interpretability and psychometric validity (Landers et al., 2022). Although digital assessment environments provide rich behavioral evidence and new opportunities for measurement innovation (Gibson, 2013), ensuring the validity and transparency of such models remains a continuing challenge. Consequently, ADLA proposes that assessment-driven analytics may offer a more interpretable and transparent learner modeling framework while preserving the diagnostic value necessary for educational decision making.

Collectively, these propositions position ADLA as a theoretically grounded alternative pathway within learning analytics. Rather than rejecting digital-trace approaches, ADLA expands the evidentiary foundations of learner modeling by demonstrating how psychometric assessment data can function as a primary analytics infrastructure. Future empirical studies are required to test these propositions and determine the conditions under which assessment-driven analytics can complement or, in some contexts, substitute for trace-based learning analytics systems.

IMPLEMENTATION FRAMEWORK

Phased Implementation Model

Successful ADLA implementation requires systematic, staged introduction rather than attempting comprehensive system deployment immediately. Phase 1 (Foundation, Months 1-3) focuses on establishing essential infrastructure through leadership engagement to secure commitment and resource allocation, teacher assessment literacy professional development addressing fundamental concepts of validity, reliability, and bias recognition, single dimension implementation typically beginning with cognitive diagnostics as the most familiar assessment type, and basic data management infrastructure including spreadsheet templates and recording protocols. Phase 2 (Expansion, Months 4-8) builds on this foundation by adding the affective assessment dimension, initiating cross-dimensional data integration and interpretation practices, establishing collaborative data team routines with bi-weekly meetings for analyzing student data and planning instructional responses, and refining instruments and procedures based on Phase 1 implementation experience and teacher feedback. Phase 3 (Full Implementation, Months 9-18) completes the system by adding the behavioral observation dimension, achieving full three-dimensional learner modeling integration across all assessment components, establishing sustainability mechanisms including new teacher onboarding protocols, ongoing calibration procedures, and quality monitoring systems, and systematically documenting implementation processes and emerging student outcome evidence to support continuous improvement and dissemination to other schools.

Professional Development Architecture

Effective ADLA implementation requires teacher development across four core competency domains. Assessment literacy encompasses understanding fundamental concepts of validity (whether assessments measure intended constructs), reliability (consistency of measurement), and bias (systematic errors favoring particular groups), along with recognizing appropriate and inappropriate assessment uses. Data interpretation skills enable teachers to read statistical summaries without requiring statistical expertise, recognize meaningful patterns in student data while distinguishing signal from noise, integrate information from multiple data sources into coherent understandings, and connect assessment evidence to specific instructional hypotheses and responses. Technical skills involve standardized assessment administration following protocols precisely, accurate scoring using answer keys and rubrics, efficient data recording and organization, and basic technology use including spreadsheets and simple databases where available. Pedagogical application competencies link assessment evidence to instructional decisions through adaptive grouping, differentiated instruction, early intervention, and strategy support.

Professional development design employs job-embedded practice rather than isolated workshops, incorporating collaborative scoring sessions where teachers apply rubrics together while discussing difficult cases, data team meetings analyzing actual classroom assessment results and planning instructional responses, peer observation focused on assessment administration and data use during instruction, and action research projects investigating effects of ADLA-informed practices in teachers' own classrooms. Scaffolded progression builds capacity incrementally by beginning with concrete skills including administering assessments and using scoring rubrics, progressing to interpretation of what data reveal about student learning, advancing application through designing instructional responses based on evidence, and culminating in innovation involving creating new assessments, conducting sophisticated analyses, and providing peer leadership. Practice with expert feedback ensures skill development through scoring sample responses with expert-scored exemplars available for comparison, interpreting practice datasets before analyzing own students' actual data, role-playing assessment administration with peer feedback on technique, and receiving coaching during initial classroom implementation. Collaborative learning communities sustain motivation and knowledge sharing through regular meetings with structured protocols for data review and instructional planning, shared resources including item banks, rubrics, analysis tools, and lesson plans, peer mentoring pairing experienced ADLA users with novices, and online communities enabling resource sharing and asynchronous collaboration.

Quality Assurance and Ethical Governance

Validity assurance requires ongoing attention to multiple validity dimensions. Content validity is maintained through expert review panels and teacher feedback mechanisms ensuring assessments comprehensively and representatively cover intended constructs. Construct validity is examined via factor analysis confirming hypothesized dimensional structure and convergent/discriminant validity studies establishing appropriate relationships with related measures while demonstrating distinctness from unrelated constructs. Consequential validity monitoring tracks both intended effects including improved instruction and student outcomes, and unintended effects such as narrowed curriculum focus, increased student anxiety, or teaching to the test.

Reliability assurance establishes and maintains measurement consistency. Internal consistency coefficients ($\alpha > .70$) document that items within scales measure common constructs. Inter-rater reliability (ICC $> .70$) ensures consistency across different scorers through systematic training and calibration procedures. Test-retest stability examines whether constructs that should remain relatively stable across brief time periods demonstrate appropriate consistency.

Bias detection and mitigation employs multiple strategies. Differential Item Functioning analysis identifies items that function differently across demographic groups despite equivalent underlying construct levels, flagging potential sources of systematic measurement bias requiring revision or removal. Culturally responsive adaptation procedures ensure instruments function appropriately across diverse populations through community participation in development, translation and back-translation protocols, and cognitive interviewing with target populations. Blind scoring procedures, when feasible, prevent scorer knowledge of student identity from influencing judgments.

The ethical framework governing ADLA implementation rests on four foundational principles. Beneficence requires that data collection and use serve students' educational interests through informing instruction that improves learning outcomes, enabling timely support for struggling students, and promoting equity rather than reproducing disadvantage. Non-maleficence demands avoiding harm through ensuring assessment burden remains balanced against educational benefit, monitoring and mitigating emotional impacts including anxiety and stereotype threat, protecting privacy through preventing unauthorized access or misuse, and avoiding punitive uses of assessment data for grades, tracking decisions, or punishment. Justice necessitates fair treatment across all student groups through culturally responsive instruments avoiding systematic bias, equal access to ADLA benefits regardless of school resources or student demographics, disaggregated analysis enabling equity monitoring, and addressing inequitable patterns when identified. Transparency requires comprehensible processes where stakeholders understand what is measured and why, appropriate consent procedures, data interpretation shared with students to foster self-understanding, and opt-out options available while maintaining system integrity where educationally appropriate.

RESEARCH AGENDA FOR EMPIRICAL VALIDATION

Although ADLA awaits large-scale empirical validation, its conceptual validity derives from integrating three mature knowledge traditions: learning analytics, educational assessment, and computational psychometrics. The novelty of the framework therefore lies not in introducing entirely new measurement constructs, but in reorganizing established constructs into a new analytics architecture.

The systematic validation of the Assessment-Driven Learning Analytics (ADLA) framework necessitates a multi-faceted research agenda. First, **Psychometric Validation Studies** must be prioritized to establish the reliability and validity of ADLA instruments across diverse educational contexts. Critical inquiries should focus on internal consistency, inter-rater reliability, and test-retest stability. Furthermore, researchers must employ Confirmatory Factor Analysis (CFA) to verify intended constructs and utilize Differential Item Functioning (DIF) analysis to ensure measurement invariance across demographic groups. Large-scale validation studies ($n > 1,000$) are recommended to establish predictive validity concerning distal outcomes such as longitudinal achievement and persistence.

Beyond psychometric properties, **Effectiveness Research** is required to determine the causal impact of ADLA on student engagement and STEM achievement. Recommended designs include cluster-randomized controlled trials and quasi-experimental methods, such as propensity score matching, to compare ADLA with business-as-usual models. Such research should also examine how ADLA influences teacher decision-making and whether it effectively mitigates achievement gaps. Additionally, **Comparative Research** should evaluate ADLA against traditional periodic testing and clickstream-based analytics to determine its relative diagnostic specificity and cost-effectiveness. Finally, **Implementation Science** must investigate the systemic factors—infrastructure, professional development models, and policy frameworks—that facilitate the scalable and sustainable adoption of ADLA.

POLICY IMPLICATIONS AND SDG 4 ALIGNMENT

The transition toward ADLA requires significant policy shifts focused on equity and assessment literacy. Policy Recommendations include mandating a technology-agnostic assessment infrastructure that prioritizes systematic learner diagnostics regardless of a school's digital maturity. Accountability systems should recognize psychometric evidence as a legitimate form of data-informed practice. Furthermore, it is essential to integrate assessment literacy into teacher preparation programs and licensure requirements. Financial support must be dedicated to developing validated instrument repositories and providing psychometric expertise to under-resourced districts.

These initiatives directly support Sustainable Development Goal 4 (SDG 4). Specifically, ADLA aligns with Target 4.1 by enabling evidence-based instruction in low-technology environments, ensuring universal quality education. By measuring both STEM competencies and non-cognitive dispositions, ADLA addresses Target 4.4 regarding workforce readiness. Most importantly, Target 4.5 is supported through ADLA's technology-independent nature, which ensures that data-informed support remains accessible to marginalized populations. International development agencies, including UNESCO and the World Bank, should therefore include ADLA in their assistance portfolios, fostering South-South cooperation and regional capacity building in psychometrics.

POLICY IMPLICATIONS AND SDG 4 ALIGNMENT

The ADLA framework represents a paradigmatic contribution to the field of learning analytics by challenging the assumption that meaningful diagnostics require high-frequency digital trace data. Theoretically, ADLA operationalizes sociocognitive constructs more directly than clickstream proxies, which often require significant inferential leaps. Functionally, it provides equivalent diagnostic value for early intervention and personalized support without the need for continuous digital infrastructure. Ethically, ADLA addresses the "analytics divide," ensuring that data justice is realized for learners in all technological contexts.

While acknowledging limitations such as the resource intensity of instrument development and lower data frequency compared to digital traces, these challenges can be mitigated through pooled resources and open-source repositories. More importantly, ADLA should be viewed as a complementary approach rather than a replacement for digital analytics. The framework encourages epistemological diversity within the field, urging researchers to move beyond atheoretical data mining toward theory-driven, interdisciplinary measurement design.

In conclusion, ADLA offers a pathway toward inclusive learning analytics. By reconceptualizing psychometric assessments as primary analytic infrastructure, this framework demonstrates that technological sophistication is not a prerequisite for data-informed excellence. Educational equity demands that the benefits of data science serve all students, not just the privileged few. Historically, learning analytics has evolved alongside technological expansion. ADLA suggests an alternative trajectory: analytics maturity need not be determined by technological sophistication but by the quality, validity, and interpretability of evidence used to support learning. This shift reframes learning analytics as an educational science rather than merely a technological enterprise. The choice for the field is clear: continue creating two-tiered systems or embrace a methodological pluralism that recognizes theoretical validity and inclusive accessibility as the true hallmarks of innovation.

LIMITATION & FURTHER RESEARCH

Despite its theoretical contributions, the ADLA framework faces several inherent limitations. First, the model's reliance on human capital—specifically teacher assessment literacy and psychometric expertise—poses a significant challenge for scalability in under-resourced regions. Second, ADLA's strategic, periodic measurement cannot achieve the granular, real-time data frequency provided by continuous digital trace logs. Furthermore, the systematic friction involved in developing and validating cross-cultural instrument repositories may hinder immediate global adoption.

Future research should prioritize large-scale empirical validation ($n > 1,000$) to ensure measurement invariance across diverse demographic groups through Differential Item Functioning (DIF) analysis. Additionally, cluster-randomized controlled trials (RCTs) are required to establish the causal impact of ADLA on longitudinal STEM achievement compared to traditional or purely digital analytics. Finally, exploring AI-assisted, semi-automated scoring for paper-based tasks could mitigate teacher workload, bridging the gap between diagnostic precision and practical feasibility.

REFERENCES

Baker, R. S., & Inventado, P. S. (2023). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (3rd ed., pp. 333–353). Cambridge University Press. https://doi.org/10.1017/9781108888295.019

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Prentice-Hall.

Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman.

CASEL. (2020). *The CASEL framework*. Collaborative for Academic, Social, and Emotional Learning. https://casel.org/fundamentals-of-sel/what-is-the-casel-framework/

Casey, K., & Azcona, D. (2017). Utilizing student activity patterns to predict performance. *International Journal of Educational Technology in Higher Education*, *14*(1), 1–15. https://doi.org/10.1186/S41239-017-0044-3

Chen, F., Azevedo, R., & Taub, M. (2023). Predicting STEM persistence through real-time modeling of cognitive-emotional dynamics. *Journal of Educational Data Mining, 15*(1), 45–67. https://doi.org/10.5281/zenodo.7892341

Chen, J., von Davier, A. A., & Mislevy, R. J. (2021). Computational psychometrics: New methodologies for a new generation of digital learning and assessment. Springer. https://doi.org/10.1007/978-3-030-74394-9

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika, 69*(3), 333–353. https://doi.org/10.1007/BF02295640

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press. https://doi.org/10.7551/mitpress/11805.001.0001

Duckworth, A. L., & Yeager, D. S. (2023). Non-cognitive skills in education: An integrative framework for measurement and intervention. *Annual Review of Psychology, 74*, 95–123. https://doi.org/10.1146/annurev-psych-032420-031851

Dweck, C. S. (2023). Mindset: The new psychology of success (Updated ed.). Ballantine Books.

Elliot, A. J., & Hulleman, C. S. (2024). Achievement goals and achievement goal complexes. In A. J. Elliot (Ed.), *Advances in motivation science* (Vol. 11, pp. 1–53). Elsevier. https://doi.org/10.1016/bs.adms.2023.10.001

Fredricks, J. A., Reschly, A. L., & Christenson, S. L. (Eds.). (2024). *Handbook of student engagement interventions* (2nd ed.). Academic Press. https://doi.org/10.1016/C2021-0-01847-3

Garcia, R., & Smith, T. (2024). Paper-based learner analytics: Designing analog solutions for data-informed instruction. *International Journal of Educational Technology in Higher Education, 21*(2), 112–130. https://doi.org/10.1186/s41239-024-00432-1

Garcia, R., Nguyen, T., & Lee, S. (2025). Advancements in psychometric evaluation: Ensuring validity and reliability in multicultural assessments. *Educational Assessment Review, 43*(1), 56–78. https://doi.org/10.1016/j.edurev.2024.100562

Gasevic, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Gray, G. (2014). A Review of Psychometric Data Analysis and Applications in Modelling of Academic Achievement in Tertiary Education. *Journal of Learning Analytics*, *1*(1), 75–106. https://doi.org/10.18608/JLA.2014.11.5

Hamre, B. K., Hamagami, A., Pianta, R. C., Baroody, A. E., & Goffin, S. G. (2023). Classroom Assessment Scoring System implementation and associations with student outcomes. *Educational Assessment, 28*(2), 98–118. https://doi.org/10.1080/10627197.2023.2198765

Holstein, K., McLaren, B. M., & Aleven, V. (2023). Reframing learning analytics to promote inclusion and equity in K–12 education. *British Journal of Educational Technology, 54*(1), 67–83. https://doi.org/10.1111/bjet.13257

Ifenthaler, D., Shibani, A., & Ferguson, R. (2022). Hybrid learning analytics: Integrating analogue and digital learning evidence. *Educational Technology Research and Development, 70*(4), 1057–1076. https://doi.org/10.1007/s11423-022-10108-6

Kim, J., & Park, E. (2023). Collaborative development of culturally responsive assessment tools in diverse classrooms. *Assessment in Education: Principles, Policy & Practice, 30*(4), 435–452. https://doi.org/10.1080/0969594X.2023.2234567

Landers, R. N., Auer, E. M., Mersy, G., Marin, S., & Blaik, J. A. (2022). You are what you click: using machine learning to model trace data for psychometric measurement. *International Journal of Testing*, *22*(3–4), 243–263. https://doi.org/10.1080/15305058.2022.2134394

Lee, H., Choi, S., & Kim, J. (2025). Diagnostic assessment of computational thinking in unplugged contexts: Development and validation. *Computers & Education, 197*, Article 104753. https://doi.org/10.1016/j.compedu.2024.104753

Liu, J., Broisin, J., & Sharma, K. (2023). Learning analytics in K–12: Redefining data strategies in hybrid classrooms. *Computers & Education: Artificial Intelligence, 4*, Article 100108. https://doi.org/10.1016/j.caeai.2023.100108

Liu, J., von Davier, A. A., & Mislevy, R. J. (2024). Psychometric models for multimodal educational data: Opportunities and challenges. *Journal of Educational Measurement, 61*(1), 1–24. https://doi.org/10.1111/jedm.12378

Lund, A., Rasmussen, I., & Smørdal, O. (2024). Classroom observation of engagement in collaborative STEM learning: A validity study. *International Journal of Science Education, 46*(3), 289–312. https://doi.org/10.1080/09500693.2023.2289456

Mislevy, R. J., Oranje, A., Bauer, M. I., von Davier, A. A., & Hao, J. (2021). *Psychometrics for education: Modern theories and applications*. Springer. https://doi.org/10.1007/978-3-030-74769-5

Mislevy, R. J., Behrens, J. T., DiCerbo, K. E., & Levy, R. (2012). Design and Discovery in Educational Assessment: Evidence-Centered Design, Psychometrics, and Educational Data Mining. *Educational Data Mining*, *4*, 11–48. https://doi.org/10.5281/ZENODO.3554641

Mullis, I. V. S., & Martin, M. O. (2017). *TIMSS 2019 assessment frameworks*. TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/timss2019/frameworks/

NCES. (2023). *Digest of education statistics 2023*. National Center for Education Statistics, U.S. Department of Education. https://nces.ed.gov/programs/digest/

OECD. (2019). PISA 2018 results (Volume III): What school life means for students' lives. OECD Publishing. https://doi.org/10.1787/acd78851-en

OECD. (2023). *PISA 2022 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en

Pekrun, R., Marsh, H. W., Elliot, A. J., Stockinger, K., Perry, R. P., Vogl, E., Goetz, T., van Tilburg, W. A. P., Lüdtke, O., & Vispoel, W. P. (2023). A three-dimensional taxonomy of achievement emotions. *Journal of Personality and Social Psychology, 124*(1), 145–178. https://doi.org/10.1037/pspp0000448

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System manual: K–3*. Paul H. Brookes Publishing.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*(4), 667–686. https://doi.org/10.1037/0022-0663.95.4.667

Prinsloo, P., & Slade, S. (2017). Ethics and learning analytics: Charting the (un)charted. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (pp. 49–57). Society for Learning Analytics Research. https://doi.org/10.18608/hla17.004

Reeve, J., & Tseng, C. (2023). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology, 36*(4), 257–267. https://doi.org/10.1016/j.cedpsych.2011.05.002

Renninger, K. A., & Hidi, S. E. (2019). *The Cambridge handbook of motivation and learning*. Cambridge University Press. https://doi.org/10.1017/9781316823279

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications.* Guilford Press.

Sailer, M., Stadler, M., Botev, J., & Fischer, F. (2021). Technology-supported learning analytics in higher education: A systematic review of academic success prediction. *Computers & Education, 173*, Article 104271. https://doi.org/10.1016/j.compedu.2021.104271

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review, 22*, 142–158. https://doi.org/10.1016/j.edurev.2017.09.003

Siemens, G., & Gasevic, D. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254. https://doi.org/10.1145/2330601.2330661

Slade, S., & Prinsloo, P. (2024). Ethics in K–12 learning analytics: Beyond privacy toward pedagogical transparency. *British Journal of Educational Technology, 55*(1), 1–15. https://doi.org/10.1111/bjet.13389

Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education, 12*(4), 618–627. https://doi.org/10.1187/cbe.13-08-0154

UNESCO. (2023). Global education monitoring report 2023: Inclusion and education—All means all. UNESCO Publishing. https://doi.org/10.54676/UZQV8501

van de Vijver, F. J., & Tanzer, N. K. (2024). Cross-cultural test adaptation and validation. In S. Guo (Ed.), *Handbook of educational measurement and assessment* (pp. 89–110). Springer. https://doi.org/10.1007/978-3-031-45981-5_6

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior, 89*, 98–110. https://doi.org/10.1016/j.chb.2018.07.027

von Davier, A. A. (2020). Computational psychometrics: New methodologies for a new generation of digital learning and assessment. Springer. https://doi.org/10.1007/978-3-030-74394-9

von Davier, A. A. (2023). Computational psychometrics: Toward ethical and explainable AI in education. Springer. https://doi.org/10.1007/978-3-031-27781-8

Wang, M. T., & Degol, J. L. (2024). Gender gap in STEM: Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review, 36*, 119–147. https://doi.org/10.1007/s10648-024-09850-1

West, D. M., Whitehurst, G. J., & Dionne, E. J. (2022). *Digital divide in education: Shifting inequities in U.S. schools*. Brookings Institution. https://www.brookings.edu/articles/the-digital-divide-in-education/

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

Yeager, D. S., Carroll, J. M., Buontempo, J., Cimpian, A., Wooden, S., Crosnoe, R., Muller, C., Murray, J., Mhatre, P., Kersting, N., Hulleman, C., Kudym, M., Murphy, M., Duckworth, A. L., Walton, G. M., & Dweck, C. S. (2023). Teacher mindsets help explain where a growth-mindset intervention does and doesn't work. *Psychological Science, 34*(5), 518–536. https://doi.org/10.1177/09567976231154 5

Zhang, S., Wang, Z., & Chen, Y. (2024). Diagnostic classification models for assessing scientific reasoning in K–12 contexts. *Applied Measurement in Education, 37*(2), 145–163. https://doi.org/10.1080/08957347.2023.2289012

Zimmerman, B. J. (2020). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2