The Technical Capability Fallacy: A Case for Pedagogy-Centered Design in Assistive STEM Education

Madjid Sadallah [a] and Benoît Encelle [b]
Universite Claude Bernard Lyon 1, CNRS, INSA Lyon, LIRIS, UMR5205, 69622, Villeurbanne, France
{*madjid.sadallah, benoit.encelle*}*@liris.cnrs.fr*

Keywords: STEM Education, Visual Impairment, Technology Adoption, Pedagogical Design, Assistive Technology

Abstract: Assistive technologies for blind and visually impaired STEM learners perform well in controlled studies yet fail to achieve sustained classroom adoption. Analysis of 25 empirical studies reveals the cause: tools are designed and evaluated against technical metrics that do not predict pedagogical success. Three patterns drive this gap—evaluation prioritizes technical performance over learning outcomes, research concentrates on a narrow range of objectives, and designs ignore institutional deployment contexts. We term this the *technical capability fallacy*: assuming technical capability ensures pedagogical utility. To address this, we propose FOCAL (*Framework for Objective–Context ALignment*), which specifies what a tool must provide for each combination of learning objective and educational context. FOCAL is not a validated instrument but a corrective heuristic grounded in learning theory. It inverts the design sequence: pedagogical diagnosis precedes technical development. Closing the adoption gap requires redefining success—not as technical accessibility, but as fit within the instructional, evaluative, and institutional constraints that govern classroom use.

1 INTRODUCTION

Educational technologies often show strong effects in controlled studies, yet these benefits rarely persist in real classrooms. The evidence that could guide adoption decisions is still limited, forcing schools and institutions to rely on local trials whose results rarely generalize (Hollands and Escueta, 2020). We call this the *technical capability fallacy*: the assumption that good performance in laboratory tasks predicts effective classroom use. A similar verification gap has been identified in AI-based accessibility systems, where technical accuracy does not guarantee users' ability to detect errors (Sadallah and Encelle, 2026b).

Multimodal tools for blind and visually impaired (BVI) students in STEM (Science, Technology, Engineering, and Mathematics) illustrate this problem. Controlled studies report promising outcomes (Brown et al., 2023; Wegwerth et al., 2023), yet classroom adoption remains rare. Teachers of students with visual impairments often report gaps in Assistive Technology (AT) competencies (Zhou et al., 2011) and confidence (Ajuwon et al., 2016)—conditions that predict brief engagement followed by discontinuation. The consistency of this pattern across tools and settings suggests a problem in design rather than implementation or training.

The consequences are concrete. When these tools do not reach classrooms, BVI students lose opportunities to develop STEM skills and access technical careers. Repeated failures intensify the problem, reducing investment, creating skepticism among teachers, and leaving educators without practical alternatives; in effect, innovation accumulates without benefiting those who need it most. These failures are not technical: tools often focus on what is easy to measure—accuracy, speed, or error rate—rather than what matters for real classroom use, such as alignment with learning objectives, curriculum requirements, and assessment practices. Research incentives reinforce this pattern, as technical novelty is more publishable than pedagogical integration, and short-term performance measures are easier to collect than long-term adoption data.

This paper advances three claims. First, current design and evaluation practices prioritize technical metrics over pedagogical requirements, creating a gap between tool capabilities and teacher needs. Second, this gap persists because research and funding systems reward technical novelty more than educational usefulness. Third, closing it requires pedagogy-

[a] https://orcid.org/0000-0001-9118-0235
[b] https://orcid.org/0000-0002-0734-6480

centered design, where learning objectives and context precede technical development.

To support these claims, we analyze 25 empirical studies on multimodal BVI STEM tools and introduce FOCAL (*Framework for Objective–Context ALignment*), a diagnostic framework examining how learning objectives and educational contexts shape design requirements. FOCAL clarifies what a tool must provide—and what constraints it must address—for each objective–context pairing. It is not a validated standard, but a pragmatic guide for prioritizing pedagogical needs over technical constraints.

2 METHODOLOGY

Our analysis draws on empirical studies of multimodal BVI STEM tools in educational contexts. A systematic search (2015–2025) across ACM Digital Library, IEEE Xplore, PubMed, OpenAlex, and Web of Science yielded 66 publications (see (Sadallah and Encelle, 2026a)); 25 focused on formal learning environments, excluding non-classroom and non-formal learning contexts. Two researchers conducted a thematic analysis (Cohen's $\kappa = 0.84$), identifying patterns in design assumptions, pedagogical alignment, and implementation conditions.

This analysis reveals three patterns through which the fallacy operates: tools are evaluated on metrics that do not predict classroom use, research attention concentrates unevenly across learning objectives, and tool design ignores the institutional contexts in which deployment must occur. Each pattern is examined in turn, followed by an analysis of why they persist.

2.1 Evaluation-Practice Mismatch

Analysis reveals that 76% of studies foreground technical performance metrics—accuracy, speed, precision—as primary outcomes, whereas only 24% assess pedagogical outcomes such as learning gains or conceptual understanding. This emphasis is reflected in research settings: 48% of studies were conducted in controlled laboratories and 48% in authentic classrooms, yet only four adopted longitudinal designs exceeding one month—adoption failures take months to emerge, yet most studies conclude before this.

Research priorities contradict reported adoption barriers. Technical usability emerges as the main obstacle in only one study, while 88% cite pedagogical or institutional constraints: curriculum alignment failures (32%), incompatible assessment outputs (28%), insufficient teacher preparation (88%), and funding instability (72%).

This pattern is clearest in studies of analytical reasoning (n=14, the largest subset): three reported quantitative task accuracy (mean = 81.4%), yet all exhibited low sustained adoption due to curriculum pacing incompatibility (32%), mismatched assessment requirements (28%), or classroom logistics constraints (24%). Strong technical performance, without attention to how tools fit classroom needs, does not predict sustained use. A clear indicator is that only two of the ten laboratory studies examined teacher training needs or autonomous classroom deployment—precisely the factors that determine sustained use. No study frames its design around a stated learning objective; tools are described by their features rather than by their pedagogical purpose.

2.2 Learning Objective Imbalance

Classifying studies by learning objective (Kilpatrick et al., 2001) reveals pronounced imbalance. Analytical reasoning dominates (56%, 14/25), reflecting ease of measuring problem-solving performance. Creative expression receives moderate attention (28%, 7/25). Procedural fluency and conceptual understanding—the building blocks of STEM learning—remain critically underrepresented (4% and 8% respectively).

This imbalance has significant consequences. Empirical studies focus predominantly on analytical reasoning, whereas guidance for procedural fluency and conceptual understanding relies mainly on theoretical extrapolation. Yet STEM learning generally develops procedural fluency and conceptual understanding first, forming the foundation for analytical reasoning (Kilpatrick et al., 2001). As a result, current research emphasizes advanced objectives while overlooking the foundational skills on which they depend.

2.3 Context-Blind Design

Tools are typically developed without explicit context specification, yet analysis reveals distinct barrier profiles and pedagogical emphases across elementary (K–5), secondary (9–12), and postsecondary levels.

Five elementary studies emphasize teacher-mediated discovery and concrete manipulation. The main barrier is teacher preparation time: one elementary math tool achieved 89% task accuracy in controlled sessions but only 12% sustained adoption because teachers could not learn mediation strategies within 45-minute classes.

Fifteen secondary studies center on assessment-driven instruction and curriculum compliance. Yet none examined whether tool outputs match required test answer formats—a striking omission given that

28% of studies cite this as a key adoption barrier. One geometry tool improved spatial reasoning by 34% but was discontinued because it could not export diagrams in the format required by standardized tests.

Eleven postsecondary studies emphasize professional workflow integration and independent use, yet only two occurred in authentic classroom settings. Barriers shift toward funding sustainability (cited in seven of the eleven) and compatibility with professional tools and standards. These differences run deep: a tool designed for independent postsecondary analytical reasoning may fail entirely in teacher-mediated elementary conceptual understanding—not because it is technically inadequate, but because it was built for a different context.

2.4 Why the Fallacy Persists

The fallacy persists because of how research is rewarded. New multimodal techniques are more publishable than studies of pedagogical integration. Accuracy and completion time are measurable in controlled sessions; meaningful learning outcomes require longitudinal, classroom-based investigation.

The result is a persistent *measurability bias*. Technical metrics (latency, accuracy, error rate) are quantitative, immediate, and well aligned with conference timelines. Pedagogical outcomes (conceptual change, transfer, autonomous use) are longitudinal, context-dependent, and harder to measure. Our corpus reflects this: only 16% of studies (4/25) extend beyond one month—too short to observe adoption failures, which typically emerge over months of classroom use (Hollands and Escueta, 2020). The mechanism is not intentional: controlled laboratory sessions yield data within weeks, whereas classroom adoption requires months of observation across teachers and cohorts. Technical metrics fill this gap not because they predict learning, but because they are available when pedagogical outcomes are not.

The result is a feedback loop: technically oriented publications set the standard, subsequent work follows the same model, and funding rewards more of the same. Breaking this loop requires a shift toward pedagogy-centered design. Technical capability is a prerequisite, not a guarantee: making a tool accessible does not make it educationally effective.

3 THE FOCAL FRAMEWORK

The three patterns we documented point to a common cause: tools are developed without first specifying the learning objective they are meant to serve and the educational context in which they must operate. FOCAL is structured around these two guiding principles.

3.1 From Features to Pedagogical Fit

Current tool development follows a feature-centered logic: implement capabilities, demonstrate proficiency, then seek applications. FOCAL reverses this process, beginning with pedagogical diagnosis—identifying the target learning objective within a given context—then defining what the tool must do, grounded in learning theory.

This reframing has three practical implications. First, pedagogical requirements vary across objective-context combinations: synchronized multimodal representations may support secondary conceptual understanding yet hinder elementary procedural fluency through split-attention effects. Second, essential versus optional features become explicit: FOCAL pushes designers to ask what a tool must do to serve the learning objective, not what it can do to impress in a demonstration. Third, implementation barriers become design priorities: if teacher preparation time is the main obstacle, minimizing training must be a design goal, not an afterthought.

3.2 Framework Architecture

FOCAL is structured as a 4×3 decision matrix linking learning objectives to educational contexts. Two findings from the corpus motivate this structure. First, pedagogical goals and contextual constraints vary enough across settings that a single design cannot serve all of them. Second, tools built without explicit context specification consistently fail when deployed outside their assumed context, indicating that these differences cannot be addressed by a single adjustable design. Each cell specifies what a tool must offer and what barriers it must address—derived from learning theory, not inferred from observed failures alone.

Learning Objectives Dimension. FOCAL's learning objectives build on Kilpatrick et al.'s framework for mathematical proficiency (Kilpatrick et al., 2001), which identifies five interrelated strands: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition. We adapt this framework for assistive STEM technology design in four ways.

First, we generalize from mathematics to STEM, as the underlying cognitive processes—principle comprehension, procedural execution, problem solving, and reasoning—extend across domains. Second,

we merge *strategic competence* and *adaptive reasoning* into *analytical reasoning*, since assistive tools typically couple hypothesis testing with evidence-based justification. Third, we exclude *productive disposition*, which concerns beliefs and engagement rather than cognitive processes that tools can directly support. Fourth, we add *creative expression*—absent from Kilpatrick's model—to capture generative STEM production (experimental design, coding, visualization) requiring features like versioning and iterative refinement.

The resulting taxonomy yields four objectives, each defined by a distinct cognitive process and corresponding tool requirement.

1. *Procedural Fluency*: executing routines accurately and efficiently—balancing equations, navigating coordinate systems, applying algorithms. The cognitive demand is repetition without overload; tools should offer streamlined interfaces that minimize unnecessary steps.
2. *Conceptual Understanding*: grasping the principles behind procedures—knowing not just how to compute a derivative, but what it represents. Transfer to new problems depends on this layer. Tools should present the same concept through multiple linked representations so that relationships, not just results, become visible.
3. *Analytical Reasoning*: using knowledge to interpret evidence, detect patterns, and construct arguments. Where conceptual understanding builds a mental model, analytical reasoning puts it to work. Tools should support open-ended exploration and hypothesis testing rather than guiding toward predetermined answers.
4. *Creative Expression* involves generating original solutions within disciplinary constraints, such as designing an experiment, writing a simulation, or constructing a visualization. Tools must support iterative work by allowing intermediate states to be saved, decisions to be reversed without penalty, and outputs to be refined progressively.

These objectives are complementary, not hierarchical. While procedural and conceptual foundations often precede analytical and creative skills (Kilpatrick et al., 2001), expert performance integrates all four. Because they impose distinct and sometimes competing design requirements, FOCAL requires identifying a target objective before development begins.

Educational Context Dimension. Educational contexts differ not merely by age but by institutional norms, assessment regimes, and adoption dynamics. Three contexts impose distinct design constraints.

- *Elementary (K–5)* emphasizes concrete reasoning and teacher-guided exploration. Tools must enable easy setup with clear pedagogical cues; even capable tools fail if teachers cannot quickly learn how to use them within a 45-minute class.
- *Secondary (9–12)* focuses on abstract reasoning and standardized assessment. Tools must follow curricular frameworks (e.g., Common Core, NGSS) and produce outputs compatible with testing contexts. Adoption depends on both instructional value and institutional recognition.
- *Postsecondary* learning emphasizes authentic practice and professional integration. Tools should align with real-world workflows (e.g., LaTeX, version control), support independent use, and remain sustainable over time, with constraints shifting from curricular alignment toward institutional and professional feasibility.

An elementary procedural tool differs from a postsecondary one not by learner age, but by what institutions require. FOCAL treats these constraints as design requirements to address upfront, not as obstacles discovered after deployment.

3.3 Theoretical Positioning

FOCAL links learning theory to concrete design choices while explaining adoption patterns and barriers that existing frameworks do not fully capture. *Cognitive Load Theory* (Sweller, 2011) explains how information presentation affects learning difficulty; FOCAL applies this by choosing how information is presented across senses—synchronized audio-tactile feedback for conceptual understanding reduces split-attention, while redundant cues for procedural tasks minimize cognitive load. *Universal Design for Learning (UDL)* (CAST, 2018) advocates multiple ways to access, engage with, and express learning; FOCAL extends UDL by specifying *which* representations support *which* learning objectives in *which* contexts, translating multiple means into actionable design prescriptions. *Activity Theory* (Engeström, 1987) analyzes how tools operate within social systems; FOCAL uses this lens to explain adoption failures—tools often fail not because of technical flaws but because they conflict with classroom constraints such as assessment requirements, teacher time, and institutional workflows. Validated primarily in sighted populations, these theories remain hypotheses when applied to multimodal BVI contexts and require empirical testing.

At the *framework level*, FOCAL complements rather than replaces existing models. While UDL prescribes *what* to provide, it offers limited guidance on

how to configure tools for specific objectives; FOCAL addresses this gap through objective–context prescriptions. The SETT Framework (Zabala, 2005) optimizes tool selection for individual students with disabilities, whereas FOCAL operates upstream, embedding contextual constraints so that technologies satisfy SETT criteria by design. Technology Acceptance Models (Venkatesh et al., 2003) predict adoption based on perceived usefulness but treat it as a single dimension; FOCAL disaggregates usefulness by objective–context, clarifying why the same tool may succeed in one setting yet fail in another. For example, a gamified arithmetic app rejected in secondary classrooms is not inherently "low usefulness"—it is an elementary procedural tool applied inappropriately to a secondary analytical context.

3.4 Framework Specification

Table 1 presents the full FOCAL matrix. Each cell specifies what a tool must provide and which barrier it must address. These prescriptions are grounded in the three theories introduced above: Cognitive Load Theory determines how information should be presented for each objective, UDL determines which representational combinations support access, and Activity Theory identifies the institutional conditions that make a tool viable in practice.

Prescriptions are provided for all twelve cells. Where empirical evidence is available—analytical reasoning (56% of studies) and creative expression (28%)—they reflect observed classroom practices. Where evidence is limited—procedural fluency (4%) and conceptual understanding (8%)—they rest on learning theory.

3.4.1 Decision Protocol

FOCAL works through five steps.

1. *Pedagogical Diagnosis* identifies the target learning objective. In mixed-objective units (e.g., algebra combining equation manipulation and function transformation), priority follows assessment emphasis. This matters: tools optimized for procedural fluency (streamlined interfaces) conflict with those targeting conceptual understanding (coordinated views across formats).
2. *Context Mapping* situates the objective within K–5, 9–12, or postsecondary settings, each defined by distinct constraints: assessment requirements in secondary education, teacher preparation time in elementary grades, and professional workflow integration at the postsecondary level.
3. *Framework Consultation* extracts required features and the main barrier from Table 1. These are necessary conditions, not optional extras: a tool missing any of them is unlikely to succeed pedagogically, regardless of its technical quality.
4. *Tool Evaluation* works eliminatively: any tool lacking a required feature is rejected. A postsecondary analytical reasoning tool without professional export formats fails regardless of its visualization quality—students cannot transfer their work to authentic professional contexts.
5. *Barrier Mitigation* addresses the main obstacle identified in the matrix—embedded tutorials for elementary preparation constraints, documentation of output formats for secondary assessment alignment, or licensing negotiation for postsecondary institutional barriers.

FOCAL is iterative: conflicting affordance requirements between steps signal that the initial objective identification needs revisiting.

3.4.2 Illustrative Cases

We examine three tools to show how objective–context alignment determines success or failure.

MathMelodies (Ahmetovic et al., 2017) targets elementary procedural fluency through game-based arithmetic practice. The tool provides musical rewards for correct answers, dissonant tones for errors, and progress tracking through level advancement. Classroom outcomes showed sustained engagement and teacher-requested continuation beyond initial pilots—consistent with FOCAL's specification of what elementary procedural fluency requires: game-like feedback, immediate error correction, and tangible progress tracking.

To demonstrate how FOCAL can be actively applied to evaluate or guide design, consider *IrisMath* (Zambrano et al., 2023) through our five-step protocol: (1) *Diagnosis*: The tool targets conceptual understanding of algebra. (2) *Context Mapping*: It is designed for postsecondary environments. (3) *Framework Consultation*: Table 1 dictates that such an objective-context pair *must* provide synchronized views across multiple formats (essential) and address ongoing professional development/learning curves (barrier). (4) *Tool Evaluation*: *IrisMath* satisfies this by coordinating audio, tactile braille, and navigable tree structures. (5) *Barrier Mitigation*: It structures learning progressively to manage the cognitive load of these novel interfaces. As a result, users' understanding of equation hierarchy improved from 42% to 78%. Because it successfully executed

Table 1: The FOCAL Decision Matrix. Each cell specifies essential features (required for pedagogical viability) and the main adoption barrier to address.

	K–5 (Elementary)	**9–12 (Secondary)**	**Postsecondary**
Procedural Fluency	**Essential:** Game-like feedback, immediate error correction, tangible progress tracking. **Barrier:** Teacher preparation time.	**Essential:** Keyboard shortcuts, test-compatible outputs, rapid input mechanisms. **Barrier:** Standardized assessment alignment.	**Essential:** LaTeX / IDE integration, customizable macros, version control. **Barrier:** Long-term funding sustainability.
Conceptual Understanding	**Essential:** Virtual manipulatives, causally linked audio-tactile feedback, guided exploration. **Barrier:** Teacher pedagogical training.	**Essential:** Synchronized views across multiple formats, dynamic parameter manipulation, comparative displays. **Barrier:** Curriculum integration.	**Essential:** Data-to-sound conversion, simulation capabilities, interaction with formal models. **Barrier:** Ongoing professional development.
Analytical Reasoning	**Essential:** Basic sorting/filtering, auditory highlighting, pattern salience. **Barrier:** Age-appropriate complexity.	**Essential:** Evidence comparison, hypothesis testing support, argument construction. **Barrier:** Evaluation rubric alignment.	**Essential:** Statistical analysis, professional export, reproducible methods. **Barrier:** Institutional licensing costs.
Creative Expression	**Essential:** Constrained building blocks, unlimited undo/redo, nondestructive exploration. **Barrier:** Open-ended assessment practices.	**Essential:** Customizable templates, version history, iterative refinement. **Barrier:** Subjective evaluation criteria.	**Essential:** Full authoring control, collaborative features, professional export. **Barrier:** Industry standards compliance.

Note: Prescriptions for analytical (56%) and creative (28%) objectives draw primarily from empirical evidence; those for procedural (4%) and conceptual (8%) draw more from learning theory.

the FOCAL pathway, its pedagogical effectiveness aligned with its technical capability.

The third case illustrates misalignment. An auto-assisting figure presentation system (Onishi et al., 2016) for educational contexts combined haptic guidance (PHANTOM Omni device), tactile displays, and speech to support exploration of 2D diagrams. Users reported that guided haptic exploration aided their understanding of spatial relationships. However, deployment faced severe adoption barriers: the system required specialized hardware (haptic device, tactile display) whose cost and maintenance burden made classroom integration prohibitive. FOCAL explains this pattern: the tool provided affordances suited to research laboratories or well-funded postsecondary programs, but was deployed in a secondary context where the critical constraint is not pedagogical sophistication but institutional feasibility—specifically, hardware costs and teacher training requirements. The system succeeded technically yet failed ecologically because it addressed the wrong set of constraints for its deployment context.

Across all three cases, the pattern holds: pedagogically aligned tools persist in their intended contexts, while contextually mismatched tools fail regardless of technical quality. These analyses remain retrospective. Establishing FOCAL's prospective value requires comparing tools designed with versus without the framework in authentic classroom settings.

4 DISCUSSION

The gap between technical capability and classroom adoption reflects a persistent misalignment: assistive technologies are optimized against metrics that do not predict pedagogical success. FOCAL addresses this by inverting the design sequence—pedagogical diagnosis becomes the precondition for technical development, rather than an afterthought—and by framing technology as a means to achieve learning goals within real institutional constraints.

4.1 Implications for Practice

This inversion carries distinct implications for each stakeholder group.

For *researchers*, prevailing evaluation protocols overvalue technical metrics whose predictive validity for classroom success remains unproven. Longitudinal, objective-aligned assessments in authentic contexts are not merely a methodological preference—they are the only way to determine whether a tool actually works. The field's overrepresentation

of analytical reasoning—at the expense of procedural fluency and conceptual understanding—further biases evidence toward what is measurable rather than what is educationally consequential.

For *developers*, feature-driven design assumes that technical capability will naturally find its pedagogical use. Pedagogy-centered design rejects this premise. It requires prior engagement with learning theory, curricular standards, and the constraints summarized in Table 1. Innovation is guided not by technical feasibility alone but by instructional relevance: only features that serve defined objectives under defined constraints warrant implementation.

For *funders*, incentive structures that privilege novelty and short-term performance gains undervalue integration, sustainability, and durability. Requiring explicit objective-context specification in proposals, supporting multi-year evaluation horizons, and funding professional development alongside deployment would realign investment with educational impact.

4.2 Methodological Tensions

Adopting a pedagogy-centered paradigm does not dissolve methodological difficulty—it relocates it. Three tensions warrant explicit acknowledgment.

The first concerns *ecological validity*. Technical metrics are insufficient predictors of adoption, yet abandoning them entirely risks replacing one inadequacy with another. The core problem is the absence of evidence linking measured performance improvements to actual learning outcomes. Until longitudinal studies map these relationships with sufficient precision, the field oscillates between metrics that are measurable but misleading and outcomes that are meaningful but elusive. FOCAL does not resolve this tension; it makes it visible and therefore researchable.

The second tension opposes *prescriptive specificity* and *generalizability*. FOCAL's detailed prescriptions—such as synchronized audio-tactile feedback for conceptual understanding objectives—provide implementable guidance, but carry the risk of overfitting to the evidential base from which they were derived. Prescriptions grounded in limited samples may fail under different population profiles, curricular contexts, or pedagogical traditions. The appropriate response is not to retreat into vague heuristics, but to treat prescriptions as working hypotheses that evidence can revise.

The third tension involves *necessary reductionism*. Representing continuous classroom complexity through a discrete matrix enables actionable guidance at the cost of simplification. The risk is not the matrix itself, but its misuse: mechanical cell-completion that satisfies the framework's categories without engaging the situated dynamics that give those categories meaning. FOCAL's value lies not in exhaustive categorization but in provoking intentional design—careful consideration of how objectives, context, and features interact in specific learning environments.

4.3 Validation and Research Agenda

FOCAL generates testable propositions at three levels. At the level of *descriptive validity*, the question is whether the objective–context distinctions it identifies correspond to genuine differences in learning processes and outcomes. At the level of *prescriptive utility*, the focus is on whether FOCAL-guided design decisions demonstrably improve tool selection, adoption rates, and classroom integration compared with feature-driven alternatives. At the level of *learning outcome validation*, controlled studies should employ objective-aligned assessments: automaticity measures for procedural fluency, transfer tasks for conceptual understanding, and structured reasoning protocols for analytical objectives.

Comprehensive validation across all twelve cells will require longitudinal effort across multiple research groups. The central question is whether designing from learning objectives and educational contexts produces better classroom outcomes than designing from technical features—a question the field can and should answer empirically.

4.4 Broader Relevance and Limitations

Although derived from BVI STEM education, the *technical capability fallacy* is not domain-specific. Across computer-supported learning environments—AI tutoring systems, adaptive recommendation engines—the same tendency recurs: computational sophistication is conflated with pedagogical effectiveness, and system performance metrics substitute for evidence of learning. FOCAL offers a transferable design approach: technological affordance must be evaluated against instructional intent, within the constraints of real classroom contexts.

Several limitations must be acknowledged. The literature FOCAL synthesizes likely overrepresents technical successes, potentially exaggerating the capability-adoption gap it diagnoses. Illustrative cases carry the risk of post-hoc rationalization—fitting the framework to available evidence rather than testing it against adverse cases. Many prescriptions rest on learning theories developed with sighted populations whose applicability to multimodal, non-visual contexts remains empirically un-

dertested. Generalization beyond BVI STEM education, while theoretically motivated, remains speculative pending direct investigation. FOCAL should be understood as a structured conceptual hypothesis—an instrument for generating and organizing research questions, not a validated decision tool for immediate deployment.

4.5 Conclusion

The persistent gap between laboratory performance and classroom adoption of assistive STEM technologies stems from a systematic misalignment: tools are optimized against technical metrics that do not predict pedagogical success. To address this, we introduced FOCAL, a framework that inverts the design sequence by placing learning objectives and educational contexts before technical development. By specifying what a tool must provide for each objective–context pair, FOCAL shifts the unit of analysis from isolated technical features to systemic fit within instructional, evaluative, and institutional constraints.

Our analysis of 25 empirical studies demonstrates that the *technical capability fallacy* is not inevitable but remediable through intentional design. For researchers, this means adopting longitudinal, objective-aligned evaluation protocols; for developers, it requires engaging with learning theory and context specifications before implementation; for funders, it demands rebalancing incentives toward sustainability and integration. FOCAL is offered as a conceptual heuristic—a structured hypothesis to be tested and refined. The coming decade must measure success not by what assistive technologies can do in controlled settings, but by what learners actually achieve with them in real classrooms.

ACKNOWLEDGEMENTS

This work is part of the COOBRA project, funded by the Agence Nationale de la Recherche (grant ANR-24-SARP-0006).

REFERENCES

Ahmetovic, D., Alampi, V., Bernareggi, C., Gerino, A., and Mascetti, S. (2017). Math melodies: Supporting visually impaired primary school students in learning math. In *Proceedings of the 14th International Web for All Conference*, pages 1–2.

Ajuwon, P. M., Meeks, M. K., Griffin-Shirley, N., and Okungu, P. A. (2016). Reflections of teachers of visually impaired students on their assistive technology competencies. *Journal of Visual Impairment & Blindness*, 110(2):128–140.

Brown, J. R., Doore, S. A., Dimmel, J. K., and Giudice, N. A. (2023). Comparing natural language and vibro-audio modalities for inclusive stem learning with blind and low vision users. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '23. ACM.

CAST (2018). *Universal Design for Learning Guidelines version 2.2*. CAST, Wakefield, MA.

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Orienta-Konsultit, Helsinki, Finland.

Hollands, F. and Escueta, M. (2020). How research informs educational technology decision-making in higher education: the role of external research versus internal research. *Educational Technology Research and Development*, 68(1):163–180.

Kilpatrick, J., Swafford, J., and Findell, B. (2001). *Adding it up: Helping children learn mathematics*. National Academy Press, Washington, DC. National Research Council.

Onishi, J., Sakai, T., Sakajiri, M., Ogata, A., Miura, T., Handa, T., Shimizu, T., and Ono, T. (2016). Auto-assisting figure presentation system for inclusion education. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 000176–000181. IEEE.

Sadallah, M. and Encelle, B. (2026a). Beyond additive design: An empirical taxonomy of multimodal STEM accessibility systems. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA 2026)*, Barcelona, Spain. ACM.

Sadallah, M. and Encelle, B. (2026b). The verification gap in AI accessibility. In *Proceedings of the 23rd International Web for All Conference (W4A 2026)*, Dubai, UAE. ACM.

Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation*, 55:37–76.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3):425–478.

Wegwerth, S. E., Olson, E., Sarupria, S., Ballard, H. E., and Wilson, M. A. (2023). Feasibility of using multimodal virtual reality to facilitate molecular biology learning for blind and low vision students. *Journal of Chemical Education*, 100(2):730–739.

Zabala, J. (2005). Using the sett framework to level the learning field for students with disabilities. *Closing the Gap*, 23(1):112–120.

Zambrano, A. M., Pilacuán, D. I., Salvador, M. N., Grijalva, F., Garzon, N. O., and Mora, H. C. (2023). Irismath: a blind-friendly web-based computer algebra system. *IEEE Access*, 11:71766–71776.

Zhou, L., Parker, A. T., Smith, D. W., and Griffin-Shirley, N. (2011). Assistive technology for students with visual impairments: Challenges and needs in teachers' preparation programs and practice. *Journal of Visual Impairment & Blindness*, 105(4):197–210.