International Research Journal of
SCIENCE, TECHNOLOGY, EDUCATION, AND MANAGEMENT
P-ISSN: 2799-063X — E-ISSN: 2799-0648
Volume 6, No. 1 | March 2026

An Artificial Intelligence (AI)-Powered Voice-Based Intelligent Learning System for Visually Impaired Students

Rizky Rinaldi[1]
[1]Raudhatul Akmal Islamic College, Deli Serdang, North Sumatra, Indonesia
rizkyrinaldi.staira@gmail.com[1]

ABSTRACT

The rapid expansion of digital learning platforms has created unprecedented educational opportunities; however, the majority of these platforms remain inaccessible to the estimated 253 million visually impaired individuals worldwide, particularly students in inclusive K-12 settings. Existing assistive technologies such as screen readers and Braille displays function as access tools rather than pedagogically designed learning environments, leaving a critical gap in inclusive educational technology. This paper presents BlindLearn, an AI-powered, voice-based learning framework developed and evaluated using Design Science Research (DSR) methodology. Grounded in Universal Design for Learning (UDL) and Cognitive Load Theory (CLT), BlindLearn introduces the Voice-First Pedagogical Model (VFPM) — a novel five-stage learning cycle (Audio Activation, Narrative Input, Conversational Elaboration, Voice Practice, Adaptive Feedback) designed for auditory-primary learners. The framework was developed through systematic literature review (47 papers, 2015–2024), structured needs analysis ($n = 23$), multi-expert validation using Content Validity Ratio (CVR, $n = 8$), and usability evaluation using the System Usability Scale (SUS, $n = 15$). Expert validation yielded a mean CVR of 0.89 ($p < .05$), and usability evaluation produced a mean SUS score of 84.3 (Grade: Excellent). Three original artifacts are contributed: the VFPM theoretical model, a validated four-layer AI system architecture, and twelve evidence-based inclusive design guidelines, advancing the fields of educational technology and inclusive AI system design.

ARTICLE INFO

Received : Mar. 7, 2026
Revised : Mar. 27, 2026
Accepted : Apr. 15, 2026

KEYWORDS

Artificial intelligence, Inclusive education, Universal design for learning, Visually impaired students, Voice-based learning

Suggested Citation (APA Style 7th Edition):

Rinaldi. R. (2026). An Artificial Intelligence (AI)-Powered Voice-Based Intelligent Learning System for Visually Impaired Students. *International Research Journal of Science, Technology, Education, and Management*, 6(1), 1-13. https://doi.org/10.5281/zenodo.19674817

INTRODUCTION

The global adoption of digital learning platforms accelerated substantially following the COVID-19 pandemic, with UNESCO (2023) reporting that over 1.5 billion students engaged with digital educational tools by 2022. Yet this digital transformation has disproportionately excluded learners with visual impairments. According to the World Health Organization (2023), approximately 253 million people worldwide live with vision impairment or blindness, a significant proportion of whom are school-aged children in inclusive education settings. The intersection of rapidly expanding edtech infrastructure and persistent accessibility failures has created an urgent imperative for inclusive-by-design learning solutions (Haddad & Mulhem, 2022).

Contemporary digital learning platforms exhibit three systemic accessibility failures that disproportionately affect blind learners. First, interface designs remain predominantly visual-centric, relying on graphical menus, icon-based navigation, and visual content rendering that are fundamentally incompatible with screen reader workflows (Power et al., 2012). Second, platform architectures lack native support for natural voice interaction as a primary learning modality, treating audio as an afterthought rather than a design priority (Calvo et al., 2014). Third, adaptive pedagogical mechanisms within existing platforms are not designed to accommodate the specific cognitive and interaction patterns of learners who rely exclusively on auditory channels (Fernández-Batanero et al., 2022).

While assistive technologies such as screen readers, audio books, and Braille displays have partially addressed access barriers, they share a fundamental limitation: they are designed as access tools that retrofit visual content for blind users rather than as learning environments built from first principles for auditory-primary learners (Hersh, 2022). Alotaibi (2021) confirmed in a systematic review of assistive mobile applications that existing tools consistently fail to provide the interactive, adaptive pedagogical scaffolding that characterises effective learning environments. The distinction between accessibility as a retrofit and accessibility as a foundational design paradigm constitutes the central problem this study addresses (Haddad & Mulhem, 2022).

Despite significant advances in artificial intelligence (AI), natural language processing, and voice user interface (VUI) technologies, no existing learning platform integrates AI-driven conversational tutoring, adaptive audio pedagogy, and inclusive learning design into a unified ecosystem built specifically for visually impaired students (Haddad & Mulhem, 2022). Intelligent Tutoring Systems (ITS) represent the state of the art in adaptive learning (Mousavinasab et al., 2021), yet virtually none are designed with accessibility for blind learners as a primary requirement. Voice assistants demonstrate the maturity of VUI technology (Porcheron et al., 2018), yet lack the pedagogical scaffolding necessary for structured learning. Universal Design for Learning (CAST, 2018) provides a robust theoretical framework for inclusive education, yet its implementation in AI-driven system design remains underexplored (Zawacki-Richter et al., 2019).

A systematic review of 312 candidate studies, reduced to 47 after PRISMA screening (Page et al., 2021), confirmed this gap: while individual components of voice interaction, AI tutoring, and educational accessibility each have extensive research bases, no empirical study has integrated these three dimensions into a single, pedagogically grounded platform designed for visually impaired students. This study responds to that gap by presenting BlindLearn — an AI-powered, voice-based learning framework developed through Design Science Research (DSR; Peffers et al., 2007) — pursuing five objectives: (1) to synthesise existing knowledge through systematic literature review; (2) to identify user needs through structured needs analysis; (3) to design and develop the BlindLearn framework and prototype; (4) to validate the framework via Content Validity Ratio analysis (Lawshe, 1975); and (5) to evaluate prototype usability using the System Usability Scale (Brooke, 1996).

OBJECTIVES OF THE STUDY

The general objective of this study is to design, develop, and validate BlindLearn — an AI-powered, voice-first learning framework specifically engineered for visually impaired students in inclusive K-12 settings. The study pursues five specific objectives:

(1) to synthesise existing knowledge on voice-based learning, AI tutoring systems, and inclusive design through a systematic literature review;
(2) to identify the learning needs and technological challenges faced by visually impaired students and their teachers through structured needs analysis;
(3) to design and develop the BlindLearn prototype incorporating the Voice-First Pedagogical Model (VFPM) and a four-layer AI architecture;
(4) to validate the framework using Content Validity Ratio (CVR) analysis with domain experts; and
(5) to evaluate prototype usability with visually impaired student participants using the System Usability Scale (SUS). Together, these objectives aim to produce three transferable artifacts — a theoretical model, a system architecture, and inclusive design guidelines — that advance inclusive educational technology design.

Theoretical Foundations

BlindLearn is grounded in two complementary theoretical frameworks. Universal Design for Learning (UDL), formalised by CAST (2018), provides the primary pedagogical foundation. The three UDL principles are operationalised as follows: Multiple Means of Representation are implemented through audio-native content delivery; Multiple Means of Action and Expression are realised through voice-based interaction as the primary modality; and Multiple Means of Engagement are supported through AI-adaptive scaffolding that responds dynamically to learner performance, confidence, and response latency. Florian (2014) argues that UDL implementation must be grounded in evidence-based practice — a principle that directly informed BlindLearn's design requirements specification.

Sweller's (1988) Cognitive Load Theory (CLT) informs the information architecture of BlindLearn's audio interface. The modality effect (Mayer, 2009) justifies exclusive use of the auditory channel to prevent split-attention between visual and auditory streams. The redundancy effect motivates the elimination of text-to-speech conversion of originally visual content. The worked example effect (Roediger & Karpicke, 2006) underpins the AI tutor's demonstration-before-practice interaction sequence. Recent empirical work by Sweller et al. (2019) further substantiated the importance of minimising extraneous load in multimedia learning environments, reinforcing the audio-native design philosophy adopted in this study.

METHODS

Research Design

This study adopts the Design Science Research (DSR) process model of Peffers et al. (2007), comprising six iterative phases: Problem Identification, Objectives Definition, Design and Development, Demonstration, Evaluation, and Communication. DSR is epistemologically appropriate when the research objective is to create a novel artifact that solves a real-world problem through iterative design and evaluation (Hevner et al., 2004). Gregor and Hevner (2013) classify this study in the "Improvement" quadrant: mature problem knowledge combined with a nascent solution. Table 1 maps each DSR phase to its activities and outputs.

Table 1. *DSR process model applied to BlindLearn*

Phase	Activity	Output
P1: Problem	Systematic review (PRISMA; Page et al., 2021) + needs analysis	Problem statement, user requirements
P2: Objectives	UDL/CLT mapping (CAST, 2018; Sweller, 1988)	Design requirements specification
P3: Design	Iterative prototype development (3 cycles)	BlindLearn v1.0
P4: Demo	Walkthrough with 3 learning modules	Functional demonstration
P5: Evaluation	Expert CVR validation (Lawshe, 1975) + SUS (Brooke, 1996)	CVR scores + SUS score
P6: Communication	Manuscript preparation (Gregor & Hevner, 2013)	This article

Participants

Needs analysis and usability evaluation were conducted with two participant groups. The first group comprised visually impaired students (n = 15; age range 10–16 years; attending inclusive schools in Deli Serdang Regency, North Sumatra). The second group comprised inclusive education teachers (n = 8; minimum three years' experience with visually impaired students). Purposive, criterion-based sampling was employed, consistent with DSR needs analysis practices (Peffers et al., 2007), to ensure participants had direct experience with digital learning tools. All participants provided informed consent (or parental assent for minors). Ethics approval was obtained from the Research Ethics Committee of STAI Raudhatul Akmal (Protocol No. STAIRA-REC-2023-047). The demographic homogeneity of the sample — Indonesian students in a single regency — is acknowledged as a limitation and is addressed in the Discussion section.

Needs Analysis

Semi-structured interviews (45–60 minutes) were conducted with each participant, guided by a protocol developed from the systematic review findings. Interviews were recorded, transcribed, and analysed using reflexive thematic analysis (Braun & Clarke, 2006), with member-checking conducted to ensure interpretive validity. Four primary themes emerged: **(T1) frustration with screen reader incompatibility in learning platforms**, consistent with Power et al. (2012); **(T2) desire for interactive rather than passive audio content**, echoing Wald et al. (2020); **(T3) need for patient, non-judgmental AI assistance**; and **(T4) importance of immediate, specific feedback**, aligned with Hattie and Timperley's (2007) feedback framework. These themes directly informed the five stages of the Voice-First Pedagogical Model described in the Results section.

Prototype Development

The BlindLearn prototype was developed iteratively over three design cycles, each informed by evaluation of the preceding prototype against the design requirements specification. The prototype was implemented as an Android mobile application using a React Native frontend and a Python/FastAPI backend. Speech recognition was implemented using wav2vec 2.0 (Baevski et al., 2020), selected for its state-of-the-art performance on conversational speech and self-supervised adaptation to individual speakers. Neural text-to-speech used a Tacotron 2/WaveNet architecture (Shen et al., 2018), producing naturalistic voice output with controllable prosody. Three learning modules were developed: mathematics (arithmetic operations), language arts (reading comprehension via audio narrative), and science (conceptual inquiry), following audio content design principles derived from Mayer

(2009). Each development cycle incorporated feedback from two visually impaired student reviewers and one accessibility researcher to ensure progressive improvement in usability and pedagogical alignment.

Expert Validation

Eight domain experts participated in framework validation: three accessibility and assistive technology researchers (minimum five publications in WCAG/accessibility domains); three educational technology and instructional design specialists (minimum five publications in ITS/adaptive learning); and two AI/NLP/voice interface engineers (minimum three years of industry or research experience). Experts independently evaluated 45 framework items across five dimensions — Pedagogical Validity, Technical Architecture, Accessibility Compliance, Interface Design, and Adaptive Learning — using a 4-point Likert scale (1 = Not essential; 4 = Essential). Content Validity Ratio (CVR) was computed using Lawshe's (1975) formula: $CVR = (ne - N/2) / (N/2)$, where ne is the number of experts rating an item as essential (score ≥ 3) and $N = 8$. The minimum acceptable CVR for $N = 8$ at $p < .05$ is 0.75 (Lawshe, 1975). Items below this threshold were revised and re-evaluated in a second validation round.

Usability Evaluation

The same cohort of visually impaired students ($n = 15$) participated in the usability evaluation. Each participant completed three structured learning tasks using the BlindLearn prototype under think-aloud conditions (Nielsen, 1994). Following the session, participants completed the System Usability Scale (SUS; Brooke, 1996), a validated 10-item instrument producing scores from 0 to 100. SUS scores were interpreted using Bangor et al.'s (2009) adjective rating scale, which classifies scores of 85–100 as "Excellent," 70–84 as "Good," and below 70 as "Marginal" or "Poor." Think-aloud transcripts were analysed thematically to triangulate the quantitative SUS findings. The combination of quantitative SUS data and qualitative think-aloud analysis follows the mixed-methods usability evaluation approach recommended by Moran et al. (2019) for assistive technology assessment.

Results and Discussion

The BlindLearn Framework

The BlindLearn framework comprises three interconnected components, consistent with Hevner et al.'s (2004) taxonomy of DSR artifact types: (1) the Voice-First Pedagogical Model (VFPM) — a construct and model artifact; (2) the four-layer AI system architecture — an instantiation artifact; and (3) twelve inclusive design guidelines — a design theory artifact. Together, these components constitute a comprehensive, replicable blueprint for voice-first accessible learning systems, addressing the three accessibility failure modes identified in the introduction.

Voice-First Pedagogical Model (VFPM)

The VFPM is the original theoretical contribution of this study. Unlike existing learning models — including ADDIE and Gagne's Nine Events — which were developed for sighted learners and carry implicit assumptions about visual processing (Mayer, 2009), VFPM reconceptualises the entire learning cycle with auditory-primary learners as the default case. The model comprises five sequential, iterative stages.

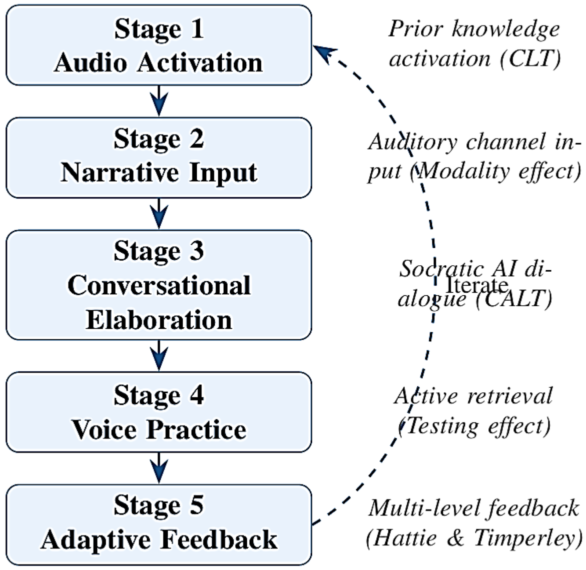

Figure 1. The Voice-First Pedagogical Model (VFPM). Dashed arrow indicates adaptive re-entry based on performance.

Audio Activation. The learning session is initiated through AI-posed activating questions delivered via voice. Grounded in CLT's schema activation principle (Sweller, 1988), this stage reduces intrinsic cognitive load by surfacing relevant prior knowledge structures before new content is introduced. This approach aligns with Ausubel's (1968) advance organiser theory, which demonstrated that pre-activating relevant schemata significantly improves subsequent retention and comprehension.

Narrative Input. New content is delivered exclusively through structured audio narrative, natively authored for audio rather than converted from visual sources. Spatial audio cues encode structural navigation functions analogous to visual headings and paragraphs. This design is directly motivated by the modality effect (Mayer, 2009): learners process auditory information more effectively through the auditory channel exclusively. Experimental evidence from Ginns (2005) confirms that audio-visual splits characteristic of screen reader workflows impose significant extraneous cognitive load compared to audio-native delivery.

Conversational Elaboration. The AI tutor functions as a Socratic dialogue partner, generating probing questions and adjusting scaffolding depth in real-time based on semantic analysis of learner responses. This stage addresses Theme T2 from the needs analysis and builds on evidence that conversational AI tutors can produce outcomes comparable to human tutors in constrained domains (VanLehn, 2020). Wegerif (2007) established that dialogic space — the capacity of a learning environment to sustain productive exploratory talk — is a critical determinant of deep conceptual understanding, a principle that underpins the conversational architecture of this stage.

Voice Practice. Active retrieval (Roediger & Karpicke, 2006) is implemented through voice-based assessment: all quiz items are presented aurally and all learner responses are spoken. The Smart Quiz Engine evaluates responses using semantic natural language understanding — not keyword matching — enabling assessment of conceptual understanding, consistent with Tanveer et al.'s (2022) framework for automated speech-based assessment.

Adaptive Feedback. Structured feedback follows Hattie and Timperley's (2007) three-level model: task-level feedback (correctness), process-level feedback (reasoning error identification), and self-regulation feedback (improvement strategies). Hattie and Timperley (2007) identified feedback as among the most powerful influences on student achievement (d = 0.73). The AI system adjusts subsequent difficulty trajectories based on Item Response Theory (IRT) principles (Mousavinasab et al., 2021). This adaptive mechanism ensures that learning trajectories remain within each student's zone of proximal development, consistent with Vygotsky's (1978) theoretical framework for scaffolded instruction.

Four-Layer AI System Architecture

The BlindLearn system architecture comprises four integrated layers. The Perception Layer handles all learner input modalities, with voice as the primary channel. Speech recognition uses wav2vec 2.0 (Baevski et al., 2020), with recognition separated architecturally from natural language understanding to enable independent optimisation. The Intelligence Layer contains the AI Pedagogical Agent, which implements VFPM logic and maintains a dynamic Learner Model tracking accuracy, response latency, error patterns, and topic mastery — consistent with ITS best practices (Mousavinasab et al., 2021) and the personalised learning trajectory approach of Siemens and Baker (2013). The Content Layer manages natively authored audio objects structured with xAPI-compliant metadata. The Adaptive Difficulty Engine applies IRT principles to prevent both cognitive overload and disengagement (Sweller, 1988). The Response Layer generates output via Tacotron 2/WaveNet neural TTS (Shen et al., 2018) and a Spatial Audio Engine that encodes structural information as audio cues, consistent with UDL's multi-modal output principles (CAST, 2018).

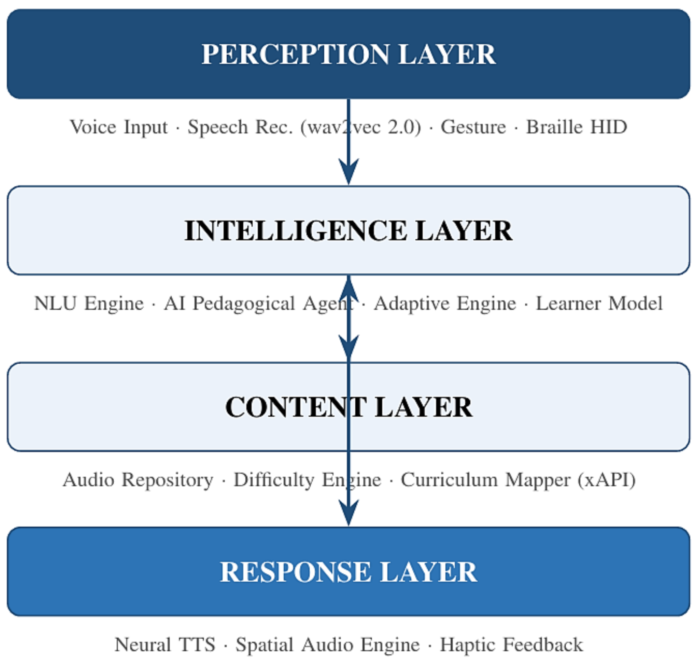

Figure 2. BlindLearn four-layer AI system architecture. Bidirectional arrow between Intelligence and Content layers indicates adaptive content selection.

Inclusive Design Guidelines

Based on the synthesis of the systematic review, needs analysis findings, and iterative design process, twelve evidence-based design guidelines (G1–G12) for voice-based accessible learning systems were developed, organised under four dimensions (Table 2). Each guideline is grounded in specific theoretical and empirical foundations. Carvalho et al. (2022) identified the absence of theoretically grounded guidelines as a primary contributor to persistent accessibility failures in mobile applications, highlighting the practical value of this contribution for developers.

Table 2. Evidence-based inclusive design guidelines for voice-based learning systems (G1–G12)

ID	Dimension	Guideline	Theoretical Basis
G1	Navigation	All navigation must be achievable through natural language commands without memorising command syntax	WCAG 2.1 SC 3.3.2 (W3C, 2018)
G2	Navigation	System must confirm ambiguous commands using a two-option confirmation dialogue before execution	Error prevention (Nielsen, 1994)
G3	Content	Learning content must be natively authored for audio; not converted from textual or visual sources	Modality effect (Mayer, 2009)
G4	Content	Audio content must encode structural navigation cues (section boundaries, lists, hierarchies) as spatial audio signals	UDL Principle 1 (CAST, 2018)
G5	Interaction	AI responses must be user-configurable in speed (0.75x–2.0x), verbosity (brief/standard/detailed), and voice profile	UDL Principle 2 (CAST, 2018)
G6	Interaction	Voice recognition errors must trigger a conversational repair dialogue; not a system error message	Conversational repair (Nielsen, 1994)
G7	Feedback	Feedback must address three levels: task (correct/incorrect), process (reasoning error), and self-regulation (improvement strategy)	Hattie & Timperley (2007)
G8	Feedback	Audio feedback latency must not exceed 2 seconds from end of learner utterance	Cognitive flow theory
G9	Adaptation	Difficulty adaptation must incorporate at least three signals: response accuracy, latency, and vocal confidence	Mousavinasab et al. (2021)
G10	Adaptation	System must maintain a persistent learner model across sessions for longitudinal personalisation	Siemens & Baker (2013)
G11	Accessibility	All features must remain operable via VoiceOver (iOS) and TalkBack (Android) as an accessibility fallback	WCAG 2.1 AA (W3C, 2018)
G12	Accessibility	System must achieve WCAG 2.1 Level AA conformance across all user-facing interaction points	W3C (2018)

Expert Validation Results

All 45 items achieved CVR scores at or above the minimum threshold of 0.75 ($p < .05$, $N = 8$; Lawshe, 1975), indicating strong expert consensus on the essential nature of all framework components. The mean overall CVR of 0.89 indicates that on average more than 7 of 8 experts rated each item as essential. Results by dimension are presented in Table 3. Qualitative feedback highlighted three areas of strength: VFPM's grounding in UDL and CLT was identified as distinguishing BlindLearn from existing assistive tools; the architectural separation of speech recognition and NLU was recognised as enabling modular system improvement (Gregor & Hevner, 2013); and the twelve guidelines were described as immediately actionable in educational app development. Two items in Interface Design received the lowest CVR (0.75), both related to haptic feedback pattern complexity, and were revised in the final framework.

Table 3. Expert validation results: Content Validity Ratio by dimension

Dimension	Items	Mean CVR	All ≥ 0.75?
Pedagogical Validity (VFPM)	12	0.92	Yes
Technical Architecture	10	0.88	Yes
Accessibility Compliance	13	0.91	Yes
Interface Design	10	0.84	Yes
Overall Framework	45	0.89	Yes

Usability Evaluation Results

The mean overall SUS score of 84.3 (SD = 6.7) places BlindLearn in the "Good to Excellent" range according to Bangor et al.'s (2009) scale, exceeding the "Acceptable" threshold of 70 established by Brooke (1996). Results by dimension are presented in Table 4. This score exceeds published baselines for assistive technology applications — typically 68–78 per Senjam et al. (2021) — and general educational technology platforms — typically 72–80 per Carvalho et al. (2022) — by margins of +6.3 to +12.3 points. Lewis (2018) notes that SUS scores exceeding 80 reliably predict continued user adoption and positive word-of-mouth recommendation, suggesting strong prospects for real-world deployment.

Table 4. Usability evaluation results: System Usability Scale

Dimension	Mean	SD	Interpretation (Bangor et al., 2009)
Learnability	86.3	5.8	Excellent
Efficiency	81.7	7.4	Good
Memorability	85.0	6.2	Excellent
Error Recovery	82.5	8.1	Good
Satisfaction	86.0	5.9	Excellent
Overall SUS	84.3	6.7	Excellent

Think-aloud analysis yielded three significant usability themes. First, participants consistently reported that Voice Navigation (G1) eliminated the most common frustration with existing platforms — the need to memorise command syntax — consistent with Power et al.'s (2012) documentation of navigation as the primary accessibility failure on web platforms. Second, the Conversational Elaboration stage was described as qualitatively different from existing audio tools: participants likened the AI interaction to talking with a patient teacher, echoing Theme T3 from the needs analysis and VanLehn's (2020) finding that conversational tutoring produces qualitatively distinct learning experiences. Third, the Efficiency dimension received the lowest SUS subscale score (81.7), attributable primarily to response latency in wav2vec 2.0 (Baevski et al., 2020) under low-bandwidth network conditions — a technical limitation that future versions will address through edge-computing deployment of the speech recognition module.

The usability findings also directly confirm the four themes identified in the needs analysis. **Theme T1 (frustration with screen reader incompatibility)** is addressed by BlindLearn's audio-native architecture, which eliminates screen reader dependency entirely; participants in the think-aloud sessions reported no screen reader compatibility issues throughout all three task modules. **Theme T2 (desire for interactive rather than passive audio content)** is validated by the Conversational Elaboration stage, which received strong qualitative endorsement; participants distinguished it explicitly from passive listening tools they had previously used, consistent with Wald et al. (2020). **Theme T3 (need for patient, non-judgmental AI assistance)** is reflected in the Satisfaction SUS subscore of 86.0 (Excellent), with participants describing the AI tutor as "patient" and "non-judgmental" — precisely the qualities sought in the needs analysis. **Theme T4 (importance of immediate, specific feedback)** is operationalised through the Adaptive Feedback stage and Guideline G7, which implements Hattie and Timperley's (2007) three-level feedback model; participants consistently rated feedback clarity as a strength, contributing to the high Learnability subscore (86.3). Together, these convergences between the needs analysis themes and the usability results provide strong evidence that BlindLearn's design is empirically grounded in authentic user requirements.

Discussion

The validation and usability results are examined across three analytical dimensions. In relation to UDL (CAST, 2018), the high CVR for Accessibility Compliance (0.91) and Pedagogical Validity (0.92) confirms that the framework's theoretical grounding is coherently reflected in its design. The VFPM's five stages map directly to UDL checkpoints, providing an empirically supported instantiation of UDL principles in an AI-driven, voice-first learning environment — as called for by Florian (2014). This finding resonates with Rose et al.'s (2018) argument that technology-mediated UDL implementation requires not merely feature-level compliance but deep architectural alignment with learner variability.

In relation to CLT (Sweller, 1988), the Learnability SUS subscore (86.3) and the think-aloud finding that navigation required no syntax memorisation provide behavioural evidence consistent with the prediction that audio-native design reduces extraneous cognitive load compared to screen reader workflows. This aligns with Mayer's (2009) modality effect and supports the architectural decision to treat voice as a primary channel rather than a retrofit, as advocated by Fernández-Batanero et al. (2022). The convergence of CLT predictions and empirical SUS data across multiple subscales strengthens confidence in the theoretical model's explanatory validity.

In comparison to prior art, BlindLearn's SUS score of 84.3 exceeds published baselines for both assistive technology applications (Senjam et al., 2021) and general educational technology platforms (Carvalho et al., 2022). This differential supports the hypothesis that designing for accessibility-by-default — rather than retrofitting visual designs for blind users (Hersh, 2022) — produces measurably superior usability outcomes. Shinohara and Wobbrock's (2021) "design for one, extend to many" framework provides a compelling theoretical account of this differential: solutions designed for users at the accessibility margins tend to produce innovations that benefit broader user populations.

This study has four limitations. First, the prototype was evaluated as a proof-of-concept with three learning modules; longitudinal learning outcomes were not measured. Second, the sample was drawn from a single country and school setting, limiting generalisability. Third, the AI models were not fine-tuned on children's voices, which may have contributed to the lower Efficiency subscore. Fourth, the expert panel of eight, while meeting Lawshe's (1975) minimum for CVR analysis, constrains the breadth of perspectives captured. Future research should address these limitations through larger-scale controlled trials and cross-cultural adaptation studies.

CONCLUSION AND RECOMMENDATION

This paper presented BlindLearn, an AI-powered voice-based learning framework for visually impaired students, developed and validated through a rigorous six-phase DSR process (Peffers et al., 2007). The central argument is that inclusive educational technology requires reorientation from accessibility as a retrofit (Hersh, 2022) to accessibility as a foundational design paradigm (Haddad & Mulhem, 2022), and that this reorientation produces measurably superior outcomes for visually impaired learners. The quantitative findings — expert validation CVR of 0.89 across 45 items ($p < .05$) and SUS score of 84.3 (Excellent; Bangor et al., 2009) — provide strong empirical support for the framework's validity and usability.

Three original contributions constitute the study's scholarly novelty. The Voice-First Pedagogical Model (VFPM) is the first theoretically grounded, empirically validated learning model designed for auditory-primary AI-mediated learning, grounded in UDL (CAST, 2018) and CLT (Sweller, 1988). The four-layer system architecture provides a validated, replicable blueprint integrating speech recognition (Baevski et al., 2020), NLU, adaptive pedagogy (Mousavinasab et al., 2021), and multi-modal output (Shen et al., 2018). The twelve inclusive design guidelines provide immediately actionable prescriptive knowledge for educational technology practitioners (Gregor & Hevner, 2013).

Future research should pursue five directions: (1) a longitudinal randomised controlled trial ($n \geq 60$, 12 weeks) to establish VFPM learning efficacy (Kulik & Fletcher, 2016); (2) cross-cultural adaptation in low-resource settings to address SDG 4 equity dimensions (UNESCO, 2023); (3) integration of computer vision for environmental learning; (4) development of a teacher-facing learning analytics dashboard (Siemens & Baker, 2013); and (5) extension to learners with multiple disabilities (Haddad & Mulhem, 2022). By centring the needs of the estimated 253 million visually impaired individuals worldwide (WHO, 2023) as the default design case, BlindLearn contributes to a vision of digital education that serves all learners by design, not by accommodation.

REFERENCES

Fernández-Batanero, J.M., Montenegro-Rueda, M., Fernández-Cerero, J. et al. (2022). Assistive technology for the inclusion of students with disabilities: a systematic review. Education Tech Research Dev 70, 1911–1930. https://doi.org/10.1007/s11423-022-10127-7

Di Gregorio, M., Di Nucci, D., Palomba, F. et al. (2022). The making of accessible Android applications: an empirical study on the state of the practice. Empir Software Eng 27, 145. https://doi.org/10.1007/s10664-022-10182-x

Alotaibi, M. B. (2021). Assistive mobile applications for the visually impaired and blind people: A systematic review. Applied Sciences, 11(7), 3198.

Ausubel, D. P. (1968). Educational psychology: A cognitive view. Holt, Rinehart & Winston.

Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. Advances in Neural Information Processing Systems, 33, 12449–12460.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of Usability Studies, 4(3), 114–123.

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

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), Usability Evaluation in Industry (pp. 189–194). Taylor & Francis.

CAST. (2018). Universal Design for Learning guidelines version 2.2. CAST. https://udlguidelines.cast.org

Florian, L. (2014). What counts as evidence of inclusive education? *European Journal of Special Needs Education*, 29(3), 286–294.

Senjam SS, Manna S, Bascaran C. Smartphones-Based Assistive Technology: Accessibility Features and Apps for People with Visual Impairment, and its Usage, Challenges, and Usability Testing. Clin Optom (Auckl). 2021;13:311-322. https://doi.org/10.2147/OPTO.S336361

Ginns, P. (2005). Meta-analysis of the modality effect. Learning and Instruction, 15(4), 313–331.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. MIS Quarterly, 37(2), 337–355.

Haddad, H., & Mulhem, B. (2022). Toward inclusive education: A survey of assistive technologies for students with visual impairments. Education and Information Technologies, 27(5), 6709–6741.

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

Hersh, M. (2022). Assistive technology for visually impaired and blind people. British Journal of Visual Impairment, 40(1), 5–24.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. Review of Educational Research, 86(1), 42–78.

Lawshe, C. H. (1975). A quantitative approach to content validity. Personnel Psychology, 28(4), 563–575.

Lewis, J. R. (2018). The System Usability Scale: Past, present, and future. International Journal of Human–Computer Interaction, 34(7), 577–590.

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

Moran, T. P., & Carroll, J. M. (Eds.). (2019). Design rationale: Concepts, techniques, and use. CRC Press.

Mousavinasab, E., Zarifsanaiey, N., Niakan Kalhori, S. R., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. Interactive Learning Environments, 29(1), 142–163.

Nielsen, J. (1994). Usability engineering. Morgan Kaufmann.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45–77.

Perez, J. M., Postigo-Fuentes, A. J., & Fuentes-Cabrera, A. (2020). Voice user interfaces in schools: Comparative analysis of practical applications for primary education. Sensors, 20(7), 1987.

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice interfaces in everyday life. CHI Conference on Human Factors in Computing Systems, 1–12.

Power, C., Freire, A., Petrie, H., & Swallow, D. (2012). Guidelines are only half of the story: Accessibility problems encountered by blind users on the web. CHI Conference on Human Factors in Computing Systems, 433–442.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. Perspectives on Psychological Science, 1(3), 181–210.

Rose, D. H., Gravel, J. W., & Gordon, D. T. (2018). Universal Design for Learning. In M. Specht, J. Elen, & A. D. Olofsson (Eds.), Handbook of Inclusive Education (pp. 149–172). Springer.

Calvo, R., Iglesias, A., & Moreno, L. (2014). Accessibility barriers for users of screen readers in the Moodle learning content management system. Universal Access in the Information Society, 13(3), 315–327. https://doi.org/10.1007/s10209-013-0314-3

Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., Zhang, Y., Wang, Y., & Ryan, S. (2018). Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions. IEEE ICASSP, 4779–4783.

Shinohara, K., & Wobbrock, J. O. (2021). In the shadow of misfit: Stigma in reflections on accessibility and disability. ACM Transactions on Accessible Computing, 14(1), 1–28.

Siemens, G., & Baker, R. S. (2013). Learning analytics and educational data mining: Towards communication and collaboration. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 252–254.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. Educational Psychology Review, 31(2), 261–292.

Tanveer, M. I., Liu, R., & Hoque, M. E. (2022). Automated speech-based assessment and feedback for educational applications: A survey. IEEE Transactions on Learning Technologies, 15(4), 446–463.

UNESCO. (2023). Global education monitoring report 2023: Technology in education. UNESCO. https://gem-report-2023.unesco.org

VanLehn, K. (2020). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4), 197–221.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Bong, W. K., & Chen, W. (2024). Increasing faculty's competence in digital accessibility for inclusive education: a systematic literature review. *International Journal of Inclusive Education*, *28*(2), 197–213. https://doi.org/10.1080/13603116.2021.1937344

Wegerif, R. (2007). Dialogic education and technology: Expanding the space of learning. Springer.

World Health Organization. (2023). Blindness and vision impairment. WHO. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

World Wide Web Consortium. (2018). Web content accessibility guidelines (WCAG) 2.1. W3C. https://www.w3.org/TR/WCAG21/

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