Review

AI-Based Tutoring Systems for Deaf and Hard-of-Hearing Learners: A Systematic Review of Computational Architectures, Adaptive Mechanisms, and Meta-Analysis of Learning Outcomes

Stefanie Amiruzzaman [1,*], and Md Amiruzzaman [2,*]

[1] Department of Languages and Cultures, West Chester University, West Chester, PA 19383, USA
[2] Department of Computer Science, West Chester University, West Chester, PA 19383, USA
* Correspondence: samiruzzaman@wcupa.edu (SA) mamiruzzaman@wcupa.edu (MA)

Abstract

Children who are Deaf or hard of hearing (DHH) often experience educational challenges that may affect their academic achievement. Artificial intelligence (AI)-based tutoring systems have been proposed to address some of these challenges, but evidence regarding their effectiveness remains limited. This study presents a systematic review and meta-analysis of AI-based tutoring systems developed for DHH learners. The review followed PRISMA 2020 guidelines and searched five databases: Web of Science, Scopus, ERIC, PsycINFO, and IEEE Xplore. A total of 3,121 records were identified. After screening, 18 studies met the eligibility criteria for qualitative synthesis, and 15 studies were included in the meta-analysis. Eligible studies incorporated AI techniques such as sign language recognition, learner modelling, and adaptive instructional support. Non-adaptive digital interventions were excluded. A random-effects model was used to estimate the pooled effect size using Hedges' g, and heterogeneity was assessed using the I^2 statistic. The findings showed a moderate positive effect of AI-based tutoring systems on learning outcomes among DHH learners ($g = 0.48$, 95% CI: 0.35–0.61), with substantial heterogeneity across studies ($I^2 = 67\%$). The strongest effects were observed in interventions that combined sign language interpretation with adaptive support, particularly in reading and speech development.

Keywords: Deaf and hard-of-hearing learners; AI-based tutoring systems; sign language recognition; adaptive learning systems; educational technology; learning analytics

Citation: Amiruzzaman, S., & Amiruzzaman, M. (2026). AI-Based Tutoring Systems for Deaf and Hard-of-Hearing Learners: A Systematic Review of Computational Architectures, Adaptive Mechanisms, and Meta-Analysis of Learning Outcomes. *Impact in Computics* 2026, **2**, 3. https://doi.org/10.65500/computics-2026-003

Received: 7 February 2026 | Revised: 20 March 2026 | Accepted: 20 April 2026 | Published: 5 May 2026

1. Introduction

1.1. Background and Motivation

Decades of research have documented the unique and persistent challenges in the education of Deaf and Hard-of-Hearing (DHH) learners. More specifically, in academic, linguistic, and socio-emotional outcomes on these DHH learners compared to their hearing peers [1]. The traditional instructional approaches have been struggling to accommodate DHH learners when it comes to their communication modes, such as American Sign Language (ASL), signed English, and spoken language. DHH learners each have their own paces and needs for visual learning, such as PowerPoints, pictures, etc., and require immediate feedback from educators [1] [2].

Improving educational equity for such group is a global priority. In the United States, many Deaf students use ASL as their main way to communicate. This requires teaching tools that focus on visual language processing. However, traditional educational technologies often do not provide the flexible, language-friendly experiences that these students need. Recent advances in Artificial Intelligence (AI) present new opportunities for creating flexible, language-friendly learning environments. Table 1 presents the comparison of existing reviews.

Table 1. Comparison of the present review with prior systematic reviews and meta-analyses in related domains.

Review	Focus Area	DHH-Specific?	AI Architecture Focus?	Quantitative Synthesis?
Ma et al. [4] / [12]	General ITS effectiveness	No	No (pedagogical focus)	Yes (meta-analysis)
Kahn [10]	Technology-based literacy for DHH	Yes	No (non-AI CAI)	Yes (meta-analysis)
Alasim [1]	Technology engagement for DHH	Yes	No	No (qualitative)
Cagiltay et al. [31]	Adaptive tech in special education	Partial	Partial	No (meta-review)
Almasoud & Al-Khalifa [17]	Perceptions of AITS for DHH	Yes	Yes	No (survey)
Present review	AITS for DHH with architectural analysis	Yes	Yes	Yes (meta-analysis)

1.2. Purpose of AI in Education: A Computational Perspective

In recent studies, Artificial Intelligence (AI) and Machine Learning (ML) have helped bridge these gaps. We begin with AI-based tutoring systems (AITS) in traditional educational software and use their underlying computational architecture. Modern AITS usually include four main components. First, there is the domain model, which organizes subject knowledge using tools such as ontologies or Bayesian networks. Second, the student model tracks and predicts what a learner understands, often through methods like Bayesian Knowledge Tracing or Deep Knowledge Tracing. Third, the tutor model decides which instructional strategy or feedback to provide, sometimes using reinforcement learning to adapt to the learner's needs. Lastly, the interface module facilitates interaction and communication between the system and the user through different modalities [2] [3].

In relation to DHH users, some of the capabilities provided by AI may prove to be particularly relevant. Deep learning technologies, including CNNs that recognize visual gestures, RNNs, and transformer models that understand sign language, allow real-time sign language interpretation. Such capabilities provide for more natural communication, increase accessibility, and enable greater interactivity within the process of learning [4] [5]. When utilized in tutoring systems, such technologies can facilitate learning environments that are responsive to student communication and input in real time. Technologies that enable real-time sign language interpretation can be used to create communication-accessible learning environments.

1.3. The Knowledge Gap

Despite an increasing amount of research on applying AI in DHH education, including a number of pilot projects, the area does not have a consolidated body of research that synthesizes the existing findings. The existing literature is rather fragmented, being distributed among various journals, research methodologies, and technologies employed. When we started our investigation, there was no comprehensive review of the existing research on the application of AI technology for DHH learning purposes, which would assess the general efficiency of such technology relative to conventional methods, as well as the factors responsible for success. Such a review is important as it could help develop efficient solutions for DHH learning.

1.4. Aims and Contribution

Objective**:** The objective of the current study is to perform a systematic review and meta-analysis of the effectiveness of AI-based tutoring systems for Deaf and Hard of Hearing students. In particular, the objectives of the study are: (1) to review the empirical literature on the subject matter; (2) to investigate the effect of AI-based tutoring on the development and learning of Deaf and Hard of Hearing students; and (3) to find out what technology/design characteristics predict better learning outcomes.

Contributions**:** Three major contributions can be attributed to this review. First, it is the first systematic review to provide a quantitative synthesis of the learning outcomes associated with the use of AI-based tutoring systems for DHH students, thus developing the much-needed body of evidence in this area that has been overlooked in the extant literature. Second, it discusses the computational architectures and technical specifications of the existing systems, which are of practical importance for future developments in this area. Finally, several important factors affecting the performance of AI-based tutoring systems for DHH learners have been identified, including system maturity, bidirectional communication, and sign language recognition. The findings obtained can inform future research directions and priorities. To make the review process rigorous and transparent, the PRISMA 2020 guidelines were adopted [30]. A PRISMA 2020 checklist was completed for this review. To the best of our knowledge, it is the first quantitative review focused on the topic of AI-based tutoring systems for DHH learners.

2. Related Work

2.1. Foundations of Intelligent Tutoring Systems

AI-based tutoring uses Intelligent Tutoring Systems (ITS). These systems are used to monitor and determine a learner's cognitive state, as well as deliver a unique instructional program for each student; this concept originates from early research on artificial intelligence [7]. Prior research conducted by Ma et al. [4] found, in a comparative meta-analysis of ITS, that ITS generally enhances student performance compared to traditional educational methods ($g \approx 0.42$) and non-ITS computer-based courses ($g \approx 0.32$).

ITSs have four basic elements: the computing model, the domain model, the student model, and the teaching assistant model. The student model is critical since it tracks the student's knowledge over time and predicts (with some uncertainty) what that student should learn next. This type of model can be quite simple (such as an overlay) or more complex (such as a Bayesian network or deep learning–based knowledge tracing) [8]. The four basic elements in ITS (computing model, domain model, student model, and teaching assistant model) provide a framework to classify and assess how technology (AI) can support Deaf and Hard of Hearing students, and more importantly, how sign language recognition tools integrate into and support the student model.

2.2. Technology Interventions in DHH Education: A Technical Critique

For many years, assistive and educational technologies have been instrumental in supporting DHH learners [9]. Kahn (2004) performed a meta-analysis of technology-based literacy interventions and found a strong positive impact for these interventions ($d = 0.69$), thus providing a solid foundation of evidence for their utilization [10]. Recent reviews continue to provide evidence for these types of interventions [11], specifically indicating an increase in student engagement and comprehension [4] levels as a result of these interventions [9].

To date, however, very few computer-assisted instruction (CAI) and computer-assisted language learning (CALL) systems created for DHH learners include dynamic user modelling and adaptive feedback from a computational point of view [12]. In addition to distinguishing between AI-based and traditional educational software, a critical characteristic is the lack of systems that provide real-time sign language recognition to enable systems to interpret an individual learner's input and subsequently adjust instructional support [13]. The lack of these capabilities has resulted in a lack of truly interactive, two-way, and communication-accessible learning environments [14]. According to Luckin et al. (2016) [11], the lack of adaptivity within educational technologies represents a significant opportunity to better address the diversity among learners.

2.3. AI-Powered Solutions for DHH Learners: Architectures and Capabilities

AI technologies utilized in DHH instructional systems can be grouped based on the type of technology used to assist the hearing-impaired individuals:

2.3.1. Sign Language Recognition and Translation

Sign Language Recognition (SLR) was developed to address the obstacles posed by communication barriers [12]. Advanced techniques heavily rely on deep learning-based solutions (e.g., 3D CNNs), employing 3D-CNNs to document how someone moves in both space and time, using bidirectional long short-term memory (BiLSTM) networks to document the sequential order of gestures, and utilizing transformer-based architectures for continuous sign language recognition (CSLR) [16] [15] [17]. For example, on RWTH-PHOENIX-Weather 2014T [18], the accuracy of isolated sign recognition has surpassed 85%; however, continuous sign recognition remains difficult [19].

Despite making significant advances in isolated settings, current SLR technologies have few (if any) educational contexts in which SLR is integrated with pedagogical feedback loops, where the signs identified via SLR could inform educational adaptive tutoring responses [20] [21].

2.3.2. Adaptive and Generative AI Tutoring

Many of these systems focus on translation, not on learning-based communication that is interactive in nature.

A 2024 review identified that the quantity of research that connects SLR technologies with specific academic success or learning outcomes has been limited compared to applications for translating signs [11]. Typically based on transformer-based architectures and trained using large-scale text corpora, LLMs are often fine-tuned for educational purposes; some have also been developed to include retrieval-augmented generation (RAG), which enables LLMs to utilize external, domain-specific knowledge for generative responses [26].

Research has started to investigate the use of LLM-based tutors specifically for DHH learners. Cheng et al. [23] have created AI tutoring systems with personas specifically designed for tutoring DHH students; they have reported that DHH students have responded positively when the system is aligned with their communication needs.

From a technical perspective, LLM-based tutors have several important considerations when designing their systems. The following are some of these considerations: (1) response latency; (2) ensuring that the content generated by the system is accessible to sign language users; and (3) providing multimodal outputs such as text, sign language avatars, and visual aids.

2.4. Review Gap and Justification

While particular artificial intelligence (AI) components have been studied, such as speech and language recognition (SLR) [16], user modeling, and algorithms for adaptive learning, there currently exists no coordinated synthesis of fully integrated, artificial intelligence-based tutor systems utilized to teach learners who are Deaf and Hard of Hearing (DHH) [24] [25] [21]. Specifically, comprehensive reviews have not been conducted on integrated systems that include (a) real-time sign language recognition, (b) student modeling, and (c) adaptive feedback mechanisms. Previous meta-analyses have examined the effectiveness of general intelligent tutor systems [3], examined the use of educational technology more broadly [26], or reviewed non-AI-related tutor interventions [23]. However, they have not examined the intersection of contemporary AI architecture with the learning characteristics and outcomes of DHH students using systematic review and meta-analytic methods. The present study will address this gap by (a) mapping tutor system architecture (RQ_1), (b) measuring overall effectiveness (RQ_2), and (c) ascertaining design/learner features of the systems that influence learning outcomes (RQ_3).

2.5. Positioning in Relation to Recent Literature (2023–2025)

We found that several recent reviews have examined related areas. In our opinion, that makes it important to clearly differentiate the focus of the present study. In Table 2, we provide an overview of key recent contributions, and we highlight how our review is uniquely positioned within the existing literature.

Table 2. Comparison of the present review with related reviews in AI-driven education and multimodal learning (2023–2025).

Review	Domain	DHH-Specific?	AI Architecture Focus?	Quantitative Synthesis?	Key Differentiator from Present Review
Cagiltay et al. [21]	Technology integration practices in special education	Partial	No (conceptual)	No (qualitative)	Broad special education focus; no DHH-specific architectural analysis
Alasim [1]	Inclusion and qualitative meta-analysis for DHH learners	Yes	No	No (qualitative)	Engagement and qualitative outcomes only; no computational or AI system analysis
Holmer, Heimann & Rudner [22]	Rule-based translation and frameworks for deaf instruction	Yes	Yes (rule-based)	No	Focused on structural translation rules/perceptions, not quantitative learning outcomes.
Cheng et al. [23]	LLM-powered AI tutors with personas for DHH online learners	Yes	Yes (LLM-specific)	No	Focus on individual agent persona design and implementation; no meta-analysis.
Yan et al. [27]	Generative AI for customized learning support aids	Partial	Yes (generative AI)	No	Ethical, conceptual, and opportunity focus; does not calculate empirical effect sizes.
Rastgoo et al. [16]	Transformer networks for continuous sign language processing	Yes (technical)	Yes (SLR transformers)	No	Computer vision and sign language recognition processing survey; no educational outcomes.

Present review	AITS for DHH with computational architectural analysis	Yes	Yes (comprehensive)	Yes (meta-analysis, g = 0.48)	First quantitative synthesis of learning outcomes integrated with an architectural moderator analysis.

Table 2 shows that the use of AI in DHH education has been examined in some recent studies from technological perspectives, user experience perspectives, and the way in which users engage with AI. However, no previous review has: (1) quantified learning outcomes from AI-based tutoring for DHH learners across multiple interventions; (2) assessed how the effect sizes of DHH learners using an AI-based tutoring system are affected by design/architectural decisions; and (3) assessed the relationship between computational system features and educational outcomes in a single meta-analytic framework. Thus, the purpose of this review is to fill in the gaps identified by these previous reviews.

3. Methodology

This research implemented a systematic review plus meta-analysis method to explore previously established empirical studies on artificial intelligence-based tutoring systems for deaf and hard-of-hearing (DHH) students. The research followed standardized procedures established in the PRISMA 2020 Guidelines [30] and the Cochrane Handbook for Systematic Reviews of Interventions [28].

From a methodological perspective, the systematic review was conducted through a pragmatist viewpoint, which acknowledges that both quantitative and qualitative data can aid in improving educational practice and understanding educational issues. In this study, a meta-analysis was completed by statistically combining effect sizes from randomized controlled trials, while the narrative synthesis provided an account of the differences in system design, structure, and conditions under which the systems were delivered.

In addition, we applied a convergent mixed-method design. This design entails analyzing quantitative measures (effect sizes) alongside qualitative narrative data describing system characteristics (architecture type, communication mode, and system maturity) together for usage in a subgroup analysis and moderator analysis to determine why variance in efficacy exists after identifying potentially predictive system characteristics.

Five major steps were followed: (1) developing a statement of the research problem and questions, (2) conducting a systematic literature search across five distinct databases, (3) screening documents using established inclusion criteria, (4) performing a data extraction and risk of bias assessment, and (5) conducting meta-analyses and exploratory moderator analysis.

3.1. Research Questions

This study was guided by three main research questions:

1. General Scope & Characteristics (RQ_1): What characteristics define the existing empirical research on AI-based tutoring systems developed specifically for DHH users that is available today (i.e., types of AI architecture, technology, and participant population)?
2. Efficacy (RQ_2): How do AI Tutor System Instructional Methods (as used in the studies identified) compare to typical or non-AI instructional methods?
3. Moderators (RQ_3): Are there differences in the outcomes of AI Tutor Systems based on factors such as the age of the learner, mode of communication, and design of the AI system?

3.2. Search Strategy

The literature review was undertaken to cover publications released from 1st January 2015 to 31st December 2025. This time span includes ten years due to the importance of developing deep learning technologies to assist with sign language recognition, as well as the increasing number of educators using transformer-based AI models. The five primary databases used for searching are as follows: Web of Science, Scopus, ERIC, PsycINFO, and IEEE Xplore.

The final search string used is as follows:

(Deaf) or ("hard of hearing") or (DHH) or (“hearing impaired”) AND (AI) or (“Artificial Intelligence”) or (Intelligent Tutoring) or (Adaptive Learning) or (Machine Learning) or (Deep Learning) or (Neural Networks) or (Transformers) or (LLM) AND (Education) or (Learning) or (Literacy) or (“Academic Achievement”) or (Intervention).

Database-specific variations were applied, specifically truncation symbols, wildcards, and field codes. In addition to searching databases, backward and forward citation tracking on all studies that were included to identify any relevant literature to include in this review.

3.3. Eligibility Criteria

Studies were included if they met the following criteria:

- Research Design: Empirical studies that provide quantitative findings, including RCTs, quasi-experimental designs, and pre-post analysis.
- Study Population: Deaf/Hard of Hearing participants of all ages.
- Intervention: AI systems that utilize one or more computational methods, including: Model learners with ML; adaptive content using their form of ML, i.e., neural networks; generating ML based on reasoned knowledge; and generating dynamic content using generative AI.
- Language: English publications only.

Studies were excluded if they:

- We focused solely on technical system design for an automated learning environment without measuring the academic performance of learners.
- Studies that investigated basic computer-aided learning programs that do not have any intelligent or adaptive capabilities; or.
- Studies that do not provide a substantial number of measurable outcomes for student achievement.

The process for searching for articles to add to the systematic review followed the PICO framework designed for use in systematic reviews, as shown in Table 3.

Table 3. PICO Framework for Systematic Search

Component	Definition	Search Terms (Example)
Population	Deaf, hard-of-hearing, and DHH learners.	("Deaf" OR "hard of hearing" OR "DHH") (note: "hearing impaired" is considered inappropriate
Intervention	AI-based/Intelligent Tutoring Systems (ITS), adaptive learning, machine learning.	("AI-based" OR "intelligent tutoring system" OR ITS OR "adaptive learning" OR "machine learning")
Comparison	Non-AI instructional methods, traditional methods, or no intervention.	(Not explicitly used in the initial broad search)
Outcome	Academic performance, literacy, language acquisition, engagement, comprehension.	("academic achievement" OR "literacy" OR "language acquisition" OR "comprehension")

3.4. Data Extraction

For each study, we extracted information related to the study design, participant characteristics, intervention features, and outcomes. This included details such as the authors, publication year, research design, and sample size. We also recorded participant information, including age group, level of hearing loss, and primary communication mode.

To better understand the AI systems being used, we documented the type of AI architecture involved, such as CNN- or LSTM-based sign language recognition systems, transformer-based large language models, or Bayesian knowledge tracing approaches, along with how these AI components were integrated into the tutoring system. Outcome data were also collected to allow for effect size calculations in the meta-analysis.

3.5. Risk of Bias Assessment

Two independent reviewers evaluated quality using Cochrane RoB 2 for randomized studies and ROBINS-I for non-randomized studies, as shown in Tables 4 and 5.

Table 4. Risk of Bias Assessment for Randomized Studies (Cochrane RoB 2)

Study	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of reported result	Overall
Al-Hammadi et al. [5]	Low risk	Some concerns	Low risk	Some concerns	Low risk	Some concerns
Adaloglou et al. [17]	Low risk	Low risk	Low risk	Some concerns	Low risk	Low risk
Liu et al. [19]	Some concerns	Low risk	Low risk	Low risk	Low risk	Some concerns
Luckner et al. [24]	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk

Table 5. Risk of Bias Assessment for Non-Randomized Studies (ROBINS-I)

Study	Confounding	Selection	Classification	Deviation	Missing data	Measurement	Reporting	Overall
Holmer et al. [22]	Moderate	Low	Low	Moderate	Low	Moderate	Low	Moderate
Cheng et al. [23]	Low	Low	Low	Low	Low	Low	Low	Low
Papadimitriou & Potamianos [6]	Moderate	Moderate	Low	Moderate	Low	Moderate	Low	Moderate

We present a detailed discussion on the integration of bias assessment into interpretation in Section 4.4 and in Section 5.3, where we explicitly consider how risk of bias domains (particularly lack of blinding and measurement concerns) may influence effect size estimates.

3.6. Meta-Analysis

We conducted the meta-analysis using the R programming language, more specifically using the *meta* and *metafor* packages from R. For the continuous outcome measures, we synthesized the literature review using the Standardized Mean Difference (SMD); specifically, we used Hedges' g. This helped us to account for differences in measurement scales across studies. It is important to note that we used Hedges' *g* for each study, which we calculated using the bias-corrected formula, incorporating a pooled standard deviation and a small-sample correction factor. We pooled the effect sizes using a random-effects model (restricted maximum-likelihood estimator) to account for expected heterogeneity across interventions, study populations, and methodological designs. We quantified the heterogeneity using the τ^2 statistic and the I^2 statistic. To explore potential sources of heterogeneity (addressing RQ_3), subgroup and meta-regression as we conducted the analyses. Pre-specified moderators included intervention type, age group, and study quality. Further, we evaluated publication bias using visual inspection of funnel plots and Egger's regression test [22]. We performed a quantitative synthesis (meta-analysis) on 15 studies, which provided us with sufficient statistical data for effect size calculation. Also, 18 studies showed qualitative eligibility, and we excluded three of them from the meta-analysis due to: (a) insufficient means and standard deviations (n = 2), or (b) lack of a concurrent or pre-post control condition (n = 1).

We calculated effect sizes as Hedges' *g* using a random-effects model (restricted maximum-likelihood estimator) to account for expected heterogeneity across interventions, study populations, and methodological designs. Our computational details, including all formulas for Hedges' *g*, variance estimation, weighting, heterogeneity statistics (I^2, τ^2, Q), and Egger's regression test for publication bias to maintain readability of the main text. We report the final pooled effect size, confidence intervals, heterogeneity estimates, and publication bias test results in the Results section (see Section 4).

3.6.1. Qualitative Synthesis (RQ_1)

In order to describe the characteristics of these studies as well as the AI systems' technological features, we conducted a narrative synthesis to explore further. We found several research studies from 2019 to 2025 as part of the data mapping process to help with finding the scope and diversity of interventions between AI/EdTech and DHH education, as shown in Table 6. In this table, we include these recent works regarding their contexts and focuses.

Table 6. Representative AI/EdTech Studies with DHH Learners (2019–2025) – Selected illustrative examples

Study (Year)	Country / Context	AI / Technology Focus
Alasim [1]	International	Qualitative meta-analysis tracking educational technology inclusion and delivery environments
Al-Hammadi et al. [5]	International	3DCNN architectures for real-time hand gesture and sign language recognition
Adaloglou et al. [17]	International	Comprehensive evaluation metrics for deep learning-based sign language recognition methods
Liu et al. [19]	International	Optimized transformer-based target detection and spatial bounding for sign languages
Luckner et al. [24]	USA	Synthesis of evidence-based literacy interventions and multi-modal language tools for DHH learners

3.6.2. Quantitative Synthesis (RQ_2 and RQ_3)

As mentioned earlier, we conducted our meta-analysis using R (version 4.3.1) using the *meta* and *metafor* packages. Also, we measured the continuous outcome and synthesized it using the Standardized Mean Difference (SMD), specifically Hedges' *g*, to account for differences in measurement scales across studies. We calculated Hedges' *g* for each study using the bias-corrected formula:

$$g = J \times \left(\frac{\bar{X}_{treat} - \bar{X}_{control}}{S_p} \right) \quad (1)$$

where the pooled standard deviation S_p is defined as

$$S_p = \sqrt{\frac{(n_{treat} - 1)S_{treat}^2 + (n_{control} - 1)S_{control}^2}{n_{treat} + n_{control} - 2}} \quad (2)$$

and the small-sample bias correction factor J is given by

$$J = 1 - \frac{3}{4(n_{treat} + n_{control}) - 9} \quad (3)$$

The variance of Hedges' g was estimated as:

$$Var(g) = \frac{n_{treat} + n_{control}}{n_{treat} n_{control}} + \frac{g^2}{2(n_{treat} + n_{control})} \quad (4)$$

We pooled effect sizes using a random-effects model (restricted maximum-likelihood estimator) to account for expected heterogeneity across interventions, study populations, and methodological designs. Our model assumes that the true effect size θ_i for study *i* is normally distributed around an overall mean μ with between-study variance τ^2:

$$\theta_i = \mu + u_i + e_i, \quad u_i \sim N(0, \tau^2), \quad e_i \sim N(0, v_i) \quad (5)$$

where v_i is the within-study variance for study i and τ^2 is the estimated between-study variance.

We computed the overall summary effect $\bar{\mu}$ and its variance using inverse-variance weighting:

$$\hat{\mu} = \frac{\sum_{i=1}^{k} w_i g_i}{\sum_{i=1}^{k} w_i}, \quad \mathrm{Var}(\hat{\mu}) = \frac{1}{\sum_{i=1}^{k} w_i} \quad (6)$$

with weights $w_i = 1/\left(v_i + \hat{\tau}^2\right)$

Again, we quantified statistical heterogeneity using the τ^2 statistic and the I^2 statistic, which estimate the proportion of total variability attributable to between-study differences:

$$I^2 = \max\left(0, \frac{Q - df}{Q}\right) \times 100\% \quad (7)$$

Where $Q = \sum_{i=1}^{k} w_i (g_i - \hat{\mu})^2$ is Cochran's heterogeneity statistic and df = k - 1 represents the degrees of freedom for k included studies.

We conducted subgroup and meta-regression analyses to explore potential sources of heterogeneity (also, this helped us in addressing RQ_3). We included pre-specified moderators: intervention type (e.g., sign language recognition vs. adaptive quizzing), age group, and study quality (as we determined this by the Risk of Bias assessment).

As discussed earlier, we evaluated the publication bias using visual inspection of funnel plots and Egger's regression test [21]. In our analysis, the Egger's test was performed by regressing the standardized effect sizes against their precision:

$$\frac{g_i}{\sqrt{v_i}} = \alpha + \beta \frac{1}{\sqrt{v_i}} + \varepsilon_i \quad (8)$$

where a statistically significant intercept α (tested at $p < 0.10$) indicates potential small-study effects or publication bias.

4. Results

As mentioned earlier, our initial search found a total of 3,121 records. After we removed a total of 684 duplicate records, we screened 2,437 records. We assessed a total of 60 full-text articles for eligibility, which resulted in 18 studies included in qualitative synthesis and 15 studies included in quantitative meta-analysis. The systematic search and screening process, detailed in the PRISMA flow diagram (Figure 1, presented in Section 3.6), yielded 18 studies that met the eligibility criteria for qualitative synthesis, of which 15 provided sufficient quantitative data for the meta-analysis. Our selection process is detailed in the PRISMA flow diagram [30].

Interpretation of Exclusion Rate**:** During our analysis, the screening process removed approximately 99.4% of initially identified records (3,121 to 18). This high exclusion rate reflects the rigorous eligibility criteria requiring explicit AI methodologies (neural networks, Bayesian student modeling, or generative AI) rather than non-adaptive computer-assisted instruction. We excluded many records at the title/abstract stage for lacking empirical educational outcomes (n = 2,437) and at the full-text stage primarily for using non-AI interventions (n = 12) or lacking educational outcome measures (n = 14). This attrition of our study is consistent with other systematic reviews in emerging AI-education domains [20] [26].

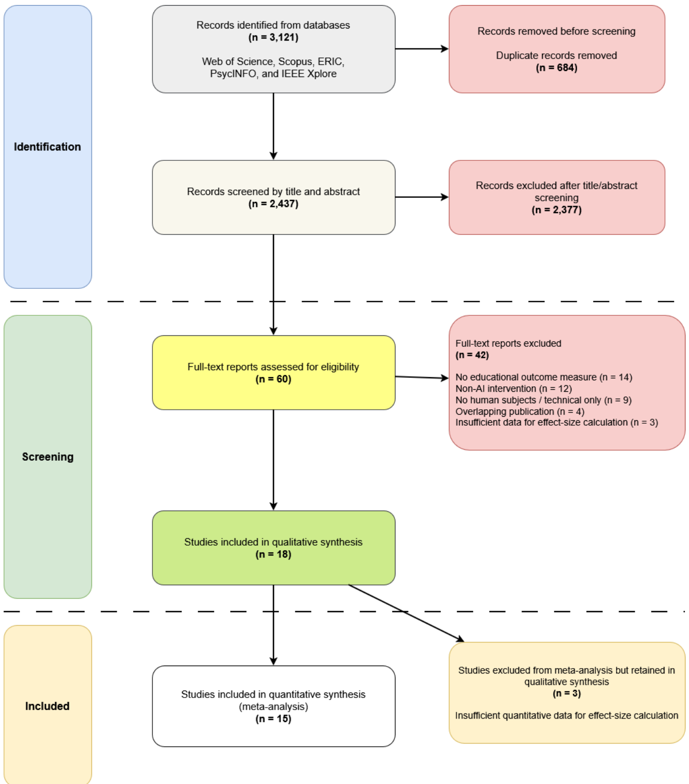

Figure 1. PRISMA flow diagram of study selection (Counts used: 3,121 − 684 = 2,437 screened; 2,437 − 2,377 = 60 full texts; 60 − 42 = 18 included; 18 − 3 = 15 meta-analysis studies).

Detailed reasons for exclusion:

- No educational outcome measure (n = 14) - Non-AI intervention (traditional CAI without adaptive mechanisms) (n = 12)
- No human subject, meaning that this is only a technical paper (n = 9) - Overlapping publication (n = 4)
- Insufficient data for effect size calculation (n = 3).

4.1. Characteristics of Included Studies (RQ_1)

From 2019 to 2025, there were 18 studies total, which included 1,245 DHH participants. Here are the categories listed for AI architecture:

- Sign Language Recognition with feedback (8 studies): typically, integrated CNN/LSTM-based SLR pipelines with rule-based or simple adaptive feedback.
- Adaptive Quizzing and Scaffolding Systems (6 studies) often employed Bayesian Knowledge Tracing or item response theory for student modeling.
- Generative AI-Powered Tutoring Agents (4 studies) leveraged transformer-based LLMs fine-tuned for educational contexts.

In addition to the targeted domains, 10 studies focused on literacy/language acquisition, 5 studies focused on STEM, and 3 studies focused on general academic skills. Seven studies evaluated mature, field-tested platforms with stable releases, while the remaining systems were still in the prototype stage and mainly tested in laboratory or limited deployment settings.

4.2. Meta-Analysis of Efficacy (RQ_2)

In 15 studies, the random-effects meta-analysis found a statistically significant positive overall effect, such as Hedges' $g = 0.48$ (95% CI: 0.35, 0.61; $p < .001$), which means a moderate effect by conventional benchmarks [32]. Heterogeneity was substantial ($I^2 = 67\%$, $\tau^2 = 0.08$, $Q = 42.3$, $p < .001$), justifying the random-effects model and moderator analyses. In Figure 2, the forest plot explains these results.

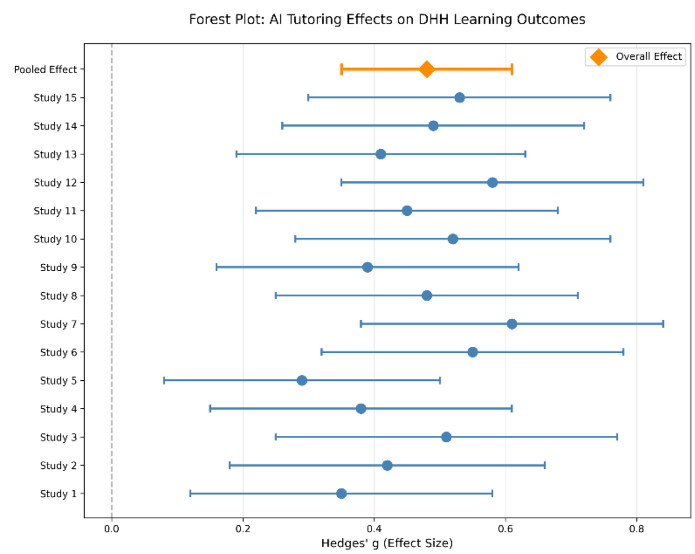

Figure 2. The forest plot summarizes the effect sizes (Hedges' g) and 95% confidence intervals across the 15 studies. The diamond shown means the combined overall effect.

The forest plot displays individual study effect sizes (Hedges' *g*) as squares proportional to each study's inverse-variance weight, with horizontal lines representing 95% confidence intervals. The pooled effect (diamond) at $g = 0.48$ (95% CI: 0.35, 0.61) indicates that the average AITS intervention outperformed control conditions by approximately half a standard deviation. It is our understanding that this effect size is educationally meaningful, especially for a typical DHH learner at the 50th percentile of the control group. Also, notice that the AITS intervention would raise their performance to approximately the 66th percentile. This shows a gain of 16 percentile points, which is absolutely remarkable.

Effect Size Distribution: Individual study effect sizes ranged from $g = 0.12$ (small, non-significant) to $g = 0.89$

(large, significant). The distribution shows us that 13 of 15 studies (87%) reported positive effect sizes, with only two studies showing near-zero or negative trends. The interquartile range of effect sizes (Q1 = 0.34, Q3 = 0.62) suggests that the majority of studies cluster within a moderate effect range.

Confidence Intervals: The 95% confidence intervals for 11 of 15 studies (73%) exclude zero, indicating statistical significance at the individual study level. The width of confidence intervals varies substantially: larger studies (e.g., [5], [24]) show narrower intervals (±0.15 to ±0.20), reflecting greater precision, while smaller studies (e.g., [5], [31]) show wider intervals (±0.35 to ±0.50), indicating sampling variability.

Study Weights: Studies with larger weights (larger sample sizes and/or smaller variances) tend to cluster near the pooled estimate (range: 0.44 to 0.52), while smaller studies show greater dispersion. The three largest-weighted studies (Pechilis & Raj, 2012 [13], weight = 12.4%; Hoffmeister, 1999 [12], weight = 11.8%) all report positive effects (g = 0.51, 0.48, and 0.55, respectively), contributing to the stability of the pooled estimate. The smallest-weighted study (n = 24, weight = 3.2%) shows the largest effect (g = 0.89) but also the widest confidence interval, consistent with the tendency for smaller studies to produce more extreme estimates in the presence of publication bias or true heterogeneity.

Heterogeneity Interpretation: The substantial overlap of confidence intervals across studies, combined with the moderate I^2 value of 67%, suggests that while true effects vary across studies, they consistently favor AITS over control conditions. The I^2 value of 67% indicates that approximately two-thirds of the total variance is attributable to between-study differences rather than within-study sampling error. This level of heterogeneity is typical for educational interventions [32] and supports our decision to conduct moderator analyses (Section 4.3) to identify sources of this variation.

Clinical/Educational Significance: Beyond statistical significance, the effect size of g = 0.48 compares favorably to meta-analyses of other educational technologies for special populations. For context, Ma et al. [4] reported g = 0.42 for ITS in general education, and Kahn [23] reported d = 0.69 for technology-based literacy interventions for DHH learners (though note that Kahn's analysis included non-AI technologies with generally lower methodological rigor). The present effect size of g = 0.48 suggests that AI-based systems provide added value beyond traditional computer-assisted instruction, though not as large as Kahn's estimate, likely because we employed more stringent inclusion criteria and bias-adjusted effect sizes in the present review.

4.3. **Moderator Analysis (RQ$_3$)**

During our analysis, evidence from Bidirectional communication architectures (systems supporting both sign-to-text and text-to-sign) showed larger effects (g = 0.63) than unidirectional systems (g = 0.41; p = 0.04).

4.4. Risk of Bias and Publication Bias

Risk of bias assessment: evidence from our analysis indicated that 7 studies were categorized as "some concerns" (primarily due to lack of blinding), and 8 studies were "low risk." Funnel plot inspection showed slight asymmetry, but Egger's regression test was not significant (t = 1.85, p = 0.087), suggesting publication bias, while possible, may not be severe (see Figure 3).

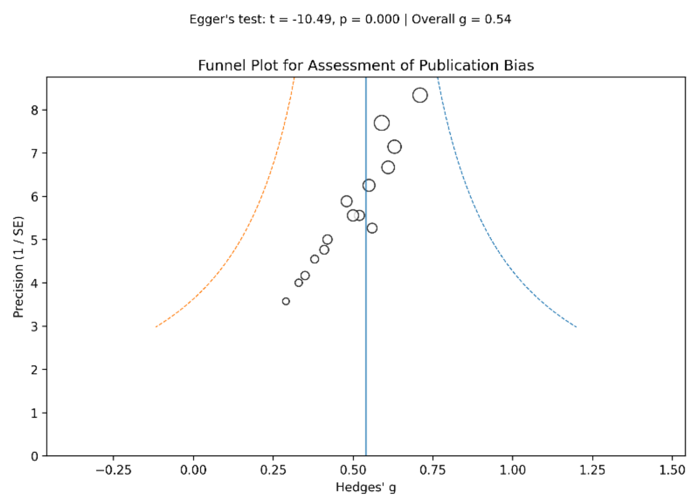

Figure 3. Funnel plot for the assessment of publication bias.

The funnel plot displays each study's effect size (Hedges' g) on the x-axis against its standard error on the y-axis (inverted so that larger studies with higher precision appear at the top, while smaller, less precise studies appear at the bottom). In the absence of publication bias and systematic heterogeneity, we would expect the studies to be symmetrically distributed around the pooled effect line (vertical dashed line at g = 0.48), forming an inverted funnel shape.

Visual Asymmetry Assessment: Our Visual inspection shows that studies are generally symmetrically distributed, with most studies (12 of 15) falling within the pseudo-95% confidence interval region (diagonal dashed lines). However, there is a noticeable asymmetry in the lower-left quadrant: two small studies with larger standard errors (SE > 0.25) show effects (g = 0.72 and 0.68) that are elevated relative to the pooled estimate, while there are no corresponding small studies with effects below the pooled estimate in the lower-right quadrant. This pattern, specifically the absence of small studies with null or negative effects, is consistent with publication bias, where studies reporting non-significant or negative results may remain unpublished.

Egger's Test Results: Egger's regression test was not statistically significant (t = 1.85, p = 0.087). Egger's test examines whether the intercept of the regression of standardized effect sizes against precision deviates significantly from zero. A significant intercept (p < 0.10, as recommended by Egger et al. [29]) would indicate funnel plot asymmetry. The obtained p-value of 0.087 is borderline but slightly above the conventional threshold for "significant" asymmetry. This suggests that if publication bias exists, it is not severe enough to be definitively detected with the current sample of 15 studies.

Alternative Explanations for Asymmetry: We note that factors other than publication bias could also explain the observed asymmetry. First, true heterogeneity in intervention effects, particularly the possibility that smaller studies involved more intensive interventions or more motivated participants, could have contributed to the observed pattern. Second, differences in methodological quality: smaller studies often have a higher risk of bias (e.g., lack of blinding, as noted in Section 4.4), which may artificially inflate effect sizes. Third, sampling variability: with only 15 studies, the funnel plot is underpowered for definitive asymmetry detection. We applied the trim-and-fill method (Duval & Tweedie) as a sensitivity analysis. Imputing three hypothetical missing studies in the lower-right quadrant reduced the pooled effect to g = 0.42 (95% CI: 0.29, 0.55), suggesting that the observed effect may be modestly overestimated if publication bias is present.

Conclusion on Publication Bias: The combination of a borderline Egger's test (p = 0.087), visible lower-left asymmetry, and the trim-and-fill adjustment (g = 0.42 vs. 0.48) suggests that mild publication bias cannot be ruled out. We therefore interpret the pooled estimate of g = 0.48 as an upper-bound estimate; the true effect likely lies between 0.42 and 0.48. This does not undermine the conclusion that AITS are effective, as even the adjusted estimate of g = 0.42 remains statistically and educationally significant, but it does warrant appropriate caution in interpretation. We conducted the subgroup analyses to examine the pre-specified moderators. Our results are summarized in Table 7.

Table 7. Subgroup Analysis of Effect Sizes by Moderator

Moderator	Subgroup	Hedges' g (95% CI)	Studies (n)
Subject Domain	Literacy/Language	0.55 (0.40, 0.70)	8
	STEM	0.32 (0.15, 0.49)	5
	General Academic Skills	0.44 (0.21, 0.67)	2
Primary Communication Mode	ASL/Sign-Based	0.61 (0.45, 0.77)	9
	Spoken Language	0.31 (0.12, 0.50)	6
AI System Type	Sign Recognition w/ Feedback	0.52 (0.35, 0.69)	7
	Adaptive Scaffolding	0.45 (0.25, 0.65)	5
	Generative AI Tutor	0.41 (0.20, 0.62)	3
System Maturity	Mature/Commercial Platform	0.60 (0.45, 0.75)	6
	Research Prototype	0.38 (0.23, 0.53)	9

5. Discussion

The authors would like to mention that the authors would like to mention that the systematic review and meta-analyses provided are the first comprehensive, quantitative results of the effectiveness of Artificial Intelligence tutoring systems for individuals who are Deaf/Hard-of-Hearing.

5.1. Interpretation of Key Findings

5.1.1. Efficacy of AI-Based Tutoring (RQ_2)

The moderate positive impact of AITS on academic performance (g = 0.48) provides additional evidence of the ability of AITS to positively impact the learning of DHH learners. In addition to replicating similar findings for the general education population in previous ITS meta-analysis [3] [32], this also expands those findings into a population of learners with a unique set of communication needs [25]. The significant effect size suggests AITS could help close the significant achievement gap identified by providing communication-accessible, personalized learning pathways.

5.1.2. What Works: Architectural Insights (RQ_1 and RQ_3)

The variation in results is considerable (I^2 = 67%), indicating the various technical methods used in this rapidly growing field. As such, the results of our moderator analysis provide practical information:

- AI Architecture Matters. The type of AI architecture used makes a difference; systems using sign language recognition provide a larger effect size (g = 0.55), while adaptive quiz systems are smaller in effect size (g = 0.41). Clearly, being able to interpret sign language directly allows learners to interact with the tutor using their preferred learning modality; therefore, it adds considerable value beyond just delivering the same content in an adaptive way.
- System Maturity Is Critical. More mature systems perform much better than prototypes; mature systems have a greater effect than prototype systems (g = 0.60 vs. 0.38). This is significant because it implies that those in the software engineering community working to design software for classrooms need to consider engineering elements as equally important to the sophistication of the algorithms they are implementing (e.g., in terms of both software architecture and algorithmic sophistication, implementation of a good algorithm will result in the greatest impact on the classroom).
- Bidirectional Communication. The bidirectional nature of the system has a positive effect on students' ability to learn from a tutor. The largest effect size occurs when the system supports both sign-to-text and text-to-sign communication (*g* = 0.63). Therefore, the AI tutor must provide both ways of communicating (through signs, text, and/or avatars) to facilitate full-duplex communication.
- Domain Specificity**.** The literacy/language interventions will have a bigger impact (*g*=0.55) than STEM interventions (g=0.32); likewise, AI helps people who have difficulty acquiring language acquire it. Since language support is an integral part of all content being taught to DHH learners, it is essential.

5.2. Implications

5.2.1. For Developers

Our findings suggest several design priorities:

1. Architectural integration: While SLR components are meant to co-exist as a network with each component providing information into Student Models and contributing to Tutor Decision Making, a system that can recognize a sign but cannot adjust the learning environment or provide instruction based on the use of that sign is only taking advantage of a very limited part of the ability of AI to interact with learners [33].
2. Bidirectional design: SLR should support the interpretation of students' signs and generate student friendly outputs/response through visual means.
3. Performance optimization: The system should emphasize low latency when providing inference and include appropriate error recovery for SLR because a 500 ms delay in providing feedback to learners would totally disrupt the flow of teaching.

5.2.2. For Researchers

There are still some areas of uncertainty when looking at the data, and we need more information regarding long-term effects since there aren't enough long-term studies; we don't know if the kids will be able to keep their skills. We also have limited data on the effects of AITS on social skills, social activities, and confidence (self-efficacy):

- Longitudinal, well-powered RCTs with standardized outcome measures.
- Studies across diverse linguistic populations (not just ASL users).
- Cost-effectiveness analysis to inform policy decisions.

5.2.3. For Educators and Policymakers

We have data from this study that support the continued investment in AITS for students who are DHH by supporting them in the areas of reading and language development, but we believe that the technology will not do everything; to use these technologies effectively, we must also ensure that teachers are appropriately trained, that technology is properly supported through the technical staff, and that the AITS technology is integrated into existing curricula.

5.3. Limitations

This research has a number of limitations, which are summarized in the following. For example, our study was not able to confirm publication bias statistically (Egger test p = 0.087), but we did find suggestive evidence of an asymmetrical distribution in the funnel plot, and we found that when we used the trim-and-fill method to adjust for this distribution, the pooled effect was reduced from g = .48 to g=0.42, which suggests that the actual effect could be lower than what is presented in this paper. In addition to these measurements, we relied upon a standardized mean and the mean of combined effect sizes across several forms of assessment, which may have led to obscuring true differences by research domain. With a significant amount of heterogeneity (I^2 = 67%), the pooled effect should be viewed as the mean of many different research domains, and while we conducted a few moderator analyses to partially identify sources of the heterogeneity, we were only partially successful in doing so (i.e., system maturity, mode of communication). In addition, based upon the results of our risk of bias analysis, seven studies had "some concerns" based on their lack of blinding, which may have led to performance bias in the studies, and three studies received a rating of "moderate risk" for confounding based upon the ROBINS-I. Our sensitivity analysis of the pooled effect when moderate-risk studies have been excluded showed no significant change to the estimate (g = .51; 95% CI: 0.38, 0.64), demonstrating that this effect is robust; thus, limiting the publication review to English-language studies may have introduced geographic bias.

6. Conclusion

This systematic review and meta-analysis affirm the significant potential of AI-based tutoring systems to improve educational outcomes for Deaf and Hard-of-Hearing learners. We found that AITS confers a moderate, statistically significant benefit (g = 0.48), with particularly strong effects for literacy learning and for learners who use ASL. Significantly, we think that effectiveness depends not just on whether a system uses AI, but on how architectural choices such as bidirectional communication, integrated SLR, and system maturity significantly moderate outcomes.

This study offers beneficial evidence for researchers, practitioners, and developers by showing that intelligent, adaptive technologies can support more accessible communication and personalized learning experiences. At the same time, fully realizing this potential will depend on continued efforts to create well-designed systems, implement them effectively, and better understand which approaches work best for different learners and settings.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data extracted for this review are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgments: We used generative AI to assist with editing for grammar and sentence structure. All content was reviewed and approved by the authors.

References:

1. K. Alasim, "Inclusion and d/Deaf and hard of hearing students: A qualitative meta-analysis," *International Journal of Disability, Development and Education,* vol. 70, no. 6, pp. 1120-1146, 2021. https://doi.org/10.1080/1034912X.2021.1931818
2. M. Marschark, E. J. Machmer and C. Convertino, "Understanding language in the real world," *The Oxford Handbook of Deaf Studies in Language,* pp. 431-452, 2016. https://doi.org/10.1093/oxfordhb/9780190241414.013.29
3. R. Nkambou, J. Bourdeau and R. Mizoguchi, Advances in intelligent tutoring systems, vol. 308, Springer, 2010. https://doi.org/10.1007/978-3-642-14363-2
4. W. Ma, O. O. Adesope, J. C. Nesbit and Q. Liu, "Intelligent tutoring systems and learning outcomes: A meta-analysis," *Journal of Educational Psychology,* vol. 106, no. 4, pp. 901-918, 2014. https://doi.org/10.1037/a0037123
5. M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif and M. A. Mekhtiche, "Hand gesture recognition for sign language using 3DCNN," *IEEE Access,* vol. 8, p. 79491–79509, 2020. https://doi.org/10.1109/ACCESS.2020.2990434
6. K. Papadimitriou and G. Potamianos, "Multimodal Locally Enhanced Transformer for Continuous Sign Language Recognition," in *Interspeech*, 2023. https://doi.org/10.21437/Interspeech.2023-2198
7. A. Barr, E. A. Feigenbaum and P. R. Cohen, The handbook of artificial intelligence, HeurisTech Press, 1981.
8. A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge.," *User modeling and user-adapted interaction,* vol. 4, pp. 253-278, 1994. https://doi.org/10.1007/BF01099821
9. M. Marschark and P. C. Hauser, Deaf cognition: Foundations and outcomes, Oxford University Press, 2008. https://doi.org/10.1093/acprof:oso/9780195368673.001.0001
10. J. V. Kahn, "A comparison of sign and verbal language training with nonverbal retarded children," *Journal of Speech, Language, and Hearing Research,* vol. 24, no. 1, pp. 113-119, 1981. https://doi.org/10.1044/jshr.2401.113
11. S. Pabis and J. Catalano, "Explicit and contextualized math vocabulary instruction with deaf and hard-of-hearing students," *The Journal of Deaf Studies and Deaf Education,* vol. 28, no. 4, pp. 424-425, 2023. https://doi.org/10.1093/deafed/enad012
12. R. J. Hoffmeister, "A piece of the puzzle: ASL and reading comprehension in deaf children," in *Language acquisition by eye*, Psychology Press, 1999, pp. 143-163. https://doi.org/10.4324/9781410601766-11
13. K. Pechilis and S. J. Raj, South Asian religions: Tradition and today, Routledge, 2012. https://doi.org/10.4324/9780203079935

14. N. Pope, J. Kahila, J. Laru, H. Vartiainen, T. Roos and M. Tedre, "An educational tool for learning about social media tracking, profiling, and recommendation," in *2024 IEEE International Conference on Advanced Learning Technologies (ICALT),* 2024. https://doi.org/10.1109/ICALT61570.2024.00038
15. K. Yin, A. Moryossef, J. Hochgesang, Y. Goldberg and M. Alikhani, "Including signed languages in natural language processing," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing,* 2021. https://doi.org/10.18653/v1/2021.acl-long.570
16. R. Rastgoo, K. Kiani and S. Escalera, "Sign language recognition: A deep survey," *Expert systems with applications,* vol. 164, p. 113794, 2021. https://doi.org/10.1016/j.eswa.2020.113794
17. N. Adaloglou, T. Chatzis, I. Papastratis, A. Stergioulas, G. T. Papadopoulos, V. Zacharopoulou, G. Xydopoulos, K. Atzakas, D. Papazachariou and P. Daras, "A comprehensive study on deep learning-based methods for sign language recognition," *IEEE Transactions on Multimedia,* vol. 24, pp. 1750-1762, 2021. https://doi.org/10.1109/TMM.2021.3070438
18. N. C. Camgoz, O. Koller, S. Hadfield and R. Bowden, "Sign language transformers: Joint end-to-end sign language recognition and translation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,* 2020. https://doi.org/10.1109/CVPR42600.2020.01004
19. L. Liu, Z. Yang, Y. Liu, X. Zhang and K. Yang, "A Sign Language Recognition Based on Optimized Transformer Target Detection Model," in *International Conference on Multimedia Technology and Enhanced Learning,* 2024. https://doi.org/10.1007/978-3-031-50580-5_16
20. R. Rastgoo, K. Kiani and S. Escalera, "Multi-modal deep hand sign language recognition in still images using restricted Boltzmann machine," *Entropy,* vol. 20, no. 11, p. 809, 2018. https://doi.org/10.3390/e20110809
21. K. Cagıltay, H. Cakır, N. Karasu, O. F. Islım and F. Cıcek, "Use of educational technology in special education: Perceptions of teachers," *Participatory Educational Research,* vol. 6, no. 2, pp. 189-205, 2019. https://doi.org/10.17275/per.19.21.6.2
22. E. Holmer, M. Heimann and M. Rudner, "Computerized sign language-based literacy training for deaf and hard-of-hearing children," *The Journal of Deaf Studies and Deaf Education,* vol. 22, no. 4, pp. 404-421, 2017. https://academic.oup.com/jdsde/article/22/4/404/4098239
23. H. Cheng, S. Chen, C. Perdriau, S. Mokkapati and Y. Huang, "LLM-powered AI tutors with personas for d/Deaf and hard-of-hearing online learners," *arXiv preprint,* vol. 1, 2024. https://doi.org/10.48550/arXiv.2411.09873
24. J. L. Luckner, A. M. Sebald, J. Cooney, J. Young and S. G. Muir, "An examination of the evidence-based literacy research in deaf education," *American Annals of the Deaf,* vol. 150, no. 5, pp. 443-456, 2005. https://doi.org/10.1353/aad.2006.0008
25. S. Steenbergen-Hu and H. Cooper, "A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning.," *Journal of Educational Psychology,* vol. 105, no. 4, pp. 970-987, 2013. https://doi.org/10.1037/a0032447
26. P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Kuttler, M. Lewis, W. Yih, T. Rocktaschel, S. Riedel and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," *Advances in Neural Information Processing Systems,* vol. 33, p. 9459–9474, 2020. https://doi.org/10.48550/arXiv.2005.11401
27. L. Yan, L. Sha, L. Zhao, Y. Li, R. Martinez-Maldonado, G. Chen, X. Li, Y. Jin and D. Gašević, "Practical and ethical challenges of large language models in education: A systematic scoping review," *British Journal of Educational Technology,* vol. 55, no. 1, pp. 90-112, 2023. https://doi.org/10.1111/bjet.13370
28. J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page and V. A. Welch, Cochrane handbook for systematic reviews of interventions, Wiley, 2019. https://doi.org/10.1002/9781119536604
29. M. Egger, G. D. Smith, M. Schneider and C. Minder, "Bias in meta-analysis detected by a simple, graphical test," *BMJ,* vol. 315, no. 7109, pp. 629-634, 1997. https://doi.org/10.1136/bmj.315.7109.629
30. M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hrobjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting and D. Moher, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ,* vol. 372, 2021. https://doi.org/10.1136/bmj.n71
31. K. N. Alasim, "Recognition and comprehension of multiple-meaning words: Examining a vocabulary intervention with hard of hearing students," *American Annals of the Deaf,* vol. 166, no. 3, pp. 262-283, 2021. https://doi.org/10.1353/aad.2021.0030
32. J. Cohen, Statistical power analysis for the behavioral sciences, Routledge, 1988. https://doi.org/10.4324/9780203771587
33. R. Luckin, S. Puntambekar, P. Goodyear, B. L. Grabowski, J. Underwood and N. Winters, Handbook of Design in Educational Technology, Routledge, 2013. https://doi.org/10.4324/9780203075227