© Journal of Accessibility and Design for All (JACCES), Volume 16, Issue 2, 2026, ISSN: 2013-7087
DOI: https://doi.org/10.17411/jacces.v16i2.788

Comprehension without load, engagement without persistence

Measuring whether generative AI is accessible to disabled learners

Galasso B., Federal University of São Paulo, Brazil,
ORCID 0000-0002-3677-7650, bruno.galasso@unifesp.br

Received: 2026-07-28 | Accepted: 2026-11-01 | Publication: 2026-11-11

Abstract: Generative artificial intelligence (GenAI) tools are increasingly offered to disabled learners to make learning accessible. Accessibility is usually evaluated by conformance: checking an interface against the success criteria of a technical standard such as WCAG 2.2. Such checks show that content can be reached, not how much effort using it requires. This systematic analysis of 92 studies (2023 to March 2026) examines whether research measures how much mental effort GenAI-supported learning demands of disabled learners, whether individual learners keep using these tools across repeated sessions or stop, and whether their confidence as learners grows. Seventeen outcome constructs were coded against Cognitive Load Theory, the Technology Acceptance Model and Self-Determination Theory. Studies mostly measured immediate responses: comprehension (57.6%), engagement (53.3%) and attitudes (48.9%). Few measured the processes that decide whether these results last: cognitive load (10.9%), disengagement (4.3%) and self-efficacy (17.4%), giving asymmetries of 5.3:1, 12.3:1 and 2.8:1. Only 27.2% of studies followed learners beyond a single session. A second coder agreed poorly with the original coding and judged that 16 of 23 sampled studies did not meet all inclusion criteria, so these figures are upper estimates. Cognitive load is defined as the demand a task places on working memory; only the load that arises when the output format does not match a learner's way of accessing information is treated as an accessibility issue. A three-axis framework adds evidence from use to conformance-based evaluation.

Keywords: accessibility, universal design, disabled learners, cognitive load, generative artificial intelligence, technology-mediated learning.

1. Introduction

About 15% of the world's population lives with some form of disability (World Health Organization, 2011). Disabled students often meet barriers that make ordinary learning materials harder to process, slower to use or impossible to access. Generative artificial intelligence (GenAI) tools, especially large language models such as ChatGPT, Gemini and Claude, are now offered to these learners as a response. They can simplify a text on request, explain a concept again without social cost and give help at the moment a difficulty arises. The expectation is that a learner who could not access a text before can now understand it, and that this understanding will last. Research on the topic has grown quickly: 96.7% of the 92 studies examined here were published in 2024–2026, a pace that mirrors the wider growth of AI research in higher education (Zawacki-Richter et al., 2019).

This growth has not yet produced a detailed account of the learner. The studies show which tools exist, which disability populations have been studied and where the research took place. They show much less about what happens to a disabled learner's thinking, behaviour and motivation when learning is mediated by these tools. Most studies observe learners at one moment, when the tool appears to work, and few follow them afterwards. Comprehension is measured in 57.6% of studies, but cognitive load, which indicates how much effort that comprehension required, in only 10.9%. Engagement is measured in 53.3%, but disengagement (learners reducing or stopping their use of the tool) in only 4.3%. Attitudes towards the tool are measured in 48.9%, but self-efficacy, the learner's belief that they can use the tool successfully on their own, in only 17.4%. This pattern is referred to here as measurement asymmetry: studies regularly measure early, visible successes and rarely measure the processes that decide whether those successes last.

An example shows why this matters. A blind student who uses a screen reader receives a long, unstructured answer from a GenAI tool. She understands it, and a comprehension test records a success. The test does not record the extra effort the format required: a screen reader presents text in sequence, so she has to keep the earlier parts of the answer in mind while listening to the rest (Sweller et al., 2011). If this happens in every session, the tool comes to feel difficult, and perceived ease of use is a key predictor of whether people continue to use a technology (Davis, 1989). Repeated difficulty can then weaken her belief that she is able to learn with the tool, a belief that motivational theories identify as a condition for sustained, self-directed use (Bandura, 1977; Deci & Ryan, 2000). She may conclude that the problem lies with her rather than with the design, and stop using the tool. In the current literature, almost none of the steps after the first is measured. Because disabled learners often have fewer alternatives, the cost of this gap falls unevenly on them.

This is an accessibility question that conformance-based evaluation cannot answer on its own. WCAG 2.2 (W3C, 2023) and the design-for-all tradition set out what makes an interface perceivable, operable, understandable and robust. A GenAI tool can meet these criteria and still be hard to use in practice for the learner it was meant to serve. GenAI output is generated anew in each interaction, so its length, density and structure change from one response to the next. Conformance testing checks the interface that displays the output; it cannot check every response the model produces. W3C guidance on cognitive accessibility also notes that it addresses barriers for people with cognitive and learning disabilities that are not included in the current WCAG success criteria (W3C, 2021). Knowing whether GenAI is accessible to disabled learners in practice therefore also requires evidence from use: about effort, about whether learners keep using the tools and about their confidence. Few studies have collected this evidence.

This review asks what the existing evidence allows researchers to know about disabled learners who use GenAI. RQ1: Which aspects of disabled learners' cognition, behaviour and affect do the included studies measure? RQ2: Within each of these three domains, which theoretically important learner processes are measured, and which are not? RQ3: Do the study designs (for example, follow-up over time or measurement in several domains) make it possible to trace how a learner's experience develops? RQ4: Do these measurement patterns differ across disability populations? The article makes three contributions. It describes how disabled learners are represented in GenAI research and what this representation leaves out. It uses three established theories to specify how GenAI-mediated learning can succeed or fail for these learners over time. It proposes a three-axis framework, with practical guidance for researchers, reviewers and funders, for measuring this process.

1.1. Key terms

The following terms are used throughout the article.

- Conformance-based accessibility evaluation: checking whether a digital product meets the success criteria of a technical standard, such as WCAG 2.2, usually by inspecting the interface.
- Cognitive load: the demand that a learning task, as presented, places on working memory, the limited system that holds and processes information during learning (Sweller et al., 2011).
- Accessibility-related load: the part of design-related (extraneous) load that arises when the format of the content does not match the way a learner accesses information (Section 2.1).
- Disengagement: a learner reducing or stopping their use of a tool over time.
- Self-efficacy: a learner's belief that they can successfully carry out a task (Bandura, 1977).
- Asymmetry ratio: the number of studies that measure a visible outcome divided by the number that measure the related process (for example, comprehension compared with cognitive load).

2. Theoretical framework

Three theories from the learning and motivation sciences describe what happens inside a learner when instruction is mediated by technology. Each points to a process that visible outcomes do not show. Together they define the three coding axes of this review and the standard against which the studies' account of disabled learners is assessed.

Cognitive Load Theory (CLT) starts from the fact that working memory has a limited capacity and that learning depends on how this capacity is used (Sweller, 1988; Sweller et al., 2011). In this article, cognitive load means the demand that a learning task, as presented, places on the learner's working memory. CLT distinguishes two main sources of load (Sweller, 2010; Kalyuga, 2011; Sweller et al., 2019). Intrinsic load comes from the complexity of the material itself, that is, from how many elements must be processed together; it depends on the task and on the learner's prior knowledge, and it cannot be removed without changing what is learned. Extraneous load comes from the way information is presented, for example from unclear structure, unnecessary information or a confusing interface; it does not contribute to learning and can be reduced through better design. Germane load, a third type in earlier accounts, is now usually described as the working-memory resources a learner devotes to intrinsic load rather than as a separate source (Kalyuga, 2011; Sweller et al., 2019). Because the same comprehension score can be reached with little or with great effort, comprehension measured alone cannot show whether learning was efficient or came at a cost the learner cannot repeat (Paas & van Merriënboer, 2020).

The Technology Acceptance Model (TAM) explains technology use through two beliefs: perceived usefulness and perceived ease of use (Davis, 1989; Venkatesh & Davis, 2000). When users find a technology difficult, they are less likely to adopt it or to keep using it. Meta-analytic evidence confirms that both beliefs predict technology adoption in education (Scherer et al., 2019). Studies that measure engagement but not disengagement only observe learners who remain users. This creates a survival bias: the learners who gave up, as TAM predicts some will, are not seen. For disabled learners this matters because a student may engage with a chatbot tutor while finding it hard to use; without a measure of disengagement, a study cannot show whether that engagement continues as barriers accumulate.

Self-Determination Theory (SDT) holds that self-directed (autonomous) motivation depends on meeting three basic needs: competence, autonomy and relatedness (Deci & Ryan, 2000; Ryan & Deci, 2000; Niemiec & Ryan, 2009). Self-efficacy, a learner's belief that they can organise and carry out the actions a task requires (Bandura, 1977), is closely linked to the need for competence and helps explain whether a positive attitude turns into sustained effort. A study that measures attitudes but not self-efficacy can show that learners like a tool without showing whether they feel able to use it for their own goals.

2.1. Task load, design load and accessibility-related load

Cognitive load is not, in itself, an accessibility barrier. Some load is necessary: a learner who studies a complex topic must process many related elements, and removing this intrinsic load would mean removing the content to be learned. Other load comes from poor design and affects all learners. This review therefore distinguishes three sources of effort.

- Intrinsic load (task load): effort required by the complexity of the learning content itself, given the learner's prior knowledge. It is part of learning and should be supported, not removed.
- Extraneous load (design load): effort caused by the way content or the interface is designed, for example long, unstructured or redundant responses. It affects all learners and can be reduced by better design.
- Accessibility-related load: the part of extraneous load that arises when the format of the output does not match the way a particular learner accesses information. A long answer without headings makes a screen-reader user listen to it in sequence and keep earlier parts in mind; dense written text demands more decoding effort from a learner with dyslexia; and for a deaf learner whose first language is a sign language, complex written text is processed in a second language. **(Galasso, 2026)** The same response can therefore create little extra effort for one learner and a great deal for another.

Only the third source is treated here as an accessibility variable. It is the mechanism through which content that conforms to accessibility standards can still be difficult to use in practice: two learners who receive the same output do not receive the same access. Universal design aims for environments that the widest possible range of people can use without adaptation (Rose & Meyer, 2002; CAST, 2018). Accessibility standards check this at the level of the interface; accessibility-related load arises at the level of the content the interface delivers, which in GenAI changes with every response. Design for all can be confirmed in use only if this load is measured. The distinction also has a methodological consequence. Many cognitive load measures, such as the NASA-TLX (Hart & Staveland, 1988), record overall effort and cannot separate these sources. Separating them requires instruments that distinguish intrinsic from extraneous load (e.g., Leppink et al., 2013) and designs that compare the same content in different formats, or across learners with different access needs. As Section 4.5 shows, the included studies did neither, so the load they recorded cannot be attributed to any one source.

The three theories lead to a shared account of successful technology-mediated learning. Learning succeeds not when a learner understands something once, but when understanding is reached at a cost the learner can sustain, with a tool the learner continues to choose, and with growing belief in their own capability. The theories also predict how these processes interact. Accessibility-related load makes a tool feel difficult, which TAM identifies as a barrier to continued use. Repeated difficulty weakens the learner's sense of competence, which SDT identifies as a condition for autonomous motivation. A learner overloaded by the format of the output

(cognitive axis) may come to see the tool as too effortful (behavioural axis) and herself as unable to use it (affective axis). Studies that measure only immediate outcomes, or only one axis, cannot observe this sequence. The three axes are therefore used both to code the studies and to describe how GenAI-mediated learning can fail for disabled learners, a failure that may be wrongly attributed to the learner rather than to the design.

3. Methodology

This review is a secondary, theory-driven analysis of how outcomes are measured in 92 studies on generative AI, disability and education published between 2023 and March 2026.

3.1. Relationship to the parent study

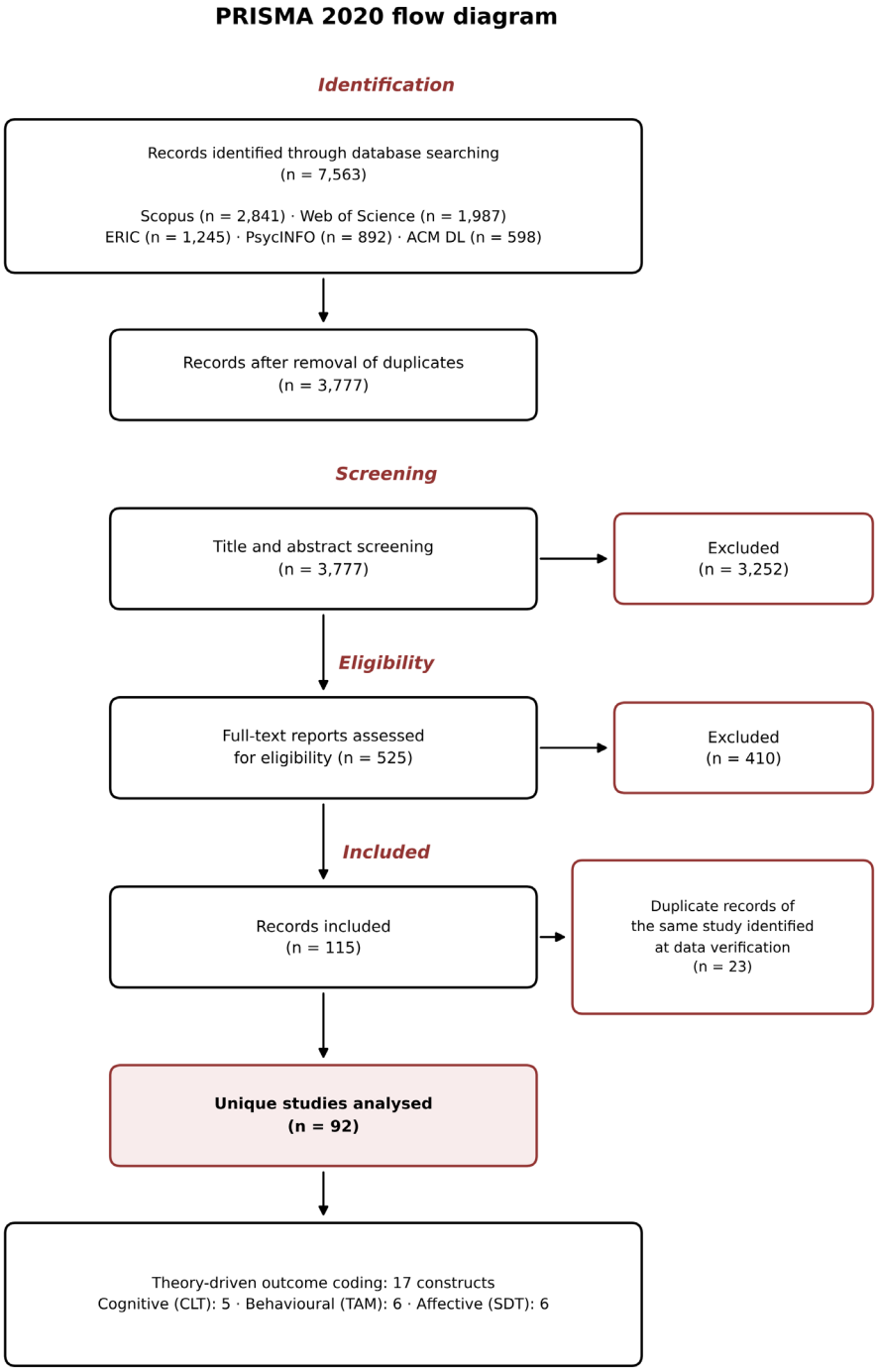
The corpus was first assembled for a companion descriptive review by the author, which has not been published. That review mapped the field: for each study it extracted, with a structured extraction form, the GenAI tool, the disability population, the country, the educational level and the type of technology, and it summarised these variables with frequency counts (population categories were assigned from a fixed list of 13 categories, such as visual impairment, autism or dyslexia, according to the population each study reported; a study could receive several categories). It did not code outcome measures. The present study re-analyses the studies included in that review to answer a different question: not what the field studies, but what it measures. The re-analysis followed four steps. First, the full texts of all included studies were retrieved again. Second, a new codebook of 17 outcome constructs was developed from the three theories (Appendix C). Third, each study was coded with this codebook, independently of the parent review's extraction (Section 3.3). Fourth, the disability population recorded in the parent review was linked to each study, solely to report results by population (Section 4.6). Apart from this population variable and bibliographic identifiers, no data from the parent review were reused. The two studies therefore use the same publications but extract different data to answer different questions.

3.2. Systematic search and corpus formation

The corpus was formed through a PRISMA-compliant systematic search (Page et al., 2021; Tricco et al., 2018; see Figure 1) executed in March 2026 across five databases: Scopus, Web of Science Core Collection, ERIC, PsycINFO, and the ACM Digital Library. A three-block Boolean strategy combined generative-AI terms (e.g., "generative AI," "ChatGPT," "large language model," "Copilot," "Gemini," "Claude"), disability and inclusive-education terms (e.g., "disabilit*," "special education," "accessibility," "autism," "dyslexia," "universal design for learning"), and educational-context terms; complete database-specific strings are available from the author on request. The search identified 7,563 records, reduced to 3,777 after deduplication; title/abstract screening excluded 3,252 records; full-text assessment of 525 reports excluded 410 against seven inclusion and seven exclusion criteria, yielding 115 records. During data verification, 23 of these records were found to duplicate another record of the same study (identical or truncated titles); they were merged, leaving 92 unique studies, which form the analytic corpus. Inclusion required a genuine generative-AI focus (not traditional or rule-based AI); an explicit, named disability population rather than general usability; an educational context; original empirical data; English language; and peer-reviewed status. The temporal distribution was concentrated: 3.3% (2023), 20.7% (2024), 62.0% (2025) and 14.1% (January to March 2026; percentages do not sum to 100

because of rounding), reflecting the field's emergence after the public release of ChatGPT in late 2022.

Figure 1. PRISMA 2020 flow diagram of the systematic search and screening process.



Note. Search conducted in five databases in March 2026. Duplicate removal before screening: 3,786 records.

3.3. Theory-driven outcome coding

Each study was coded for the presence or absence of 17 outcome constructs in three theory-aligned domains: cognitive (cognitive load, comprehension, learning gains, transfer and working memory), behavioural (adoption/intention, engagement, disengagement, persistence, self-

regulation and help-seeking) and affective (self-efficacy, motivation, satisfaction, attitudes, anxiety and frustration). Coding followed a fixed sequence: locating all named outcome measures in the Methods or Procedure section; checking the Results for further outcomes; assigning each outcome to the 17-construct taxonomy; and recording design features (longitudinal, pre-post, experimental) and, where reported, sample size. Only Methods and Results were coded, so that constructs mentioned only in an introduction were not counted. The threshold was deliberately inclusive: any construct named in Methods or Results was coded as measured. Because an inclusive threshold raises the counts of all constructs, the prevalence figures are upper estimates; the second-coder check (Section 3.4) shows that it raised the counts of the visible constructs most. Under this threshold, a construct was coded as measured when it was named as an outcome in Methods or Results, including outcomes reported by teachers or assessed on the output of the tool. The cognitive load code recorded any measure of mental effort; because no study used an instrument that separates intrinsic from extraneous load (Section 4.5), the code could not distinguish the three sources of load described in Section 2.1. The main results are prevalence percentages and asymmetry ratios. An asymmetry ratio is the number of studies that measure a visible outcome divided by the number that measure the related process on the same axis; because both counts share the denominator of 92, the ratio is computed from the counts.

For RQ4, each study was assigned to the disability population recorded in the parent review. The parent review assigned each study one or more categories from a fixed list of 13 population categories (for example, visual impairment, deaf or hard of hearing, autism, ADHD or dyslexia), plus “not identified”, according to the population the study reported. For this analysis, these categories were grouped into 11 groups (Table 2): dyslexia was grouped with general learning disabilities, and the general, multiple and neurodivergent (general) categories were grouped together. Because 64 of the 92 studies involve more than one category, a study can appear in several groups, and results by population are therefore reported as counts and are not tested statistically.

3.4. Coding reliability, quality appraisal and ethics

The author coded all 115 records. Intra-coder consistency was checked first on a calibration subset of 15 studies (13% of the 115 records), coded twice 72 hours apart, with 91% agreement across the 17 constructs; most disagreements concerned the boundary between satisfaction and attitudes and were resolved by refining the operational definitions. The full set of records was coded in four batches (30, 30, 30 and 25), each followed by a consistency check against earlier coding decisions.

To assess inter-coder reliability, a second coder with experience in inclusive education, who had not taken part in the original coding, independently coded a random subsample of 23 of the 92 unique studies (25%), drawn with a fixed random seed, using the codebook in Appendix C, without access to the first coder’s decisions. Agreement was assessed for each construct with percentage agreement and Cohen’s kappa (Cohen, 1960). Because kappa can be low for rare constructs even when agreement is high (Feinstein & Cicchetti, 1990), Gwet’s AC1 (Gwet, 2008) and positive agreement are also reported. Percentage agreement averaged 77.2% across the 17 constructs, but agreement corrected for chance was low (pooled Cohen’s kappa = 0.06; pooled AC1 = 0.71). The disagreements were systematic rather than random. Across the 391 construct decisions in the subsample, the first coder recorded 89 constructs as measured and the second coder 10, and the two coders agreed on 5 of these. The second coder recorded a construct only when it was measured as an outcome of the learners themselves, whereas the first coder had also recorded constructs assessed on AI outputs, reported by teachers or clinicians, or described

qualitatively. Full texts were not available to the second coder for four studies, which were coded from their abstracts; excluding them did not change the result (pooled kappa = 0.07). Positive agreement, the share of "measured" codes on which both coders agreed, was 0.10. The second coder also judged that 16 of the 23 studies did not meet at least one of four inclusion criteria (generative AI, a named disability population, an educational context and original empirical data). The prevalence figures reported in this article therefore rest on the first coder's inclusive threshold and should be read as upper estimates. Construct-level results are given in Appendix D, and the full construct-level statistics in the repository (see Data availability).

No formal appraisal of the methodological quality of the included studies was conducted, and this decision follows from the research questions. Appraisal tools such as the Mixed Methods Appraisal Tool (Hong et al., 2018) judge how far a study's findings can be trusted. This review does not synthesise findings: it records whether a construct was measured at all. A weaker study that measured cognitive load and a stronger study that did not are both correctly coded, and excluding or weighting studies by quality would change the description of the field rather than make it more accurate. PRISMA-ScR also treats critical appraisal as optional when the aim is to map how a field is studied (Tricco et al., 2018), which is the aim here. Three study characteristics that bear on measurement were nevertheless coded and are reported with the results: study design (longitudinal, pre-post, experimental), the reporting of sample size, and, for cognitive load, the use of validated instruments. Section 4.5 also tests whether the asymmetries hold in the methodologically stronger subsets of longitudinal and experimental studies.

The verification also bears on eligibility. Besides the second coder's judgements reported above, the parent review recorded no disability population for 8 of the 92 studies. These checks indicate that the corpus includes studies outside the stated inclusion criteria. It is therefore best read as a map of research at the intersection of generative AI, disability and education, rather than as a set of studies that all meet those criteria. The studies were not removed, because a full re-screening of the corpus was beyond the scope of this study; this is listed as a limitation.

The protocol was registered retrospectively on OSF; because registration followed completion, deviations cannot be checked against a pre-specified plan, although all protocol details are documented. As a secondary analysis of published, publicly available articles involving no human participants or personal data, the review did not require institutional ethical approval.

4. Results

4.1. Overall coverage

Across the 92 studies, cognitive outcomes were measured in 76.1% (n = 70), behavioural outcomes in 66.3% (n = 61) and affective outcomes in 75.0% (n = 69); 51.1% (n = 47) measured all three domains. At first sight, the field therefore seems to attend to the whole learner. Within each domain, however, measurement is strongly uneven (Table 1; Figure 2).

4.2. Cognitive axis: comprehension without load

Most cognitive codes concerned what disabled learners understood: comprehension in 57.6% (n = 53), learning gains in 40.2% (n = 37) and transfer in 26.1% (n = 24). Few measured what this understanding cost them: cognitive load in 10.9% (n = 10) and working memory in 5.4% (n = 5). For every study that measured cognitive load, 5.3 measured comprehension (5.3:1). Studies

usually recorded that a learner succeeded, but not how much effort the success required or whether the learner could repeat it.

4.3. Behavioural axis: engagement without persistence

Most behavioural codes concerned learners while they were using the tool: engagement in 53.3% (n = 49) and adoption or intention to use in 35.9% (n = 33). Few measured learners who stopped: disengagement in 4.3% (n = 4) and persistence in 9.8% (n = 9). Help-seeking, which shows that a learner is struggling and may need support or an accommodation, was measured in a single study (1.1%). The engagement-to-disengagement ratio is 12.3:1. As a result, the abandonment that TAM predicts when barriers outweigh benefits, and the learners for whom the tool did not work, remain largely unobserved.

4.4. Affective axis: attitudes without self-efficacy

Most affective codes concerned how learners felt about the tool: attitudes in 48.9% (n = 45) and satisfaction in 46.7% (n = 43). Fewer measured whether learners came to feel capable of using it: self-efficacy in 17.4% (n = 16) and frustration in 7.6% (n = 7). The attitudes-to-self-efficacy ratio is 2.8:1. Studies thus recorded learners' opinions of the tool more often than their confidence in using it independently.

4.5. Measurement depth and study design

The measures that would reveal a learner's difficulty while it is happening were the rarest. Help-seeking was measured in one study (1.1%), frustration in seven (7.6%), persistence in nine (9.8%) and working memory in five (5.4%). Each of these constructs was measured in fewer than 10% of the studies.

Measurement was also shallow. Only 27 studies (29.3%) reported their sample size (median 33; range 1–1,000; total N = 3,240), which points to mostly small, exploratory studies rather than confirmatory designs. Most studies (65 of 92) did not report a sample size at all, which also raises the question of whether all of them report original empirical data. None of the 10 studies that measured cognitive load used a validated cognitive load instrument such as the Leppink et al. (2013) scale or the NASA-TLX (Hart & Staveland, 1988); the instruments recorded for them were general questionnaires, rating scales or tests. None of these studies could therefore separate intrinsic, extraneous and accessibility-related load. Instrument quality was not coded for the other constructs, so the depth of measurement of self-efficacy or self-regulation, for example, cannot be assessed from the dataset. The asymmetry therefore concerns which constructs are measured and, for cognitive load at least, also how they are measured.

The pattern is not explained by weak study designs. Longitudinal designs, which are needed to test whether outcomes last, were used in 27.2% of studies (n = 25), pre-post designs in 15.2% (n = 14) and experimental designs in 22.8% (n = 21). Of the 25 longitudinal studies, 16 (64.0%) measured all three domains, more than the corpus average of 51.1%. Within each axis, however, the asymmetries remained in both stronger subsets. Among the 25 longitudinal studies, 14 measured comprehension and 3 cognitive load, 16 engagement and 3 disengagement, and 15 attitudes and 7 self-efficacy. Among the 21 experimental studies, the corresponding counts were 8 and 2, 11 and 2, and 10 and 4. In the original coding, measurement asymmetry therefore runs through the whole corpus and does not appear to be simply a sign of weaker studies.

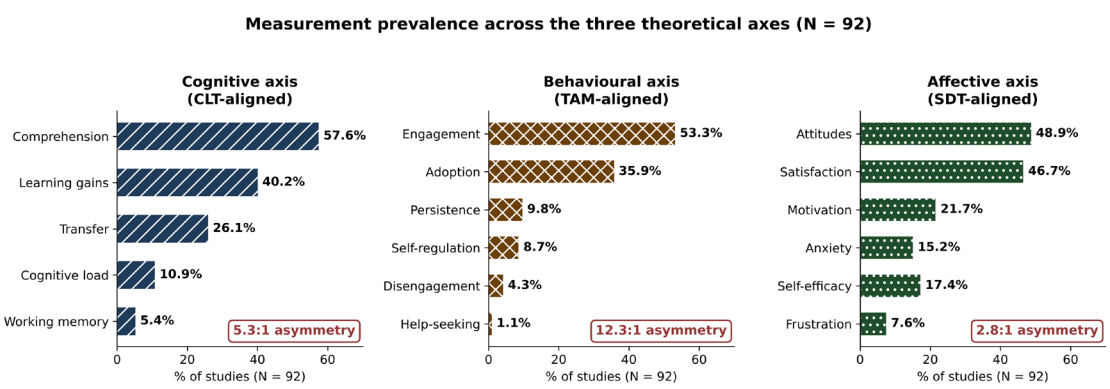
© Journal of Accessibility and Design for All (JACCES), Volume 16, Issue 2, 2026, ISSN: 2013-7087
DOI: https://doi.org/10.17411/jacces.v16i2.788

Table 1. Measurement prevalence across the 92 unique studies, by theoretical axis.

Construct	n	% of N = 92
Cognitive outcomes (total)	70	76,1
Comprehension	53	57.6
Learning gains	37	40.2
Transfer	24	26.1
Cognitive load	10	10.9
Working memory	5	5.4
Behavioural outcomes (total)	61	66,3
Engagement	49	53.3
Adoption/intention	33	35.9
Persistence	9	9.8
Self-regulation	8	8.7
Disengagement	4	4.3
Help-seeking	1	1.1
Affective outcomes (total)	69	75,0
Attitudes	45	48.9
Satisfaction	43	46.7
Motivation	20	21.7
Anxiety	14	15.2
Self-efficacy	16	17.4
Frustration	7	7.6
Design characteristics (total)	70	76,1
Multi-domain (all three)	47	51.1
Longitudinal	25	27.2
Pre-post	14	15.2
Experimental	21	22.8

Note: N = 92 unique studies. Percentages are of the full corpus; constructs are not mutually exclusive.

Figure 2. Measurement prevalence across the three theoretical axes (N = 92).



4.6. Results by disability population

Table 2 reports the measurement pattern for each disability population. The asymmetry appears in every group: in all ten groups other than "no population identified", more studies measured comprehension than cognitive load, engagement than disengagement, and attitudes than self-efficacy. For example, among the 26 studies involving deaf or hard-of-hearing learners, 13 measured comprehension and 2 cognitive load, and none measured disengagement. This gap matters because many deaf signers read GenAI output in a second language, and text-based interaction with these tools assumes written language as the default channel (Galasso, 2026). Disengagement was not measured in any study involving deaf or hard-of-hearing learners, learners with intellectual disability, learners with physical or motor disability, or learners with speech or language disability. Self-efficacy was measured most often in studies involving visually impaired learners (11 of 38). For 8 studies, the parent review identified no disability population; this is discussed in Section 3.4.

Table 2. Measurement of paired constructs by disability population (number of studies).

Population group	n	Compr.	Load	Engag.	Diseng.	Attit.	Self-eff.
Visual impairment	38	23	7	22	2	19	11
Deaf or hard of hearing	26	13	2	13	0	12	3
Autism	32	20	4	19	1	16	5
ADHD	23	17	2	13	1	11	3
Dyslexia and specific learning disabilities	33	27	5	22	2	17	6
Intellectual disability	17	10	1	11	0	6	1
Physical or motor disability	17	9	2	8	0	7	1
Speech or language disability	17	9	2	7	0	5	1
Mental health conditions	25	14	4	12	2	11	3
General, multiple or unspecified disability	45	31	6	29	3	23	6
No population identified	8	5	0	4	0	5	0

Note: n = number of studies involving the group. Populations as recorded in the parent review. Studies involving more than one population appear in several rows, so rows do not sum to 92. Compr. = comprehension; Load = cognitive load; Engag. = engagement; Diseng. = disengagement; Attit. = attitudes; Self-eff. = self-efficacy.

5. Discussion

The main finding concerns how disabled learners are represented. Across 92 studies, disabled learners are observed at the point where GenAI appears to work for them, and seldom afterwards. Studies record what learners understood but not what the understanding cost them; that they engaged, but not whether they came back; that they liked the tool, but not whether they came to trust their own ability to use it. This section discusses why this pattern exists, what it costs disabled learners and how it can be corrected.

5.1. Why surface outcomes dominate

Comprehension, engagement and attitudes are easier to measure, need shorter studies and tend to produce the positive results that publication rewards. Measuring cognitive load requires

specialised instruments and controlled conditions; measuring disengagement requires following learners over time and being willing to report negative results; and measuring self-efficacy requires domain-specific scales rather than general satisfaction items. Researchers therefore tend to measure what is easiest to measure, and publication incentives reinforce this. The field is also very new (96.7% of the studies were published in 2024–2026), and shared measurement practices have not yet developed. Similar patterns have occurred before. Early research on computer-assisted instruction focused on performance outcomes such as test scores and completion rates, and only later did research syntheses show the importance of cognitive and motivational processes (Mayer, 2009). Research on assistive technology long measured adoption without the psychological and social factors that predict abandonment (Lazar et al., 2017). Research on GenAI and disability may repeat these errors. The problem also reflects a wider concern in the psychological sciences: constructs are often measured without evidence that the measures capture what they claim to capture (Flake & Fried, 2020).

5.2. What the asymmetry costs disabled learners

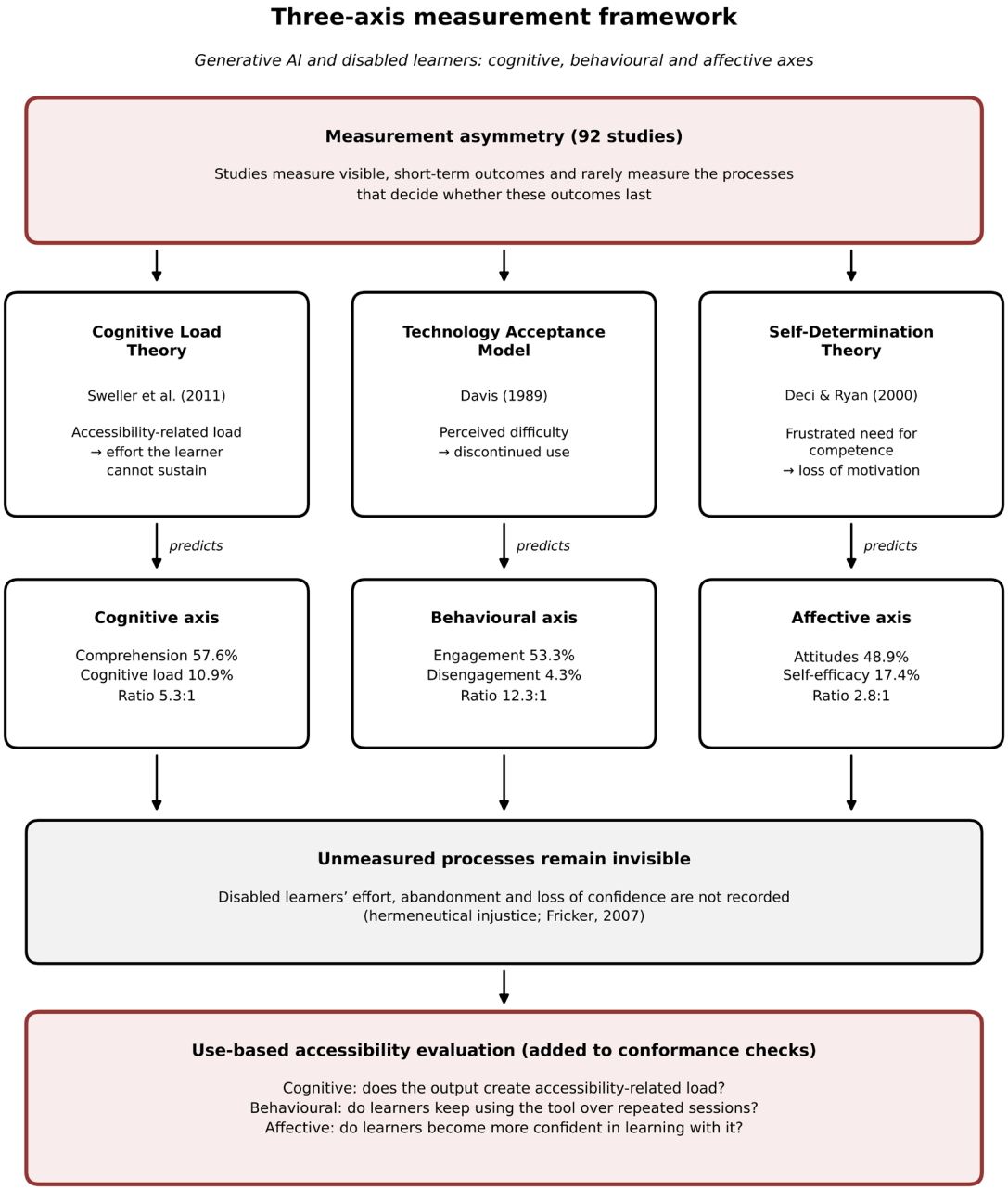
The asymmetry has direct consequences for learners. When comprehension is measured without load, a student who understands a GenAI explanation easily and a student who understands it only through great effort receive the same score, although theory predicts that the second is more likely to stop using the tool; the included studies cannot test this. When engagement is measured without disengagement, studies describe the learners who stayed and say little about those who left, who may be the very learners the tool was meant to help. When attitudes are measured without self-efficacy, a learner who enjoys a tool cannot be distinguished from one who has become able to use it independently. These gaps also reflect a long-standing tendency to measure what is convenient for researchers and designers rather than what matters to disabled people (Oliver, 1990; Shakespeare, 2006). Fricker (2007) uses the term hermeneutical injustice for a situation in which a group lacks the shared concepts needed to make its experience understood. When learners' effort, abandonment and loss of confidence are not measured, their experience of GenAI-mediated learning remains largely invisible to research and practice. The practical consequences are concrete. An institution that relies on this evidence when buying GenAI tools will find measures of comprehension, attitudes and engagement, but almost no measures of effort, abandonment or long-term motivation. Procurement decisions, accessibility evaluations and teaching recommendations will therefore rest on incomplete data about the learners they affect. The pattern is the same across disability populations (Section 4.6). For deaf or hard-of-hearing learners, for example, only 2 of 26 studies measured cognitive load and none measured disengagement, although many deaf signers read GenAI output in a second language.

5.3. Towards observing the learner

The three-axis framework (Figure 3) proposes three changes in measurement. On the cognitive axis, studies should pair comprehension measures with a validated cognitive load measure that distinguishes types of load (e.g., Leppink et al., 2013). Where possible, they should compare the same content in different output formats (for example, structured and unstructured responses) or across learners with different access needs, so that accessibility-related load can be separated from the load of the task itself. On the behavioural axis, studies should measure disengagement, non-adoption and persistence explicitly, rather than treating them as the absence of engagement, and should use longitudinal designs to test whether engagement continues after the first sessions. On the affective axis, studies should treat self-efficacy as a core outcome and measure it with scales suited to the task and population rather than with general satisfaction

ratings. These changes do not require new theories. They require applying established theories and instruments (CLT with differentiated load scales, TAM with measures of continued use, and SDT and self-efficacy theory with validated self-efficacy scales) to outcomes that are currently left unmeasured.

Figure 3. Three-axis measurement framework for generative AI and disability research.



5.4. Implications for theory, accessibility evaluation and practice

For theory, the concept of measurement asymmetry applies beyond this field. It reflects a distinction found across the learning sciences between measuring the visible products of instruction and measuring the processes that decide whether those products last. CLT separates performance from learning (Sweller et al., 2011), SDT separates compliance from autonomous engagement (Deci & Ryan, 2000), and universal design for learning emphasises sustained

engagement over one-off completion (Rose & Meyer, 2002; CAST, 2018). All three imply that understanding instructional effectiveness requires measuring processes as well as products. For accessibility evaluation, the implication is more specific. Current approaches to judging whether a GenAI tool is accessible are conformance-based and focus on the interface: they ask whether assistive technology can reach the output, not what using it costs. This is necessary but not sufficient. A design-for-all evaluation of a GenAI tool should add three questions based on use to the conformance checklist. Does the output create accessibility-related load, that is, more effort for learners who access it through a different route, such as a screen reader, than for other learners (cognitive axis)? Do learners keep using the tool over repeated sessions, or do they stop (behavioural axis)? Do learners become more confident in their ability to learn with it (affective axis)? For research practice, future studies should: (a) include at least one validated cognitive load measure alongside any comprehension assessment; (b) track disengagement, non-adoption and persistence over several sessions; (c) measure self-efficacy with scales suited to the task and the population; (d) use longitudinal or repeated-measures designs that can show whether early outcomes last, decline or reverse; and (e) report results separately for each disability population. Journal reviewers, funders and procurement bodies can use these criteria to encourage better measurement.

5.5. Limitations

Several limitations should be noted. First, the corpus is limited to peer-reviewed studies in English from five databases; grey literature, preprints and studies in other languages may show different patterns. Second, the presence/absence coding is inclusive, so the prevalence figures are upper estimates, and it cannot distinguish central from peripheral measures or judge the quality of instruments. Third, inter-coder reliability was assessed on a 25% subsample and was low (pooled kappa = 0.06), and the second coder judged that 16 of the 23 sampled studies did not meet at least one inclusion criterion. The prevalence figures are therefore upper estimates under an inclusive threshold, and the corpus includes studies outside the stated criteria; a full re-screening and double coding of the corpus is needed to confirm the findings. Fourth, no formal quality appraisal was conducted, so the findings describe how the field measures outcomes, not how strong its evidence about effects is. Fifth, because no included study used an instrument that separates types of load, the review cannot estimate how much of the reported load was accessibility-related. Sixth, several disability populations are represented by few studies, so comparisons between populations are descriptive. Seventh, the field is very new (96.7% of the studies were published in 2024–2026), so the asymmetries are best read as a baseline against which future progress can be measured. Finally, the review does not test whether better measurement would improve outcomes for learners; that would require experimental evidence beyond its scope.

6. Conclusions

The main conclusion of this review is that studies of GenAI for disabled learners mostly measure short-term comprehension, engagement and attitudes, and rarely measure whether these early results can be sustained. Across 92 studies, cognitive load was measured in 10.9%, disengagement in 4.3% and self-efficacy in 17.4%. In other words, more than 82% of the studies say nothing about whether disabled learners' confidence in their own ability changes when they learn with a GenAI tool. These gaps are not random: they fall on the processes that Cognitive Load Theory, the Technology Acceptance Model and Self-Determination Theory identify as

decisive for whether technology-mediated learning lasts. These figures are upper estimates (Section 3.4). In the original coding, the direction of the gap is the same in every subset examined, and the few constructs the second coder recorded point the same way. In the original coding, the asymmetry holds in every disability population examined. Closing these gaps does not require new theory. It requires applying existing theories and instruments to disabled learners: measuring the load that instruction imposes with scales that separate types of load, such as that of Leppink et al. (2013); recording which learners stop using a tool, when they stop and why, as studies of continued use do (Bhattacherjee, 2001); and measuring self-efficacy with validated scales (e.g., Compeau & Higgins, 1995) as an outcome in its own right. Journal editors and reviewers should expect at least one measure on each theoretical axis; funders should support the longitudinal designs (currently 27.2% of studies) that can follow learners beyond first use; and researchers should develop instruments suited to specific disability populations rather than relying only on general scales. Until then, the field cannot say whether GenAI makes learning more accessible for disabled learners in practice, or only at first contact.

7. Acknowledgements

The author thanks Maria Laura Sánchez Toca for independently coding the reliability subsample.

Fundings. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

AI disclosure. During the preparation of this article the author used Anthropic's Claude (a large language model) to support the tabulation of coded data, to assist with formatting and language editing, and to check consistency between reported figures and the coding dataset. During revision, the same tool assisted with language simplification, the computation of agreement statistics from the two coders' data and consistency checks. No part of the research design, the coding decisions, the analysis or the interpretation was delegated to an AI tool. The author reviewed and edited all AI-assisted output and takes full responsibility for the content, accuracy and originality of the article.

Competing interests. The author declares no competing interests.

Ethics. The study is a secondary analysis of published, publicly available research. No human participants or personal data were involved, and institutional ethical approval was not required.

Data availability. The coding protocol, the coding dataset (including the disability population of each study), the second coder's reliability data and the analysis materials are openly available in an Open Science Framework repository: https://doi.org/10.17605/OSF.IO/TMXG6.

8. Bibliography

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. MIS Quarterly, 25(3), 351–370. https://doi.org/10.2307/3250921

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

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46. https://doi.org/10.1177/001316446002000104

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. MIS Quarterly, 19(2), 189–211. https://doi.org/10.2307/249688

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. Journal of Clinical Epidemiology, 43(6), 543–549. https://doi.org/10.1016/0895-4356(90)90158-L

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. Advances in Methods and Practices in Psychological Science, 3(4), 456–465. https://doi.org/10.1177/2515245920952393

Fricker, M. (2007). Epistemic injustice: Power and the ethics of knowing. Oxford University Press.

Galasso, B. J. B. (2026). Visuo-spatial cognition and the redesign of human-AI interaction in generative artificial intelligence. Discover Artificial Intelligence, 6, 950. https://doi.org/10.1007/s44163-026-02040-2

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. British Journal of Mathematical and Statistical Psychology, 61(1), 29–48. https://doi.org/10.1348/000711006X126600

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Education for Information, 34(4), 285–291. https://doi.org/10.3233/EFI-180221

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? Educational Psychology Review, 23(1), 1–19. https://doi.org/10.1007/s10648-010-9150-7

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). Research methods in human-computer interaction (2nd ed.). Morgan Kaufmann.

Leppink, J., Paas, F., van der Vleuten, C. P. M., van Gog, T., & van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. Behavior Research Methods, 45(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

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

Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. Theory and Research in Education, 7(2), 133–144. https://doi.org/10.1177/1477878509104318

Oliver, M. (1990). The politics of disablement. Macmillan.

Paas, F., & van Merriënboer, J. J. G. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. Current Directions in Psychological Science, 29(4), 394–398. https://doi.org/10.1177/0963721420922183

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Rose, D. H., & Meyer, A. (2002). Teaching every student in the digital age: Universal design for learning. ASCD.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. Computers & Education, 128, 13–35. https://doi.org/10.1016/j.compedu.2018.09.009

Shakespeare, T. (2006). Disability rights and wrongs. Routledge.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. Educational Psychology Review, 22(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive load theory. Springer. https://doi.org/10.1007/978-1-4419-8126-4

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. https://doi.org/10.1007/s10648-019-09465-5

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., … Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Annals of Internal Medicine, 169(7), 467–473. https://doi.org/10.7326/M18-0850

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. Management Science, 46(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

W3C. (2021). Making content usable for people with cognitive and learning disabilities (W3C Working Group Note, 29 April 2021). World Wide Web Consortium. https://www.w3.org/TR/coga-usable

W3C. (2023). Web Content Accessibility Guidelines (WCAG) 2.2. World Wide Web Consortium. https://www.w3.org/TR/WCAG22

World Health Organization. (2011). World report on disability. WHO.

Zawacki-Richter, O., Marín, 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, 39. https://doi.org/10.1186/s41239-019-0171-0

Appendix A. Systematic search strategy

The corpus was assembled through a PRISMA-compliant systematic search executed in March 2026 across five databases: Scopus, Web of Science Core Collection, ERIC (EBSCOhost), PsycINFO (EBSCOhost), and the ACM Digital Library. The search combined three Boolean blocks with the AND operator; within each block, terms were combined with OR and truncation (*) was used to capture morphological variants. Database-specific field tags and syntax were adapted to each platform (e.g., TITLE-ABS-KEY in Scopus; TS in Web of Science).

Block A — Generative-AI terms (21 terms). “generative AI” OR “ChatGPT” OR “GPT-4” OR “large language model*” OR “LLM” OR “Copilot” OR “Gemini” OR “Claude” OR “AI chatbot*” OR “conversational AI” (among 21 terms).

Block B — Disability and inclusive-education terms (45 terms). “disabilit*” OR “special education” OR “inclusive education” OR “accessibility” OR “deaf*” OR “blind*” OR “dyslexia” OR “autism” OR “ADHD” OR “neurodiverg*” OR “assistive technolog*” OR “universal design for learning” (among 45 terms).

Block C — Educational-context terms (14 terms). “education” OR “learning” OR “teaching” OR “student*” OR “classroom” (among 14 terms).

The search identified 7,563 records, reduced to 3,777 after deduplication. Title-and-abstract screening excluded 3,252 records. Full-text assessment of 525 reports excluded 410 against the criteria below, yielding 115 records, which correspond to 92 unique studies after removal of 23 duplicate records (Section 3.2). The complete database-specific query strings, the deduplication log and the exclusion log are available from the author on request.

Appendix B. Inclusion and exclusion criteria

Inclusion criteria. A study was included if it (IC1) focused on generative AI rather than traditional machine-learning or rule-based systems; (IC2) addressed an explicit, named disability or accessibility population rather than general usability; (IC3) was set in an educational context, formal or informal; (IC4) reported original empirical data; (IC5) was written in English; (IC6) was published between January 2022 and March 2026; and (IC7) was a peer-reviewed journal article, a conference full paper (≥4,000 words), or a book chapter.

Exclusion criteria. A report was excluded if it involved (EC1) traditional AI without a generative component; (EC2) a general “accessibility” focus without disability specificity; (EC3) a non-educational application; (EC4) non-empirical work (commentary, editorial, protocol); (EC5) a review or meta-analysis; (EC6) an inaccessible full text; or (EC7) insufficient methodological reporting to permit outcome coding.

Appendix C. Construct taxonomy and operational coding definitions

Each study was coded for the presence or absence of 17 outcome constructs organised across three theory-driven axes. The threshold for coding a construct as measured was deliberately inclusive: a construct was recorded as present when a study reported any named outcome measure, instrument, or dependent variable corresponding to it in the Methods or Results

section. Coding was restricted to Methods and Results to avoid false positives from constructs merely discussed in introductions or literature reviews.

C.1. Cognitive axis (Cognitive Load Theory)

- Cognitive load — any measure of mental effort or working-memory burden imposed by the task or tool (e.g., validated subjective load scales, dual-task measures, physiological indices, or explicit load-related self-report items). The code does not distinguish intrinsic, extraneous and accessibility-related load (Section 2.1), because no included study used an instrument that separates them.
- Comprehension — measures of understanding of presented material (e.g., comprehension tests, quiz accuracy, recall of content).
- Learning gains — pre-to-post change in knowledge or skill attributable to the intervention.
- Learning transfer — application of learned material to novel problems or contexts not seen during instruction.
- Working memory — direct measures of working-memory demand or capacity (e.g., span tasks, working-memory load indices).

C.2. Behavioural axis (Technology Acceptance Model)

- Technology adoption / intention to use — measures of uptake, acceptance, or stated intention to continue using the tool.
- Engagement — behavioural or self-reported involvement with the tool or task (e.g., time on task, participation, interaction frequency).
- Disengagement — measures of withdrawal, drop-off, non-adoption, or abandonment of the tool over time.
- Persistence — continuation of use or effort across multiple sessions or over an extended period.
- Self-regulated learning — measures of learners' planning, monitoring, or control of their own learning process.
- Help-seeking — measures of learners' requests for assistance, from the tool or from others, during learning.

C.3. Affective axis (Self-Determination Theory)

- Self-efficacy — learners' beliefs in their own capability to succeed at the task, ideally via domain-specific scales.
- Intrinsic motivation — measures of interest-driven, autonomous engagement, as distinct from externally driven compliance.
- Satisfaction — global evaluative ratings of the experience or tool.
- Attitudes — evaluative dispositions towards the tool or towards AI in learning.
- Anxiety — measures of task-, technology-, or evaluation-related anxiety.

- Frustration — measures of frustration, need-thwarting, or negative affective response during use.

Appendix D. Coding protocol and quality assurance

Coding followed a structured sequence for each of the 115 records: (a) locate all named outcome measures, instruments, and dependent variables in the Methods/Procedure section; (b) examine the Results section for any additional outcomes not listed in Methods; (c) classify each identified outcome into the 17-construct taxonomy; and (d) record study-design features (longitudinal, pre-post, experimental) and, where reported, participant count and instruments.

A two-phase quality-control procedure was applied. In the first phase, a calibration subset of 15 studies (13% of the 115 records) was coded, set aside, and recoded after a 72-hour interval to assess intra-coder consistency. Agreement between the two codings was 91% across all 17 constructs, with the few discrepancies concentrated on the boundary between satisfaction and attitudes, which share conceptual overlap in many instruments; operational definitions were refined to resolve these boundary cases. In the second phase, the full set of records was coded in four sequential batches (n = 30, 30, 30, 25), each followed by a consistency check against cumulative coding decisions to mitigate coder drift.

Inter-coder reliability was then assessed on a random subsample of 23 of the 92 unique studies (25%). The subsample was drawn with a fixed random seed from the 92 unique study identifiers. A second coder with experience in inclusive education, who had not taken part in the original coding, received the codebook (Appendix C), a short written briefing and the full texts (available for 19 of the 23 studies; the other four were coded from their abstracts), and coded the 23 studies independently, without access to the first coder's decisions. For each construct, percentage agreement, Cohen's kappa and Gwet's AC1 were computed. Agreement by construct ranged from 43.5% (engagement) to 100.0% (help-seeking, which neither coder recorded in the subsample, so that kappa could not be computed); kappa ranged from −0.17 to 0.47. Disagreements were not resolved by discussion and the first coder's decisions were not changed; both coders' decisions and the construct-level statistics are deposited in the repository so that readers can examine them.

Appendix E. Coding dataset

The accompanying file 06_Supplementary_Coding_Data.csv contains one row per unique study (N = 92); the 23 duplicate records identified during verification were merged, with a construct coded as present if it was coded in either record. The file has the following fields: a sequence identifier; the study title; the source file reference; reported participant count; the disability population recorded in the parent review; the 17 outcome-construct indicators grouped by axis (cog_*, beh_*, aff_*, coded True/False); the instruments used where specified; and three design indicators (longitudinal, pre_post, experimental). A second file contains the second coder's decisions for the 23-study reliability subsample. These files are the empirical basis for every prevalence figure, asymmetry ratio and agreement statistic reported in the manuscript and permit full reconstruction of the analysis.

© Journal of Accessibility and Design for All (JACCES), Volume 16, Issue 2, 2026, ISSN: 2013-7087
DOI: https://doi.org/10.17411/jacces.v16i2.788

How to cite this article

Galasso B. (2026). Comprehension without load, engagement without persistence: Measuring whether generative AI is accessible to disabled learners. Journal of Accessibility and Design for All, 16(2), 196-215. https://doi.org/10.17411/jacces.v16i2.788

© Journal of Accessibility and Design for All (JACCES), ISSN 2013-7087, is published by the Universitat Politècnica de Catalunya, Barcelona Tech, with the sponsoring of ONCE Foundation for Cooperation and Social Inclusion of People with Disabilities. This issue is free of charge and is available in electronic format.

This work is licensed under an Attribution-Non-Commercial 4.0 International Creative Commons License. Readers are allowed to read, download, copy, redistribute, print, search, or link to the full texts of the articles or use them for any other lawful purpose, giving appropriate credit. It must not be used for commercial purposes. To see the complete license contents, please visit http://creativecommons.org/licenses/by-nc/4.0.

Jacces is committed to providing accessible publication to all, regardless of technology or ability. The present document grants vital accessibility since it applies to WCAG 2.2 and PDF/UA recommendations. The evaluation tool used has been Adobe Acrobat® Accessibility Checker. If you encounter problems accessing the content of this document, you can contact us at jacces@catac.upc.edu

Jacces is published under the DOAS (Diamond Open Access Standard) framework of the DIAMAS project, which certifies the quality and excellence of academic journals. It has been evaluated using the Self-assessment tool provided by FECYT, from the European Research Area (ERA). This article meets the requirements of the Diamond standard https://diamas.fecyt.es