Qədim Diyar Oktyabr 2026, Cild: 8, Sayı: 10, s. 41−51
Ancient Land October 2026, Volume: 8, Issue: 10, pp. 41−51
DOI: https://doi.org/10.36719/2706-6185/64/41-51

From Digital Reading Traces to Adaptive Instruction an Evidence Centered Diagnostic Model for Primary Education

Nurjamal Abdurazakova

Abstract. *This conceptual study develops an evidence centered model for using digital reading traces in the adaptive pedagogical diagnosis of primary school pupils. The problem addressed is that response accuracy, reading time, rereading, navigation, hint use, and answer revision are often treated as direct measures of comprehension although each trace is open to several interpretations. The study synthesizes evidence centered assessment design, formative assessment, process data research, digital reading studies, cognitive diagnostic adaptive testing, and ethical learning analytics. The resulting model connects six elements: a map of the reading construct, purposefully designed tasks, trace capture, evidence rules, teacher mediated instructional decisions, and continuous validation. It distinguishes observable events from inferred reading processes and requires triangulation before a learner is assigned a diagnostic profile. A construct to action matrix is proposed for word recognition, fluency, vocabulary in context, literal retrieval, inferential integration, evaluation, and metacognitive monitoring. The model also introduces safeguards for interface effects, language and cultural bias, unstable response time, privacy, opaque automation, and premature labeling. Its central claim is that adaptation should modify the next learning opportunity rather than define the child. Digital traces become educationally valuable only when they are linked to an explicit claim, generated by a suitable task, interpreted with uncertainty, and reviewed by a teacher. The framework is intended for future piloting in Karakalpak medium and other multilingual primary classrooms.*

Keywords: *reading literacy, primary education, digital traces, adaptive diagnosis, evidence centered design, formative assessment, multilingual education*

Ajiniyaz Nukus State Pedagogical Institute, PhD in Pedagogical Sciences, Nukus, Uzbekistan
E-mail : abdrnurj@mail.ru
Received: 21 April 2026; Accepted: 28 Agustos 2026; Published online: 28 October 2026
© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Rəqəmsal oxu izlərindən adaptiv tədrisə: ibtidai təhsil üçün sübutlara əsaslanan diaqnostik model

Nurcamal Abdurazakova

Xülasə. *Bu konseptual tədqiqat ibtidai sinif şagirdlərinin adaptiv pedaqoji diaqnostikasında rəqəmsal oxu izlərindən istifadə üçün sübuta əsaslanan model hazırlayır. Problem ondan ibarətdir ki, cavabın düzgünlüyü, oxu vaxtı, təkrar oxu, naviqasiya, ipucundan istifadə və cavabın dəyişdirilməsi çox vaxt birbaşa anlama göstəricisi kimi qəbul edilir, halbuki hər bir iz bir neçə cür izah oluna bilər. Tədqiqat sübuta əsaslanan qiymətləndirmə dizaynını, formativ qiymətləndirməni, proses məlumatları araşdırmalarını, rəqəmsal oxunu, koqnitiv diaqnostik adaptiv testi və etik öyrənmə analitikasını birləşdirir. Təklif olunan model altı elementi əlaqələndirir: oxu konstruktunun*

xəritəsi, məqsədli tapşırıqlar, izlərin qeydə alınması, sübut qaydaları, müəllim vasitəsilə tədris qərarları və davamlı validasiya. Model müşahidə olunan hadisələri ehtimal edilən oxu proseslərindən ayırır və şagirdə diaqnostik profil verilməzdən əvvəl məlumatların üçbucaqlı yoxlanmasını tələb edir. Söz tanıma, səlislik, kontekstdə lüğət, birbaşa məlumatın tapılması, nəticə çıxarma, qiymətləndirmə və metakoqnitiv nəzarət üçün konstrukt fəaliyyət matrisi təqdim olunur. İnterfeys təsiri, dil və mədəni qərəz, qeyri-sabit cavab vaxtı, məxfilik, qeyri-şəffaf avtomatlaşdırma və vaxtından əvvəl etiketləmə risklərinə qarşı qoruyucu şərtlər müəyyənləşdirilir. Əsas nəticə budur ki, adaptasiya uşağı müəyyənləşdirməməli, növbəti öyrənmə imkanını dəyişdirməlidir. Modelin gələcəkdə qaraqalpaq dilli və digər çoxdilli ibtidai siniflərdə sınaqdan keçirilməsi nəzərdə tutulur.

Açar sözlər: *oxu savadlılığı, ibtidai təhsil, rəqəmsal izlər, adaptiv diaqnostika, sübuta əsaslanan dizayn, formativ qiymətləndirmə, çoxdilli təhsil*

Ajiniyaz adına Nukus Dövlət Pedaqoji İnstitutu, Pedaqoji elmlər üzrə fəlsəfə doktoru, Nukus, Özbəkistan
E-poçt : abdrnurj@mail.ru
Daxil oldu: 21 Aprel 2026; Qəbul edildi: 28 Avqust 2026; Onlayn dərc edildi: 28 Oktyabr 2026
© Müəllif(lər) 2026. Bu açıq girişli məqalə Creative Commons Attribution–NonCommercial 4.0 Beynəlxalq Lisenziyasının (CC BY-NC 4.0) şərtlərinə uyğun olaraq yayımlanmışdır.

Introduction

Reading literacy in primary education is not a single score. It involves accurate word recognition, fluent coordination of decoding and meaning, vocabulary use, retrieval of explicitly stated information, integration of ideas, inference, evaluation, and monitoring of one's own understanding. International reading frameworks similarly treat comprehension as an interaction among the reader, the text, the purpose, and the context rather than as the mechanical reproduction of words (Mullis & Martin, 2019). This multidimensional character creates a diagnostic problem. A pupil may select an incorrect answer because the relevant word was unfamiliar, a sentence relation was missed, the task instruction was misunderstood, attention was interrupted, or the interface was difficult to operate. The visible outcome alone does not identify the cause.

Digital assessment appears to offer a solution because it records not only the final response but also events that occur while a pupil works. Systems can capture response time, sequence of pages, rereading, use of audio or hints, changes of answer, skipped items, and patterns of navigation. These records are commonly described as process data or digital traces. They make otherwise invisible parts of task engagement available for analysis. Recent reviews show rapid growth in the use of log based process data in computer based assessment, but they also emphasize unresolved questions of construct interpretation, validity, preprocessing, and generalizability (He & Cui, 2025). More data therefore do not automatically produce better diagnosis.

The key danger is inferential overreach. A long response time may indicate careful integration, weak decoding, distraction, or unfamiliarity with the device. Reopening a paragraph may reflect productive monitoring or confusion. A pupil who answers quickly may be fluent, may guess, or may have seen the text before. A trace is an event generated by the interaction of a pupil, a task, an interface, and a situation. It is not a psychological process in itself. If the system converts one ambiguous event directly into a label such as weak comprehension, the apparent precision of the technology conceals a weak evidential argument.

This concern is especially important in the early grades. Younger pupils differ in motor coordination, familiarity with digital devices, sustained attention, language experience, and confidence. Screen based reading can also alter comprehension conditions; a meta analysis of children's reading found

that outcomes depend on age, text type, and the design of digital features rather than on a simple opposition between paper and screen (Furenes et al., 2021). In multilingual contexts, interface language and text familiarity may add another layer of variation. A diagnostic system designed for Karakalpak medium classrooms must therefore distinguish reading difficulty from linguistic, cultural, and technological barriers.

Formative assessment offers a useful criterion for judging digital diagnosis. Assessment becomes formative when evidence is interpreted and used to decide what to do next in teaching and learning (Black & Wiliam, 2009). A dashboard that displays response times but does not support a defensible instructional response is information, not yet formative diagnosis. Conversely, a modest set of well chosen traces can be valuable if it helps a teacher select an appropriate prompt, text, grouping arrangement, or scaffold and then observe whether the pupil responds as expected.

Research on reading literacy in Uzbekistan has developed tools for evaluating pupils through text analysis. Abdurazakova (2025) describes the importance of relating assessment to the actual operations pupils perform when they work with a text. The present study extends that concern into digital environments but does not duplicate the author's dissertation topic. Its focus is the evidential bridge between trace, inference, and pedagogical action. The research question is: How can digital reading traces be organized into an adaptive diagnostic system that remains construct valid, interpretable to teachers, sensitive to multilingual contexts, and ethically responsible?

The aim is to develop a theoretical and methodological model rather than to claim that a completed algorithm has already been validated. Three objectives follow: to specify the reading constructs that a system may reasonably diagnose; to formulate a chain connecting tasks and traces to instructional decisions; and to establish validity and ethical safeguards. The contribution is an evidence centered architecture in which adaptation changes the learner's next opportunity instead of turning provisional evidence into a fixed description of the child.

Methods

The study uses qualitative conceptual synthesis and design based reasoning. It does not report an experiment, a school sample, or machine learning accuracy. Twelve sources were selected because together they cover the problem's essential domains: primary reading frameworks, digital reading, evidence centered assessment design, formative assessment, process data, cognitive diagnostic adaptive testing, learning analytics ethics, education technology governance, and the author's prior work on text based reading assessment. The corpus contains peer reviewed journal articles, an assessment framework, an OECD working paper, and an international education report.

The analytical procedure comprised five steps. First, the reading construct was decomposed into dimensions that can plausibly be elicited through primary grade tasks. Second, for each dimension, possible observable behaviors were listed without treating them as direct measures. Third, an evidence argument was specified by asking what combination of task features and traces would strengthen or weaken a claim about the pupil. Fourth, each claim was connected to a low risk instructional response that could be reversed after further evidence. Fifth, the complete chain was examined for threats arising from the interface, language, culture, privacy, and automated decision making.

Evidence centered design provided the organizing logic. It begins with claims about what a learner knows or can do, identifies evidence that would support those claims, and designs tasks capable of eliciting that evidence (Mislevy et al., 2003). This order is important. Starting with whatever a

platform can record encourages measurement by convenience. Starting with the construct requires every collected variable to justify its relationship to a reading process and an instructional purpose.

The synthesis also integrates three traditions that are often separated: data based decision making, assessment for learning, and diagnostic testing. Their thoughtful combination can produce more valid formative decisions when multiple sources of evidence are used and learning is continuously optimized (van der Kleij et al., 2015). The proposed model accordingly treats automated analysis as decision support. The teacher reviews the context, considers classroom evidence, chooses an intervention, and evaluates the pupil's response.

Conceptual adequacy was tested against four boundary cases. The first was a pupil with slow but accurate reading; the second, a pupil with fast inaccurate responses; the third, a pupil whose performance changed after audio support; and the fourth, a pupil who performed differently on culturally familiar and unfamiliar texts. A useful model had to permit more than one explanation at the first observation, specify what additional evidence was needed, and recommend an action that could help discriminate among explanations. This method guards against deterministic profiling.

Use of artificial intelligence
A generative artificial intelligence tool was used for English language editing, structural organization, and consistency checking during manuscript preparation. It was not used to create empirical observations, participant data, or bibliographic sources. The author remains responsible for the conceptual argument, source verification, translation, and final submitted text.

Results

The unit of diagnosis
The first result is a change in the unit of diagnosis. The appropriate unit is not the isolated trace and not the child as a fixed type. It is a provisional claim about a reading process under specified task conditions. For example, repeated returns to a sentence do not by themselves establish an inference problem. A defensible claim might be narrower: when causal relations are implicit in a short informational text, the pupil rereads the relevant sentences, selects distractors based on local keywords, and can answer after a teacher asks the pupil to connect two propositions. This pattern supports a tentative hypothesis about integration rather than a permanent label.

The distinction creates three layers. The event layer contains what the platform records: choices, time stamps, navigation, audio activation, highlighting, or revision. The evidence layer contains interpreted patterns that are meaningful because the task was designed to elicit a particular process. The decision layer contains a pedagogical response and a rule for checking its effect. Errors arise when a system skips the evidence layer and moves directly from event to decision. Evidence centered diagnosis makes the warrant between layers explicit.

Reading literacy should be represented as a related set of processes. For primary education, the model uses seven diagnostic dimensions: word recognition and accuracy; fluency; vocabulary in context; literal retrieval; inferential integration; evaluation and reflection; and metacognitive monitoring. These dimensions are analytically useful, but they are not assumed to be completely independent. Vocabulary affects inference, decoding affects available attention, and monitoring may cause rereading. The system must allow overlapping explanations and seek corroboration.

The six stage evidence chain
The second result is a six stage chain. Stage one defines the claim and its developmental meaning. Stage two designs a task in which the target process is necessary while irrelevant interface demands

are minimized. Stage three records only traces needed for the declared purpose. Stage four combines outcome and process evidence using transparent rules and expresses uncertainty. Stage five translates the hypothesis into a teacher review and an adaptive instructional action. Stage six evaluates the pupil's response and the fairness of the inference, then retains, revises, or rejects the hypothesis.

This chain prevents data abundance from controlling the educational question. It also creates an audit trail. A teacher or researcher should be able to ask why a variable was recorded, which claim it supports, how alternative explanations were considered, what threshold or rule was used, and how the decision will be checked. Maddox (2023) argues that process data can add value to large scale assessment but require careful definition and interpretation. In classroom diagnosis, transparency is even more important because an inference immediately shapes a child's learning opportunity.

Adaptive testing and adaptive instruction must also be distinguished. Computerized adaptive testing selects items to estimate ability efficiently. Cognitive diagnostic computerized adaptive testing can report mastery of several reading attributes and reduce unnecessary testing burden. A large primary reading study demonstrated the potential of such an approach with calibrated items and validated attributes (Li et al., 2023). Rapid adaptive assessment has likewise shown that efficient reading measurement is technically possible (Ma et al., 2025). Yet an efficient estimate is not itself an instructional plan. The present model extends adaptation beyond item selection to the next scaffold, text, question sequence, modality, or teacher conference.

Table 1

Construct to evidence to action matrix

Reading dimension	Task and possible traces	Permitted inference	Adaptive response and check
Word recognition and accuracy	Read controlled words or sentences; accuracy, self correction, audio request, item time.	A repeated decoding difficulty is plausible only across items and after interface familiarity is checked.	Provide grapheme word pattern practice and supported rereading; retest with parallel items.
Fluency	Read a short passage; pace, pauses, phrasing, accuracy, rereading.	Slow time alone is insufficient; rate must be interpreted with accuracy and prosodic evidence.	Use modeled and repeated reading; compare accuracy and phrasing on a matched passage.
Vocabulary in context	Infer word meaning from sentence and paragraph clues; choice pattern, clue reopening, revision.	Difficulty using contextual clues is plausible when decoding and task language are adequate.	Highlight semantic clues, compare meanings, then present a new word in a similar context.
Literal retrieval	Locate explicitly stated information; navigation path, paragraph return, correct selection.	The pupil may have inefficient search or weak task interpretation; reading aloud the instruction helps discriminate.	Teach question to text matching and scanning; check on a text with the same structure.

Reading dimension	Task and possible traces	Permitted inference	Adaptive response and check
Inferential integration	Connect information across sentences; rereading of both locations, distractor type, explanation.	A claim requires evidence that both propositions were understood separately but not integrated.	Use linking questions or a graphic relation; check transfer to a new causal or referential link.
Evaluation and reflection	Compare claims with evidence or purpose; annotation, justification, revision after counterevidence.	The quality of justification matters more than click sequence or time.	Model evidence based justification; request a second judgment using explicit criteria.
Metacognitive monitoring	Allow confidence rating, voluntary rereading, hint use, and correction.	Mismatch between confidence and performance is a hypothesis about monitoring, not motivation or character.	Teach stop check explain routines; observe whether calibration improves across tasks.

Source. Table compiled by the author. Rules for interpreting digital traces

The third result is a set of interpretation rules. First, every trace must be conditional on task design. Time on an item means different things when a pupil must scroll, manipulate a graphic, listen to audio, or read a continuous paragraph. Second, a claim should normally require convergence of at least two indicators, such as an outcome pattern and a process pattern, or digital evidence and a brief teacher elicitation. Third, the model should preserve alternative explanations until a discriminating task is administered. Fourth, thresholds must be calibrated for grade, language, text type, device, and administration conditions rather than copied from another population.

Fifth, evidence should be accumulated across occasions. Primary pupils' performance is sensitive to fatigue, attention, and familiarity. A single unusual response time should not change a profile. Sixth, missing traces must not automatically be coded as failure. A pupil may not use a hint because it was unnecessary, unnoticed, linguistically unclear, or socially perceived as undesirable. Seventh, uncertainty should be visible. A teacher report should distinguish supported, tentative, and unsupported claims and show which observation would most efficiently clarify the case.

These rules imply a transparent teacher facing report. The report should state the assessed process, the tasks used, the pattern observed, plausible alternatives, confidence, recommended action, and review date. It should not display a hidden risk score without explanation. Nor should it rank pupils when the purpose is individual diagnosis. Comparison with developmental expectations can be useful, but the central question is which learning opportunity is likely to help this pupil now and what evidence will show whether it helped.

Adaptive pedagogical responses

The fourth result is a distinction among four forms of adaptation. Task adaptation changes the difficulty, representation, or sequence of the next item. Scaffold adaptation changes the support, for example by adding a prompt, glossary, segmentation, or audio model. Interaction adaptation changes the social arrangement, such as teacher conferencing, peer explanation, or guided group reading.

Evidence adaptation changes what the system observes next in order to test an alternative explanation. The last form is frequently neglected, yet it is essential when traces are ambiguous.

Adaptation should be minimal, purposeful, and reversible. If a pupil answers accurately but slowly, the next step should not automatically lower text complexity. A matched task can test whether slowness comes from decoding, careful monitoring, or interface navigation. If performance improves only after vocabulary support, the system may recommend preteaching a small set of words while maintaining the comprehension demand. If an audio instruction changes performance but audio text presentation does not, the barrier may lie in task language rather than reading comprehension. Each response is therefore also a diagnostic probe.

The teacher remains a necessary participant because classroom knowledge provides context not contained in the log. The teacher may know that the pupil is new to the device, was interrupted, speaks another home language, or usually explains the same relation orally. The model does not idealize unaided teacher judgment; it structures it. Digital evidence should challenge intuition when patterns are consistent, while the teacher should challenge automated inference when contextual evidence provides a credible alternative.

Validity fairness and ethical safeguards

The fifth result is a safeguard framework. Learning analytics creates ethical questions concerning privacy, informed participation, transparency, power, and the obligation to act on information (Slade & Prinsloo, 2013). Primary pupils require heightened protection because they cannot carry the full responsibility for understanding data practices. Data minimization is therefore the default: record only what serves a specified diagnostic purpose, retain it only as long as necessary, restrict access, and prohibit secondary uses that have not been justified.

Fairness requires more than equal access to a device. A trace can be biased by text familiarity, dialect, translation, font, navigation, touch accuracy, connectivity, and prior digital experience. The digital system should include a short interface familiarization, accessible presentation, alternative input where appropriate, and checks for differential patterns across language and learner groups. UNESCO (2023) cautions that education technology should serve learners and teachers on transparent public terms rather than allow technological possibility to determine educational purpose.

The strongest ethical rule is that a diagnostic profile must not become an identity. Terms such as weak reader or low ability compress a provisional performance pattern into a personal characteristic. The model instead uses conditional language: current evidence suggests difficulty integrating information across sentences under these task conditions. This wording makes space for development, context, and error. It also aligns with formative assessment, which treats evidence as a resource for improving learning rather than as a final judgment.

Table 2

Major threats and required safeguards

Threat	Invalid shortcut	Required safeguard	Audit question
Response time ambiguity	Slow equals poor comprehension.	Interpret time with accuracy, task demand, navigation, and repeated observations.	Could the same time pattern arise from a competent strategy?

Threat	Invalid shortcut	Required safeguard	Audit question
Interface confounding	Every click reflects reading cognition.	Use familiarization tasks and minimize irrelevant interaction demands.	Would performance change on paper or with another device?
Language and cultural bias	A common item functions identically for all pupils.	Review text, instruction, and examples with local language educators; compare subgroup patterns.	Is the difficulty located in the target construct or in cultural linguistic access?
Opaque automation	The score is valid because an algorithm produced it.	Document features, rules, uncertainty, and human review; validate on the target population.	Can a teacher explain and contest the inference?
Premature labeling	One session defines the learner.	Use provisional claims, multiple occasions, and scheduled reclassification.	What new evidence could reverse the conclusion?
Privacy and surveillance	More stored data necessarily improve learning.	Minimize data, define retention and access, obtain appropriate consent, prohibit unrelated reuse.	Is each collected field necessary for a declared educational purpose?
Teacher deskilling	The dashboard replaces professional judgment.	Design reports for inquiry, context entry, decision rationale, and follow up.	Does the system strengthen the teacher's reasoning and responsibility?

Source. Table compiled by the author**.**

Discussion

The proposed model clarifies what is new and what remains to be demonstrated. Digital assessment creates a richer observational record, but its value lies in the quality of the evidence argument, not the number of logged events. Evidence centered design supplies the connection among claim, evidence, and task (Mislevy et al., 2003). Formative assessment supplies the requirement that evidence change subsequent teaching and learning (Black & Wiliam, 2009). Process data research supplies techniques and warnings about how traces are generated and analyzed (He & Cui, 2025; Maddox, 2023). The present framework integrates these strands into a classroom diagnostic cycle.

The model differs from purely predictive learning analytics. Prediction asks which pupils are likely to obtain a later outcome. Diagnosis asks which learning process is currently plausible, what alternative explanations remain, and which next action can both support learning and improve the inference. A model may predict low performance accurately using socioeconomic or behavioral correlates without identifying a teachable cause. Such prediction is unsuitable for adaptive instruction if it lowers expectations or routes pupils away from demanding texts. Actionability must be grounded in the construct, not merely in statistical association.

The model also places limits on cognitive diagnostic adaptive testing. The study by Li et al. (2023) shows that attribute level feedback can be generated efficiently when the item bank and cognitive model are carefully developed. ROAR CAT similarly demonstrates the feasibility of rapid online adaptive reading assessment (Ma et al., 2025). These advances are important, but a local classroom system still requires validated construct definitions, a sufficiently large and balanced item bank, calibration, accessibility testing, and evidence that feedback improves instruction. Importing an algorithm without these foundations would create false precision.

For Karakalpak medium classrooms, the framework has a specific methodological implication. Texts and task instructions must be developed with attention to the linguistic structure, vocabulary distribution, genres, and cultural knowledge available to pupils. Translation alone does not establish equivalence. A task intended to measure inference may instead measure knowledge of an unfamiliar setting. Local teachers and language specialists should participate in construct mapping, item review, think aloud studies, and interpretation of unexpected traces. The author's earlier text analysis work provides a relevant pedagogical base for this development (Abdurazakova, 2025).

Implementation can begin without complex artificial intelligence. A first prototype may use a small validated task set, transparent rule based combinations, and a teacher conference protocol. This is methodologically preferable to training an opaque model on a small convenience sample. The initial aim should be to learn which tasks elicit interpretable evidence and which recommendations teachers can use consistently. Statistical models can be introduced after the construct, data quality, and decision process are stable.

The framework also changes the evaluation criterion for a diagnostic platform. Technical performance such as classification accuracy is necessary but insufficient. Evaluation should examine construct representation, subgroup fairness, teacher comprehension, decision consistency, instructional uptake, learner response, privacy compliance, and unintended narrowing of the curriculum. A system that predicts well but is not understandable or actionable may have little educational value. A system that increases short term test performance by repeatedly training narrow item types may undermine broader reading literacy.

Digital reading itself must remain part of the validity discussion. Furenes et al. (2021) show that screen and paper outcomes vary with design conditions. Therefore, a digital diagnostic system should not assume medium neutrality. Some tasks may validly assess digital navigation as part of contemporary literacy, while others should minimize navigation to isolate comprehension. The intended construct must decide which digital actions are relevant and which are contamination.

Institutionally, implementation requires governance. Schools need a clear statement of purpose, named responsibility for data access, procedures for correction, a retention schedule, and a mechanism through which teachers and families can question results. Professional development should focus on reasoning from evidence rather than on operating a dashboard. Teachers should practice generating alternative explanations, selecting discriminating tasks, and documenting why an adaptation was chosen. Such capacity protects both scientific validity and professional agency.

Limitations

This article is conceptual. It does not establish the reliability, classification accuracy, usability, or learning impact of a functioning diagnostic system. The proposed reading dimensions require operational definitions and empirical validation in the target grades and languages. The review is selective rather than systematic, and the matrix does not capture every component of reading, including motivation, background knowledge, oral language, and extended written response. Digital traces are platform dependent, so the same event may not be comparable across devices or interfaces.

Ethical recommendations also require alignment with applicable national data protection and child safeguarding rules before implementation.

Future research

Future research should proceed in four linked phases. First, qualitative cognitive laboratories should examine how pupils interpret tasks and interfaces in Karakalpak, Uzbek, and other relevant languages. Second, a local item bank should be built from a construct specification and reviewed for linguistic and cultural accessibility. Third, pilot studies should test trace reliability, alternative explanations, subgroup fairness, and agreement between system hypotheses and structured teacher evidence. Fourth, a classroom trial should compare ordinary feedback with model guided adaptation, measuring not only assessment accuracy but also changes in reading performance, teacher decision quality, learner confidence, and time burden.

Research should also test whether particular traces have incremental diagnostic value beyond response accuracy. A feature should be retained only if it improves a construct relevant inference or a pedagogical decision. Studies should predefine exclusion and missing data rules, document device conditions, and report uncertainty. Participatory design with teachers, pupils, families, language experts, and data protection specialists can reveal risks that technical validation alone will miss.

Conclusion

Digital traces can support adaptive diagnosis in primary reading, but they are not evidence by default. Their meaning depends on the reading construct, the task that generated them, the interface, the learner's language and experience, and the reasoning rule used to connect observation to claim. A scientifically responsible system must therefore make this chain visible.

The evidence centered model developed here links construct mapping, task design, trace capture, transparent inference, teacher mediated adaptation, and continuous validation. It requires corroboration, preserves alternative explanations, expresses uncertainty, and uses each intervention as an opportunity to learn more about the pupil. Its ethical principle is equally clear: adaptation should change the next learning opportunity, not define the child. With local linguistic validation, careful governance, and empirical testing, the model can guide the development of useful diagnostic tools for Karakalpak medium and other multilingual primary classrooms.

Declarations

Conflict of interest. The author declares no conflict of interest.
Funding. No external funding was received for this conceptual study.
Ethics statement. The study did not involve human participants or personal data; therefore, institutional ethics approval and participant consent were not applicable.
Data availability. No empirical dataset was generated or analyzed.

References

Abdurazakova, N. S. (2025). Current state of assessment of reading literacy of primary school pupils based on text analysis. *International Journal of Pedagogics*, *5*(2), 115–117. https://doi.org/10.37547/ijp/Volume05Issue02-30?utm_source=chatgpt.com

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*, 5–31. https://doi.org/10.1007/s11092-008-9068-5

Furenes, M. I., Kucirkova, N., & Bus, A. G. (2021). A comparison of children's reading on paper versus screen: A meta-analysis. *Review of Educational Research*, *91*(4), 483–517. https://doi.org/10.3102/0034654321998074

He, S., & Cui, Y. (2025). A systematic review of the use of log-based process data in computer-based assessments. *Computers & Education*, *228*, Article 105245. https://doi.org/10.1016/j.compedu.2025.105245

Li, Y., Huang, C., & Liu, J. (2023). Diagnosing primary students' reading progression: Is cognitive diagnostic computerized adaptive testing the way forward? *Journal of Educational and Behavioral Statistics*, *48*(6), 842–865. https://doi.org/10.3102/10769986231160668

Ma, W. A., Richie-Halford, A., Burkhardt, A. K., Kanopka, K., Chou, C., Domingue, B. W., & Yeatman, J. D. (2025). ROAR-CAT: Rapid Online Assessment of Reading ability with Computerized Adaptive Testing. *Behavior Research Methods*, *57*(1), Article 56. https://doi.org/10.3758/s13428-024-02578-y

Maddox, B. (2023). *The uses of process data in large-scale educational assessments* (OECD Education Working Papers No. 286). OECD Publishing. https://doi.org/10.1787/5d9009ff-en

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series*, *2003*(1), i–29. https://doi.org/10.1002/j.2333-8504.2003.tb01908.x

Mullis, I. V. S., & Martin, M. O. (Eds.). (2019). *PIRLS 2021 assessment frameworks*. TIMSS & PIRLS International Study Center, Boston College. https://pirls2021.org/frameworks/?utm_source=chatgpt.com

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

UNESCO. (2023). *Global education monitoring report 2023: Technology in education: A tool on whose terms?* https://doi.org/10.54676/UZQV8501

van der Kleij, F. M., Vermeulen, J. A., Schildkamp, K., & Eggen, T. J. H. M. (2015). Integrating data-based decision making, assessment for learning and diagnostic testing in formative assessment. *Assessment in Education: Principles, Policy & Practice*, *22*(3), 324–343. https://doi.org/10.1080/0969594X.2014.999024