Teacher-centered technology design for educational assessment

Thomaz Edson Veloso da Silva[1], Guilherme Corredato Guerino[2], Luiz Rodrigues[3]

[1]Educometrika LTDA
Fortaleza – CE – Brazil

[2]State University of Paraná (UNESPAR)
Apucarana – PR – Brazil

[3]Federal Technological University of Paraná (UTFPR)
Apucarana – PR – Brazil

thomaz@educometrika.com.br, guilherme.guerino@ies.unespar.edu.br

luizrodrigues@utfpr.edu.br

Abstract. *This paper presents the design and evaluation of a mobile-first, teacher-centered application that enables photo-based student assessment using standard smartphones. The system generates real-time feedback dashboards tailored to teachers, principals, and school managers, promoting data-informed decisions across classroom, school, and system levels. Two empirical studies were conducted: a prototype usability test with five teachers and a real-world case study in a public school. Results highlight the app's usability, alignment with teachers' workflows, and potential to enhance pedagogical planning in under-resourced contexts. We conclude with design guidelines for scalable, low-infrastructure assessment technologies.*

1. Introduction

In recent years, the use of digital technologies to support classroom assessment has grown substantially, driven by the promise of more timely and actionable feedback for teachers and school leaders. Traditional paper-and-pencil assessments often have long turnaround times and limited scope for immediate pedagogical adaptation [Shute 2008]. Meanwhile, learning analytics platforms have shown the potential to surface patterns in student performance. Yet, many tools remain data-centric rather than teacher-centered, limiting their uptake in everyday practice [Topali et al. 2024, Bichsel 2012, Datnow and Park 2014].

Teacher-centered design emphasizes the importance of aligning technological advantages with the actual workflows, preferences, and decision-making needs of educators [Alfredo et al. 2024]. By focusing on usability, contextual relevance, and minimal disruption to existing routines, such approaches are known to increase adoption and sustained use [Xia et al. 2022, Nicol and Macfarlane-Dick 2006]. However, most existing solutions require substantial infrastructure (e.g., dedicated scanners or networked devices) or advanced data-science expertise, barriers that are particularly acute in under-resourced schools and districts [Isotani et al. 2023, Veloso et al. 2023].

To address these gaps, we developed a mobile-first prototype application, hereafter referred to as Nino, which enables teachers to collect assessment data simply by photographing student answer sheets with a standard smartphone camera. The app processes

DOI: 10.5753/sbie.2025.12616

images in the cloud to extract item-level responses, generates classroom- and school-level summaries, and delivers contextualized feedback dashboards tailored to teachers, principals, and system managers. This three-tier feedback model supports evidence-based pedagogical decisions at the classroom, school, and district levels.

Based on that context, the primary objectives of this paper are to:

1. Design and implement a photo-based assessment system centered on teachers and requiring minimal infrastructure for execution.
2. Evaluate the usability and workflow integration of the prototype with in-service teachers (Study 1).
3. Investigate the real-world integration of the system within pedagogical and administrative tasks in a school environment (Study 2).
4. Develop design guidelines for scalable teacher-centered assessment technologies in low-resource settings.

To achieve these objectives, we investigate how usable and efficient Nino is for teachers collecting and reviewing assessment data in a controlled prototype setting and what design features and workflow considerations facilitate sustained adoption of photo-based assessment technology among educators.

We conducted two empirical studies to assess our prototype's feasibility and usability. In Study 1, we engaged a cohort of five in-service teachers in a controlled prototype assessment session to collect usability ratings, task completion metrics, and qualitative insights into workflow integration. In Study 2, we deployed the system in a real school environment over two weeks, examining adoption patterns, decision-making changes, and perceived utility among teachers.

This paper provides a detailed description of a teacher-centered, photo-based assessment app requiring minimal infrastructure, presents empirical evidence from a prototype evaluation (Study 1) highlighting usability and workflow fit for educators, offers insights from a real-world deployment (Study 2) demonstrating the app's potential to inform pedagogical and administrative decisions, and derives design guidelines for scalable, teacher-centered assessment technologies in low-resource settings.

The remainder of the paper is organized as follows. Section 2 reviews related work on digital assessment and teacher-centered technology design. Section 3 describes the system architecture and design rationale of Nino. Sections 4 and 5 present the methodology and results of our two studies. Section 6 discusses implications, limitations, and future directions. Finally, Section 7 concludes the paper.

2. Related Work

The emergence of mobile technologies has spurred several innovations in digital formative assessment, enabling faster turnaround and more granular data than traditional paper-and-pencil methods. Rasiq *et al.* proposed a smartphone-based answer-sheet processing pipeline that leverages on-device OCR to extract student responses [Rasiq et al. 2019]. Similarly, Dui *et al.* evaluated a tablet-based system for automatically scoring handwritten responses in preschoolers, grade-schoolers, adults, and elderlies, highlighting the trade-offs between recognition accuracy and user training overhead [Dui et al. 2019].

In parallel, the field of learning analytics has shifted toward more teacher-centered dashboard designs, recognizing that data alone is insufficient unless it is meaningfully integrated into educators' workflows. Hoz-Ruiz *et al.* surveyed contemporary analytics platforms and distilled seven design principles that promote teacher agency, including contextualized recommendations and low-friction data capture [de La Hoz-Ruiz et al. 2024]. Schildkamp conducted a mixed-methods study of dashboard adoption in four U.S. districts, finding that alignment with existing pedagogical routines was the strongest predictor of sustained use [Schildkamp 2019].

Privacy, ethics, and infrastructural constraints have also been prominently featured in recent work. Isotani *et al.* discuss the "data zero" problem in low-resource contexts, where limited connectivity and device heterogeneity necessitate hybrid on-device/cloud architectures to preserve data integrity and student anonymity [Isotani et al. 2023]. Building on this, Fachola *et al.* propose a federated learning approach for educational data, allowing model updates without raw data leaving the device, thereby addressing privacy concerns and bandwidth limitations [Fachola et al. 2023].

Finally, there is growing interest in multi-stakeholder feedback systems that serve teachers, school leaders, and district managers. Silva *et al.* introduced a tiered reporting framework that aggregates classroom math assessments into classroom-level performance indicators, facilitating pedagogical strategic planning [Veloso et al. 2023]. However, many of these systems still rely on bulky hardware or require extensive IT support, underscoring the need for lightweight, photo-based solutions that can be deployed with minimal infrastructure.

In recent literature, our work builds on these strands by combining a mobile-first, photo-capture pipeline with a teacher-centered dashboard design and a three-tier feedback model.

2.1. Theoretical and Conceptual Framework

Our work is grounded in two complementary theoretical traditions: data-informed decision-making in education and teacher-centered learning analytics design.

Data-Informed Decision Making (DIDM) [Wang 2021] emphasizes how educators use assessment data as evidence to inform instructional strategies, curricular adjustments, and school improvement plans. DIDM frameworks typically distinguish between three levels of decision-making: classroom (micro), school (meso), and municipality/system (macro), each requiring tailored data visualizations and recommendations to be actionable by different stakeholders. This paper adopts the tri-level model, which identifies key decision points and information needs at each level and emphasizes formative assessment cycles that link data collection, interpretation, and pedagogical action.

Teacher-Centered Learning Analytics posits that tools must align with teachers' existing workflows, pedagogical beliefs, and professional agency to achieve sustained integration [Alfredo et al. 2024, Topali et al. 2024, Xia et al. 2022]. [Veloso et al. 2023] highlights how technology can either support or undermine teacher agency depending on the degree of automation versus interpretability; we build on this by ensuring our dashboards offer both automated summaries and raw item-level detail for professional inquiry.

Figure 1 illustrates our conceptual model: (1) *Data Capture* via smartphone photo, (2) *Data Processing* through on-device OCR and secure cloud pipelines, (3) *Multi-Level Analytics* generating classroom, school, and system dashboards, and (4) *Decision Support* offering contextualized feedback aligned with teacher, principal, and manager information needs. This model operationalizes DIDM and teacher-centered analytics principles into a unified workflow.

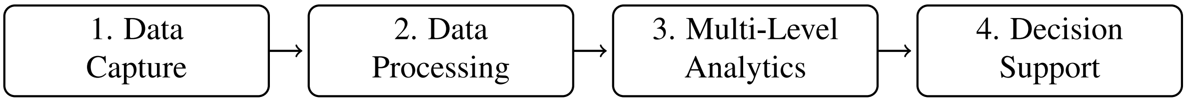

Figure 1. Conceptual Framework Diagram: from data capture through decision support across multiple stakeholder levels.

3. System Design and Implementation

3.1. Design Principles

Following past research, we guided Nino's development with a set of principles that reflect both the practical needs of teachers and the limitations and opportunities of the Brazilian school context [Guerino et al. 2024, Rodrigues et al. 2023]. The application proposes a technological solution that assists in managing assessments and automates the correction of answer sheets, promoting agility in the assessment process and facilitating pedagogical decision-making. To this end, we adopted a *teacher-centered design approach*, ensuring that design decisions were based on the real demands of teachers. From the prototyping phase to the case study, the focus was on understanding teachers' routines, limitations, and expectations to integrate the application into the school's daily routine.

Another principle was the creation of a *user-friendly interface* with an emphasis on simplicity, clarity, and ease of use. Given that many teachers still face challenges in using digital technologies [Isotani et al. 2023], we carefully designed the user experience to minimize the learning curve, enabling the application to be used intuitively, even by users unfamiliar with digital tools. Additionally, we designed Nino as a *lightweight application*, aiming for compatibility with a wide range of mobile devices, particularly those with lower processing power. This technical choice seeks to ensure accessibility and viability of use in school contexts with limited infrastructure and unstable connectivity.

Finally, we designed the application to facilitate the *rapid processing of collected data*, particularly in recognizing answer sheets, thereby supporting pedagogical decisions quickly. By making assessment results available instantly, Nino enables teachers to obtain a clear overview of student performance, identify specific areas of difficulty, and adjust their teaching strategies accordingly [Veloso et al. 2023].

3.2. System Architecture

Overall, Nino is based on a modular and scalable architecture composed of three main components: mobile front-end, back-end server, and AI system. The *mobile front-end*, developed using React Native, enables educational staff to capture images of completed answer sheets, visualize assessment status, and verify performance dashboards. It provides real-time feedback, guidance on the answer sheet collection, and submission controls to ensure data quality and usability in varied classroom settings. The *back-end server*, implemented with Node.js, manages user authentication, assessment data, session control, and request routing. It also handles communication between the front end and the AI system via secure API calls, ensuring efficient and reliable data flow. The *AI system* for recognizing multiple-choice answer sheets was developed in Python, leveraging the OpenCV library, and exposed through a RESTful API. Its primary goal is to automate the identification of filled-in responses on handwritten answer sheets, aiming for a performance-efficient and accurate application.

The AI system might be summarized as follows: To reduce computational complexity, the processing pipeline begins with image normalization, including orientation correction and grayscale conversion. Second, noise reduction techniques like Gaussian filters enhance visual clarity while preserving key structural features such as borders and markings. Third, the system identifies the predefined circular response areas using shape detection techniques. For this, edge detection and contour analysis allow for the precise localization of the document boundaries, which are then adjusted through perspective correction to simulate a flat, scanned image. Fourth, once the document is rectified, it is segmented vertically to isolate the student response regions. These regions are analyzed to detect filled circles, using intensity thresholds to differentiate between marked and unmarked options. Lastly, the selected responses are ordered and output in a structured format for integration with Nino back-end.

3.3. User Interface and Workflow

Before starting implementation, we prototyped all user interfaces of Nino to test the prototypes (described in Section 4). Figure 2 illustrates the primary interfaces of the application.

Figure 2a is the application's home screen, viewed from the teacher's login. In this screen, teachers view and manage all assessments linked to them. On the screen, teachers can verify the information for each assessment, including the school name, class, discipline, start and end dates, and the number of answer sheets recorded so far. There is also the functionality of identifying whether an evaluation is open or closed. An open assessment allows new answer sheet photos, while closed ones do not.

Figure 2b illustrates the registration screen of an answer sheet. As mentioned, the application recognizes the answers from a photo taken by a teacher of a student's answer sheet. We provided lighting and framing guidelines to ensure the photo achieved the highest possible quality.

Figure 3c shows how the feedback is displayed after the teacher takes a picture of the answer sheet. This screen displays the information on the number of answer sheets photographed for assessment and the student's name and school. Additionally, from this screen, the teacher can record another student's answer sheet (Figure 2b), access the report (Figure 2d), or report a system error.

Figure 4d displays the instant report of the results obtained for a particular assessment, separated by evaluated subject. For each photo taken, the application automatically registers the database. It classifies the student into: adequate (approved on more than 75% of questions), intermediate (approved between 50% and 75%), critical (approved between 25% and 50%) or very critical (approved less than 25%). The teacher may also export the full report, which contains information by question and student.

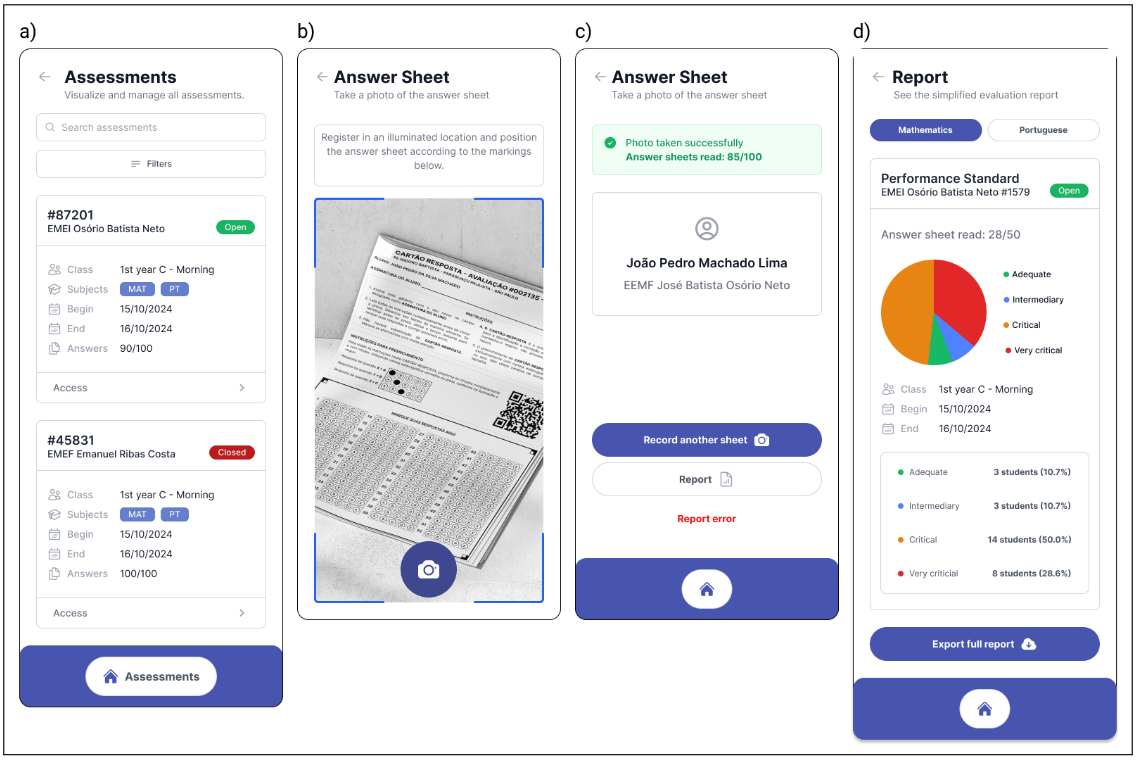

Figure 2. Main interfaces of Nino's prototype, where all assessment/student information presented is fictitious.

4. Study 1 - Prototype Assessment with Teachers

4.1. Method

We conducted a moderate remote usability test with five primary school teachers selected through convenience sampling to evaluate the prototypes and Nino's initial idea [Barbosa et al. 2021, Wohlin et al. 2012]. This study aimed to verify teachers' perceptions of Nino application prototype interface, which was created to assist in correcting evaluation templates.

We invited teachers to enter a Google Meet room individually to participate in the study. Before starting, the moderator explained the study's goal and ethical care and presented the Informed Consent Form, which was read and signed by all participants. Then, we invited the participants to perform the following tasks in the prototype through the Figma tool: i) log in to the application; ii) view the evaluations on the homepage; iii) filter evaluations; iv) take a picture of a student's feedback; v) check the report generated from the class; vi) export a complete report. Since the artifact to be evaluated was a prototype, not the developed application, we explained to the participants that the interactions were simulations for testing purposes, as recommended in the user-centered literature at early development stages [Barbosa et al. 2021].

Upon completing the tasks, participants answered a demographic questionnaire. Additionally, we conducted a semi-structured interview (see Table 1) to deepen further our qualitative understanding of the participants' perceptions of the designed prototype [Barbosa et al. 2021]. For qualitative data analysis, we transcribed the audio of the interviews. We then used thematic analysis with a deductive approach

[Clarke and Braun 2017], in which codes were categorized into three predefined groups: i) improvement of existing features, ii) possible new features, and iii) future features.

Table 1. Questions used in the semi-structured interview.

#	Question
1	How was your overall experience when viewing the Nino prototype?
2	Was there any part of the app that you found confusing or complicated to use? If so, which one(s)?
3	Which features did you find most useful?
4	Have you found any functionality that is missing in Nino?
5	How would you describe Nino's interface design?
6	Did you feel comfortable browsing the app? If not, please explain.
7	Would you have any suggestions to improve Nino?
8	What is the probability of using Nino in your template routine?

4.2. Results

Regarding the participants' demographics, three identified as female and two as male. Two participants had 6 to 10 years of experience as teachers, one between 11 and 15 years, and the other two had more than 15 years of experience. Regarding their experience with the use of applications in general, two responded that they use only basic applications (messages, social networks, and others); one answered that they use basic and advanced applications (digital banks, digital signatures, and others); the others two responded that they use basic and advanced applications and, in addition, applications that assist in work as a teacher (Padlet, Canva, Inshort, and Mentimeter applications were mentioned).

Regarding the results of the interviews, the first classification category of the thematic analysis focused on *improvements to existing features*. In this sense, participants highlighted the need to improve the evaluation screen interface, including a lighter identifier for each assessment and displaying information such as the start date, final collection date, and evaluation type (e.g., bimonthly, stage). Improvements were also suggested in identifying students after reading the feedback, such as adding photos, grades, and class, which would facilitate teacher recognition. There have been requests for adjustments to report generation, including the inclusion of complication comments, the classification of students by performance groups, and the need to improve visual accessibility through increased contrast of icons and buttons.

Regarding the *possible new features*, participants suggested unpublished resources that could enrich the system. The importance of allowing the inclusion and treatment of students with disabilities and foreigners, with the collection of specific information (such as reports) and appropriate signaling in the reports. The possibility of inserting comments into reviews, visible only to the Secretariat, was also highlighted as relevant. Other suggested features include the differentiation of access by CPF to allow users with different functions (such as secretaries and teachers) to act in the same assessment with different permissions and the shuffling of tests, maintaining the coherence between feedback and version of the applied test. It also mentioned the reopening of evaluations for collecting absent students.

The latest category, called *future features*, included ideas that were still in the early stages or whose viability required further study. Participants suggested, for example, creating a web version for more robust reports and interactive dashboards and the possibility of comparing evaluations over time for analysis of class performance evolution. It also discussed the customization of the proficiency scale, allowing schools to define their evaluation criteria through spreadsheets and improving the initial navigation experience, such as including explanatory texts in icons and buttons.

5. Study 2 - Case Study in a Real Environment

5.1. Method

This study investigated how integrating the Nino influenced the assessment of multiple-choice tests in a real-world school environment. We conducted a case study [Barbosa et al. 2021] involving two elementary school teachers. Considering the novelty of Nino and that this was one of the first times used in an authentic learning environment, we opted for a small-scale study, aiming for initial insights into how the App would contribute to the school's staff assessment-related activities.

This study included four main steps. *First*, we recruited two teachers through convenience sampling [Wohlin et al. 2012]. Both teachers were male, between 20 and 25 years old, and had over three years of teaching experience in public middle schools. Moreover, both of them had previously engaged in educational technology initiatives. One teacher specialized in Mathematics, while the other taught the Portuguese language. Both agreed to integrate Nino into assessments to be applied across a total of 6 classes, consisting of approximately 20 students each, spanning 6th and 9th grades in an urban school district. Nevertheless, these assessments were not mandatory, so not all students completed them.

Second, we co-designed the assessments with the participating teachers. The goal of this step was twofold: i) to ensure that the assessments were pedagogically aligned with each teacher's curriculum and classroom practices and ii) to format the assessments in a manner compatible with the design requirements of the Nino, particularly the visual and spatial constraints necessary for automated recognition expected by the App (see Figure 2). Given that both teachers voluntarily agreed to participate in the case study, this step was essential to ensure the feasibility of collecting data in an authentic learning environment.

Third, the teachers used Nino to collect the assessments. The students completed 42 answer sheets and assessed with Nino. Importantly, no student-identifying information was gathered. The app captured only images of completed answer sheets, which were anonymized at the point of capture and stored without metadata linking them to individual students. Additionally, no individual performance data were collected or analyzed. As such, this study did not involve human subject data, so no institutional review board approval was required.

Lastly, we investigated how the app contributed to the assessment process in the investigated learning settings. Here, we followed a methodology similar to Study 1 regarding the tasks accomplished with the App (see Section 4). The difference is that this study was enhanced by using a functional version of the Nino to collect and analyze log data, aiming to understand the app's contribution to assessment time and how its AI system performed under authentic classroom conditions. Particularly, our focus was on reliability and robustness to classroom-specific variations (e.g., lighting, handwriting).

5.2. Results

Regarding the performance of the app's AI system, the log data revealed promising insights. Out of the 42 answer sheets, the AI system performed no error in 30 of them (71%). Those answer sheets included 356 questions to be marked by the students, where the app correctly recognized and graded 315 of them (88%). Importantly, by further investigating answer sheets in which the AI system failed to acknowledge any item, we noted that many of those did not follow the marking instructions. For instance, some cases included marking the answer sheet with pink or orange-colored pens and failing to fulfill the marking area completely. If we disregard those cases, the AI system's performance goes up to 93% and 79% in terms of recognizing marked questions and successfully grading the complete answer sheet. Moreover, the log data indicated that teachers spent an average of 20 seconds to collect each answer sheet. Overall, these findings suggest the AI system from Nino is robust to most authentic classroom scenarios, despite it presents room for improvements in handling unexpected patterns likely to happen in practice.

6. Discussion

Our finding provides promising evidence of the Nino potential. The controlled prototype evaluation with five in-service teachers (Study 1) revealed positive perceptions of the Nino's core workflow (photo-based capture, automated processing, and dashboard review). Participants highlighted the ease of capturing answer sheets (20 seconds per sheet in Study 2 logs) and appreciated the concept of rapid, in-situ feedback. However, qualitative feedback pinpointed specific refinements: clearer identifiers for assessments (e.g., display of dates and types), improved student-recognition mechanisms (e.g., photos or class labels post-processing), and enhanced report visual accessibility (contrast, icons) and customization (performance groupings, comments). These findings align with prior work emphasizing minimal friction and contextual relevance in teacher-centered analytics tools [Rodrigues et al. 2023, Alfredo et al. 2024]. While the photo-based pipeline concept is intuitive, careful attention to interface details and integration into existing routines is crucial for acceptance [Schildkamp 2019, Xia et al. 2022].

In the two-week deployment with two teachers and 42 anonymized answer sheets, the system demonstrated technical robustness under authentic classroom conditions: an overall sheet-level accuracy of 71% (30/42 sheets without any recognition error) and item-level recognition accuracy of 88%, rising to approximately 93% when excluding sheets with marking deviations (e.g., non-standard pen colors or incomplete fills). These performance metrics indicate that the OCR and image-processing pipeline can handle common variations in lighting and handwriting, consistent with findings in similar photo-based assessment research [Freitas et al. 2022, Chevtchenko et al. 2024]. From a pedagogical perspective, teachers reported that rapid access to aggregated class results facilitated reflection on item-level patterns (e.g., identifying misconceived topics) and spurred informal discussions about instructional adjustments. Although the sample and duration were limited, these anecdotal insights suggest the system can meaningfully contribute to formative assessment cycles [Shute 2008].

Furthermore, the three-tier feedback model (classroom, school, system) holds promise: even in this small-scale study, teachers indicated potential interest in sharing summary data with peers or school leaders to inform collaborative planning. This resonates with data-informed decision-making frameworks distinguishing micro (classroom) and meso (school) levels [Wang 2021]. However, actual uptake at broader levels would require institutional buy-in, clear protocols for data sharing, and alignment with existing assessment policies.

Consistent with teacher-centered learning analytics principles, co-design activities (e.g., aligning assessment formats to app constraints) were instrumental in ensuring feasibility. Future deployments should embed iterative co-design workshops to refine UI elements (e.g., onboarding tutorials, icons with explanatory tooltips) and feature sets (e.g., proficiency scale customization, support for students with disabilities or multilingual backgrounds). Low-resource settings particularly benefit from lightweight, mobile-first approaches but demand attention to devise heterogeneity, intermittent connectivity, and minimal training overhead [Isotani et al. 2023, Guerino et al. 2024].

The observed recognition errors often stemmed from deviations in marking (e.g., colored pens, incomplete fills). To mitigate this, the system should incorporate in-app guidance (e.g., overlay templates or real-time visual guidance when capturing) and possibly lightweight pre-capture validation (e.g., immediate feedback if a sheet appears misaligned or poorly marked). In contexts where teachers may be less familiar with strict marking protocols, brief training materials or integrated tutorials can improve data quality. Moreover, adaptive algorithms (e.g., threshold adjustment or confidence-based prompts) could flag uncertain cases for manual review, balancing automation with teacher agency.

For sustained adoption beyond prototype stages, the app must integrate with existing school workflows and information systems (e.g., LMS, SIS). Simplified export formats and APIs can enable seamless sharing of aggregated reports with school leaders and district managers. Additionally, embedding the app within broader professional development initiatives, where coaches and instructional coordinators guide teachers in interpreting analytics, can reinforce evidence-based practices.

7. Conclusion

This paper presented the design, implementation, and evaluation of a mobile-first, teacher-centered application (Nino) for photo-based assessment in under-resourced contexts. Through two empirical studies, a controlled prototype usability test with in-service teachers and a short-term real-world deployment, we demonstrated that the app's core workflow (capturing answer sheets via smartphone, automated OCR processing, and multi-level feedback dashboards) aligns well with educators' routines and can operate reliably under typical classroom conditions. Usability feedback highlighted both the concept's promise (rapid capture and near-real-time insights) and areas for refinement (interface clarity, guidance on marking conventions, and customization of reports). The case study further suggested that even a brief deployment can prompt reflective practices and signal the potential value of sharing aggregated data with school leaders for evidence-based decision-making.

Despite promising initial results, the studies were limited by small samples and short duration, precluding definitive claims about long-term adoption or direct impacts on student learning. Future work should involve larger-scale, longitudinal deployments across diverse settings, integration with existing school information systems, enhanced analytics features, and rigorous evaluation of educational outcomes. Attention to privacy, data governance, and co-design with stakeholders remains critical. Overall, this work contributes design guidelines for lightweight, scalable assessment technologies that foreground teacher agency and pave the way for further research into how rapid, low-infrastructure feedback can support equitable, data-informed pedagogical practices.

References

Alfredo, R. D., Echeverria, V., Zhao, L., Lawrence, L., Fan, J. X., Yan, L., Li, X., Swiecki, Z., Gašević, D., and Martinez-Maldonado, R. (2024). Designing a human-centred learning analytics dashboard in-use. *Journal of Learning Analytics*, 11(3):62–81.

Barbosa, S. D. J., Silva, B. d., Silveira, M. S., Gasparini, I., Darin, T., and Barbosa, G. D. J. (2021). Interação humano-computador e experiência do usuario. *Auto publicação*.

Bichsel, J. (2012). Analytics in higher education: Benefits, barriers, progress, and recommendations. *ECAR Research Study*.

Chevtchenko, S. F., Carvalho, R., Rodrigues, L., Souza, E., Rosa, D., Cordeiro, F., Pereira, C., Vieira, T., Marinho, M., Dermeval, D., et al. (2024). An optimization approach for elementary school handwritten mathematical expression recognition. In *International Conference on Artificial Intelligence in Education*, pages 234–241. Springer.

Clarke, V. and Braun, V. (2017). Thematic analysis. *The journal of positive psychology*, 12(3):297–298.

Datnow, A. and Park, V. (2014). Data-driven leadership. *Jossey-Bass Education Series*.

de La Hoz-Ruiz, J., Khalil, M., Domingo Segovia, J., and Liu, Q. (2024). Learning analytics for enhanced professional capital development: a systematic review. *Frontiers in Psychology*, 15.

Dui, L. G., Lunardini, F., and Ferrante, S. (2019). A tablet-based application to study the speed-accuracy tradeoff in handwriting throughout lifespan. In *2019 E-Health and Bioengineering Conference (EHB)*, pages 1–4.

Fachola, C., Tornaría, A., Bermolen, P., Capdehourat, G., Etcheverry, L., and Fariello, M. I. (2023). Federated learning for data analytics in education. *Data*, 8(2).

Freitas, E., Batista, H. H., Barbosa, G. A., Wenceslau, M., Portela, C., Isotani, S., Cordeiro, T., Bittencourt, I. I., Yasojima, K., Sobrinho, A., et al. (2022). Learning analytics desconectada: Um estudo de caso em análise de produçoes textuais. In *Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil (WAPLA)*, pages 40–49. SBC.

Guerino, G. C., Rodrigues, L. A. L., Oliveira, L. S., Marinho, M., da Silva, T. E. V., Amorim, L., Dermeval, D., da Penha, R. S., Bittencourt, I. I., and Isotani, S. (2024). We see you: Understanding math teachers from brazilian public schools to design equitable educational technology. *Revista Brasileira de Informática na Educação*, 32:336–358.

Isotani, S., Bittencourt, I. I., Challco, G. C., Dermeval, D., and Mello, R. F. (2023). Aied unplugged: Leapfrogging the digital divide to reach the underserved. In Wang, N., Rebolledo-Mendez, G., Dimitrova, V., Matsuda, N., and Santos, O. C., editors, *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky*, pages 772–779, Cham. Springer Nature Switzerland.

Nicol, D. J. and Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2):199–218.

Rasiq, G. R. I., Al Sefat, A., and Hasnain, M. F. (2019). Mobile based mcq answer sheet analysis and evaluation application. In *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, pages 144–147.

Rodrigues, L., Guerino, G., Challco, G. C., Veloso, T. E., Oliveira, L., da Penha, R. S., Melo, R. F., Vieira, T., Marinho, M., Macario, V., et al. (2023). Teacher-centered intelligent tutoring systems: Design considerations from brazilian, public school teachers. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1419–1430. SBC.

Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational Research*, 61(3):257–273.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1):153–189.

Topali, P., Ortega-Arranz, A., Rodríguez-Triana, M. J., Er, E., Khalil, M., and Açkapınar, G. (2024). Designing human-centered learning analytics and artificial intelligence in education solutions: a systematic literature review. *Behaviour & Information Technology*, pages 1–28.

Veloso, T. E., Chalco Challco, G., Rogrigues, L., Versuti, F. M., Sena da Penha, R., Silva Oliveira, L., Corredato Guerino, G., Cavalcanti de Amorim, L. F., Monteiro Marinho, M. L., Macario, V., Dermeval, D., Bittencourt, I. I., and Isotani, S. (2023). Its unplugged: Leapfrogging the digital divide for teaching numeracy skills in underserved populations. In *Towards the Future of AI-augmented Human Tutoring in Math Learning 2023 - Proceedings of the Workshop on International Conference of Artificial Intelligence in Education co-located with The 24th International Conference on Artificial Intelligence in Education (AIED 2023)*. Springer.

Wang, Y. (2021). When artificial intelligence meets educational leaders' data-informed decision-making: A cautionary tale. *Studies in Educational Evaluation*, 69:100872. From Data-Driven to Data-informed Decision Making:Progress in the Field to Improve Educators and Education.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., Wesslén, A., et al. (2012). *Experimentation in software engineering*, volume 236. Springer.

Xia, Q., Chiu, T. K., Zhou, X., Chai, C. S., and Cheng, M. (2022). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, page 100118.