DOI: https://doi.org/10.21580/jieed.v6i1.31141

Journal of Integrated Elementary Education

eISSN 2776-1657 | pISSN 2828-223X | Vol. 6 No. 1 (2026) | pp. 46-64 | OPEN ACCESS

Supporting elementary teachers' pedagogical decision-making through machine learning–based behavioral diagnostics

Taqwa Nur Ibad[1]*, Istiningsih[2], Sumarsono[2], Alfiatus Safaah[3], Hartono[4], Ihya'Ulumuddin[3]

[1]Universitas Islam Zainul Hasan Genggong Probolinggo, Indonesia
[2]Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia
[3]Universitas Islam Syarifuddin, Lumajang, Indonesia
[4]Universitas Islam Negeri Kiai Haji Achmad Siddiq Jember, Indonesia

*Corresponding author's email: ibadyangsukses@gmail.com

Abstract

This study investigates how elementary school teachers interpret and use a machine–learning–based behavioral diagnostic system to support pedagogical decision-making regarding students' non-cognitive competencies. Employing a descriptive-analytical approach grounded in the principles of human-centered artificial intelligence, data were gathered through questionnaires, semi-structured interviews, and document analysis. The findings indicate that teachers were able to clearly interpret the diagnostic outputs and deemed them relevant to their classroom practices. Additionally, there were high levels of perceived usefulness in identifying students' learning needs (M = 4.25) and supporting instructional planning (M = 4.22). Over 80% of teachers reported using the system to differentiate learning activities and adjust instructional strategies. Moreover, the diagnostic system served primarily as a decision-support tool, assisting teachers in validating their professional intuition, designing more responsive differentiation strategies, and engaging in deeper reflective practice regarding the development of students' non-cognitive skills. Importantly, teachers maintained full professional autonomy by selectively interpreting and contextualizing system recommendations based on their students' knowledge. These findings suggest that machine learning-based behavioral diagnostics can effectively enhance teachers' pedagogical reasoning when designed within a human-centered framework. Rather than replacing teacher judgment, the system improved teachers' awareness of students' motivation, engagement, self-regulation, and perseverance, thereby facilitating more informed and responsive instructional decision-making in elementary education.

ARTICLE HISTORY

Received:
2 February 2026
Accepted:
10 March 2026
Published:
31 March 2026

KEYWORDS

Machine Learning; Behavioral Insight Diagnostics; Non-Cognitive Competencies; Elementary School.

1. Introduction

Elementary educators are increasingly expected to address student development in an integrated manner, balancing academic achievement with non-cognitive competencies such as motivation, engagement, self-regulation, and perseverance—factors widely recognized as critical for long-term academic success (OECD, 2021; Zhou, 2023; Heikkinen et al., 2023). In everyday classrooms, however, teachers typically rely on informal observation and professional intuition to gauge these attributes. While experiential knowledge is invaluable, this approach is prone to subjectivity and inconsistency, particularly given the complexity of modern classrooms with their diverse learners and packed curricula (van der Lans, et al., 2022; Herppich, et al., 2023). Without systematic diagnostic support, subtle but important behavioral patterns often go unnoticed, leading to instructional responses that may miss their intended mark (Revuelta-Domínguez et al., 2022). This issue becomes especially urgent in primary education, where unacknowledged non-cognitive difficulties can develop into decreased academic participation, and teachers may inadvertently delay or misdirect their interventions.

Recent advances in artificial intelligence, particularly machine learning, which addresses how computers improve automatically through experience at the intersection of computer science, statistics, and data science (Cukurova et al., 2025)—offer new possibilities for augmenting teachers' interpretive capabilities rather than replacing their judgment (Holmes, et al., 2022; Ouyang & Jiao, 2021). Machine learning enables the transformation of informal teacher observations into structured, analyzable data by providing intuitive interfaces for recording behaviors such as task initiation, attention duration, help-seeking patterns, peer collaboration, and persistence on difficult tasks. More importantly, its algorithms can detect subtle psychological patterns that even experienced teachers might miss—identifying correlations between time of day and student engagement, or sequences of behaviors that precede disengagement.

The strategic role of AI lies in serving as a cognitive amplifier, making invisible patterns visible through visualizations like radar charts and trend lines, and enabling teachers to track psychological development longitudinally and intervene with precision (Benjamin Balraj, 2025). Systems designed with a human-centered philosophy emphasize transparency, pedagogical relevance, and teacher control, offering significant insights while concurrently preserving teachers' discretion (Ouyang & Jiao, 2021).

One emerging application combines machine learning with behavioral insights to build diagnostic profiles of students' non-cognitive learning behaviors by analyzing observable classroom actions. Early evidence suggests these profiles can help teachers see beyond surface behavior and recognize underlying needs (Khosravi, 2023). However, the effectiveness of such systems depends not on data precision but on teachers' ability to interpret results wisely—analytics become meaningful only when they enhance teacher sense-making (Wise & Jung, 2019).

Despite a surge of research on AI-based assessment and learning analytics, most studies to date have focused on system development, predictive accuracy, and technical validation, with recurring limitations such as small-scale evaluations, a lack of causal evidence linking predictions to pedagogical interventions, and insufficient attention to ethical considerations like bias (Cabral et al., 2025). Comparatively little work has examined how teachers actually interpret and use these diagnostic outputs to shape their instruction, especially in elementary settings (Crompton & Burke, 2018). The interpretation of results by instructors remains a critical blind spot; without it, the pedagogical value of AI-based diagnosis remains purely speculative.

Recent reviews have begun mapping this landscape, arguing that the real promise of AI lies in its ability to provide actionable analytics that teachers can use to make more informed, personalized instructional decisions. These reviews also stress that ethics and sustained professional development are non-negotiable prerequisites for meaningful classroom use (Akintolu, 2025). Furthermore, any integration of AI must be governed by a human-centered framework that places pedagogical values, fairness, and teacher autonomy at the core; otherwise, risks such as algorithmic bias and opaque decision-making may undermine teacher professionalism (Efendi et al., 2025). The most productive relationship between teachers and AI is one of complementarity rather than competition or substitution, where AI handles routine tasks while teachers focus on higher-order pedagogical work, such as critical thinking, emotional support, and value-based assessment (Benjamin Balraj, 2025).

This study fills the identified gap by exploring how elementary teachers at the Elementary School of Nurul Islam Kota Lumajang, Indonesia, used a machine-learning-based behavioral diagnostic system to inform their decisions about students' non-cognitive skills. Notably, we do not assess the technical performance of the system but focus on the teachers themselves: how they perceived the diagnostic information, judged its credibility, and translated it into instructional actions.

Two guiding theoretical frameworks inform our investigation: human-centered AI, the principle that technology should augment, not replace, human expertise (Shneiderman, 2022); and pedagogical reasoning, which involves how teachers transform their knowledge about students into effective classroom practices (Shulman, 1987; Datnow, 2021). As Ibad et al. (2024) remind us, the benefits of artificial intelligence depend heavily on how users manage it ethically and wisely; machine learning systems will ultimately prove useful only if teachers remain in control. By foregrounding the position of teachers, this study contributes to the broader discussion on the ethics and pedagogy of AI in schools, arguing for a vision of technology that supports rather than undermines teacher agency.

2. Methods

This study employed a descriptive-analytical research design within a human-centered educational technology framework to examine how teachers interpret and utilize outputs from a machine learning–based behavioral diagnostic system to inform pedagogical decisions related to students' non-cognitive competencies (Ouyang & Jiao, 2021). The research was conducted at Madrasah Ibtidaiyah Nurul Islam, Lumajang, East Java, Indonesia, a private Islamic elementary school that had implemented a machine learning–based behavioral diagnostic system for approximately one semester prior to data collection.

Purposive sampling was used to select participants who had direct experience using the system. The participants consisted of six classroom teachers, one school principal, and twenty-eight third-grade students whose behavioral data were recorded in the diagnostic system. The teacher participants were S.M. (female, 38 years old, 15 years of teaching experience), A.F. (male, 42 years old, 18 years), N.H. (female, 35 years, 10 years), D.L. (female, 40 years, 16 years), R.H. (male, 37 years, 12 years), and S.A. (female, 44 years, 20 years). All teachers had used the system for at least six months and had received basic training on interpreting the outputs. The school principal, D.W.P. (female, 48 years old, 25 years of teaching experience), participated as an institutional informant but was not a direct user of the system. The principal and the Grade III D teacher agreed to be identified by name, while the other teacher participants are referred to by pseudonyms to protect their privacy.

The diagnostic system was based on a supervised machine learning model (Random Forest classifier) trained on approximately 2,500 labeled behavioral instances from elementary school students. Teachers recorded daily observations of five behavioral indicators: task initiation, attention duration, help-seeking behavior, peer collaboration, and persistence on challenging tasks. The system generated a bi-weekly non-cognitive profile visualized through radar charts across four dimensions: motivation, engagement, self-regulation, and perseverance, along with trend analyses and pedagogical suggestions.

The main features of the diagnostic system used in this study are summarized in Table 1. The system allows teachers to record observable student behaviors, visualize non-cognitive learning profiles, monitor behavioral trends over time, and receive pedagogical suggestions based on detected patterns. Importantly, the system was designed as a decision-support tool rather than an automated decision-maker. Teachers are explicitly encouraged to use their professional judgment when interpreting and applying the system's recommendations.

Table 1.
Key Features of the System

Feature	Description
Daily Behavior Logging	Teachers record five observable behaviors per student using dropdown menus: task initiation, attention duration, help-seeking, peer collaboration, and persistence on difficult tasks.
Non-Cognitive Profile Visualization	Every two weeks, the system generates a radar chart for each student, displaying scores across four dimensions: motivation, learning engagement, self-regulation, and perseverance.
Trend Analysis	A line graph shows changes in each dimension over the past 4–6 weeks, helping teachers identify patterns (e.g., declining persistence, improving engagement).
Pedagogical Suggestions	Based on pattern recognition, the system offers suggestions framed as possibilities, not prescriptions. Example: "This student shows declining persistence. Consider reducing task complexity or providing more frequent encouragement."
Decision-Support Design	All outputs are advisory. Teachers are explicitly told that their professional judgment should override any system recommendation when they deem it appropriate.

To capture teachers' experiences with the system, three complementary instruments were employed. A 20-item questionnaire measured teachers' understanding of the diagnostic outputs, their perceived usefulness, and the relevance of the information to instructional decision-making using a five-point Likert scale. Semi-structured interviews explored how teachers interpreted diagnostic information and how it influenced lesson planning, differentiation, and reflective practice. In addition, document analysis was conducted on 24 lesson plans, 28 student progress notes, and teachers' reflective journals.

Questionnaire data were analyzed using descriptive statistics, including means and percentages. Interview transcripts and documents were analyzed using thematic analysis following the six-phase procedure proposed by Braun and Clarke (2021). To enhance trustworthiness, several strategies were applied, including data triangulation, member checking, and peer debriefing. Ethical clearance was obtained from the school and the university, and all student data were anonymized.

3. Results

3.1 Teachers' Use of the Machine Learning–Based Diagnostic System

The machine learning–based behavioral insight diagnostic system was implemented to support teachers in identifying students' learning profiles and informing instructional decision-making. The system processes student behavioral learning data and presents diagnostic outputs through an interactive dashboard that visualizes key dimensions of student learning behavior. The main interface of the system is presented in Figure 1.

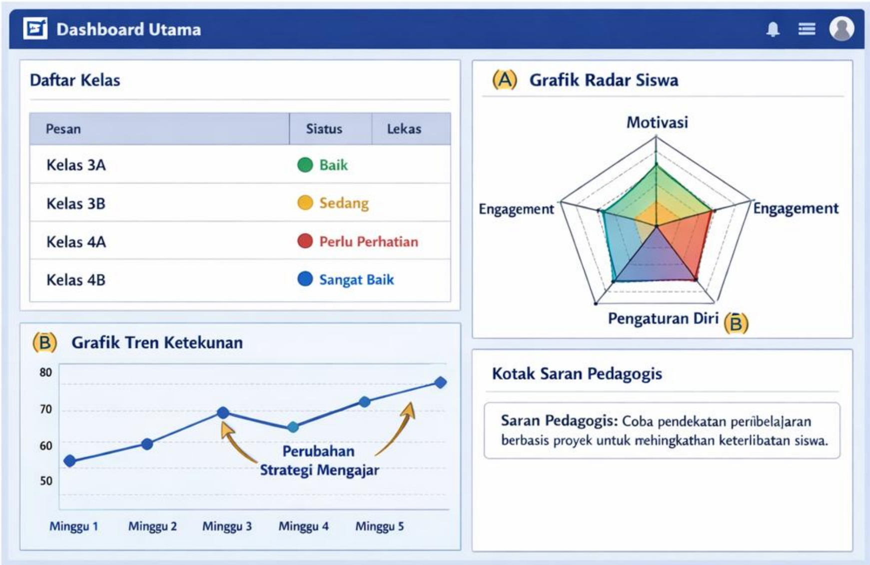

Fig. 1. Main dashboard of the ML-based behavioral diagnostic system.

The dashboard summarizes behavioral indicators derived from machine learning analysis of student activity data. Teachers can view information related to student engagement, persistence in completing tasks, collaboration patterns, and indicators associated with learning strategies. By presenting these indicators in a visual format, the dashboard enables teachers to quickly obtain an overview of classroom learning dynamics and identify students who may require additional instructional support.

In addition to the dashboard overview, the system provides a visualization of individual student behavioral profiles through radar chart representations.

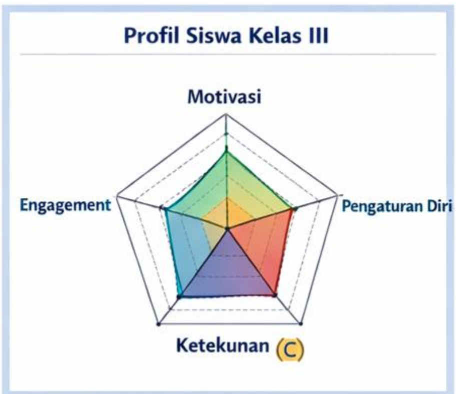

Fig. 2. Student behavioral profile visualization using a radar chart.

Figure 2 illustrates how the system represents student behavioral indicators across multiple learning dimensions. Each axis of the radar chart corresponds to a specific behavioral indicator, such as engagement, persistence, collaboration, and self-regulated learning. The radar visualization allows teachers to easily compare strengths and weaknesses across these dimensions. Through this graphical representation, teachers can quickly identify areas in which students demonstrate strong learning behavior as well as areas that require additional support. To support continuous monitoring of students' learning development, the system also provides a trend analysis feature.

Fig. 3. Learning behavior trend analysis interface.

As shown in Figure 3, the trend analysis interface displays changes in student behavioral indicators over time. Teachers can observe whether students' engagement, persistence, and learning strategies improve, remain stable, or decline during the learning process. This feature helps teachers evaluate student progress and identify patterns that may not be immediately visible through classroom observation alone.

System usage records indicate that teachers actively accessed the dashboard to review diagnostic outputs. As presented in Table 2, teachers primarily used the system to monitor student learning behavior, identify potential learning difficulties, and obtain additional information to support instructional planning.

Table 2.
Patterns of use by grade level.

Grade Level	Frequency of Use	Primary Purpose
I (Ibu Nurul)	Weekly	Identifying students needing extra support
II (Ibu Dewi)	Twice weekly	Monitoring progress after interventions
III D (Ibu Sri)	Three times weekly	Differentiating instruction: preparing for parent meetings
IV (Bapak Rudi)	Weekly	Grouping students for collaborative projects
V (Ibu Siti)	Weekly	Tracking engagement trends
VI (Bapak Ahmad)	Bi-weekly	Preparing students for transition to secondary school

Interview data further illustrate how teachers used the system in their classroom practice. One teacher explained: *"The dashboard helped me quickly see which students were struggling and which ones were more independent in their learning."* Another teacher noted: *"Before using the system, I mostly relied on observation. Now the data helps confirm what I see in class."* These findings indicate that the diagnostic system served as a practical analytical tool to support teachers in monitoring classroom learning dynamics.

3.2 Teachers' Understanding of Diagnostic Outputs

An important aspect of the system implementation was teachers' ability to interpret the diagnostic outputs generated by the machine learning model. Table 3 presents the quantitative results related to teachers' perceptions of the clarity and usefulness of the diagnostic information provided by the system.

Table 3.
Teachers' understanding of diagnostic outputs.

Indicator	Mean	SD	Interpretation
Clarity of behavioral indicators	4.21	0.48	High
Ease of interpreting non-cognitive profiles	4.08	0.58	High
Alignment with classroom observations	4.15	0.53	High
Confidence in understanding output summaries	4.02	0.61	High

Overall, teachers reported that the diagnostic outputs were relatively easy to interpret. The visual representations of behavioral indicators enabled teachers to understand the results without requiring advanced technical knowledge of machine learning algorithms.

Qualitative findings indicate that the system often validated teachers' professional intuition. Rather than presenting entirely new information, the diagnostic outputs helped teachers clarify observations they had already noticed in classroom practice. For example, Ibu Sri Mu'awana described how the system helped her better understand one of her students:

> *"I have a student—Abrisam—who is actually very bright but rarely finishes his work. I used to think he was just lazy. But the system showed a pattern: he gives up quickly when the task is too long. It wasn't laziness; it was task-length anxiety. So I started breaking the tasks into smaller chunks, and now he completes them."*

Similarly, Ibu Nurul Hidayati explained how the radar chart visualization helped identify specific behavioral learning needs: *"Sometimes you know a child is struggling, but you can't put your finger on why. The radar chart made it visual—I could see that his self-regulation was low, even though his motivation was high."*

These qualitative insights are supported by the quantitative results. Teachers rated the alignment between system outputs and their classroom observations relatively high (M = 4.15). In addition, teachers reported increased awareness of students' non-cognitive learning needs (M = 4.27), suggesting that the system helped transform implicit classroom observations into clearer diagnostic insights.

3.3 Perceived Pedagogical Usefulness

Beyond understanding the diagnostic outputs, teachers also reported using the information to guide instructional decisions. Tables 4 and 5 present quantitative findings on how teachers applied the diagnostic insights in their teaching practices.

Table 4.
Perceived pedagogical usefulness of diagnostic information.

Aspect of Usefulness	Mean	SD	Interpretation
Identifying students' learning needs	4.25	0.45	Very High
Assisting lesson differentiation	4.18	0.54	High
Supporting classroom management decisions	3.97	0.62	Moderate–High
Contributing to instructional planning	4.22	0.49	Very High

The results indicate that teachers perceived the system as useful for supporting instructional planning and lesson differentiation. The behavioral profiles provided by the system allowed teachers to better understand variations in student learning behavior.

Table 5.
Forms of pedagogical decision-making supported by diagnostic outputs.

Decision Area	Number of Teachers	Percentage
Differentiation of learning activities	5	83.3%
Adjustment of instructional strategies	5	83.3%
Grouping students based on learning profiles	4	66.7%
Selection of learning resources	4	66.7%
Monitoring students' non-cognitive progress	5	83.3%

The findings show that a large proportion of teachers used the diagnostic outputs to support differentiated instruction. Teachers reported adjusting learning tasks, grouping strategies, and levels of instructional guidance based on the behavioral learning profiles generated by the system.

Qualitative evidence illustrates how these adjustments were implemented in classroom practice. For example, Ibu Dewi Lestari explained: *"I have two students with similar math scores but completely different non-cognitive profiles. One is highly motivated but gets anxious, while the other is confident but easily bored. The system helped me see that they need different kinds of challenges."* Another teacher, Ibu Sri Mu'awana, described how the system supported more flexible student grouping: *"The profiles helped me group students not just by* *ability, but by how they learn. Some students need frequent feedback, while others need more autonomy."*

Although teachers found the system helpful, they did not rely solely on the automated recommendations. Instead, they combined the diagnostic outputs with their professional judgment. As one teacher explained, *"The system said one of my students needed easier tasks. But I know this child well. Instead of simplifying the task, I gave more guidance."* These findings suggest that teachers used the diagnostic system as a supportive reference while maintaining their professional autonomy in instructional decision-making.

3.4 Instructional Decisions Informed by Diagnostic Outputs

The implementation of the diagnostic system also influenced teachers' reflective practice. Table 6 presents the quantitative results related to teachers' perceptions of how the system supported reflection on student learning behavior.

Table 6.
Teachers' reflections on professional use of diagnostic information.

Reflective Dimension	Mean	SD	Interpretation
Increased awareness of students' non-cognitive needs	4.27	0.46	Very High
Support for reflective teaching practices	4.19	0.54	High
Confidence in data-informed decision-making	4.05	0.62	High
Perceived professional relevance	4.23	0.50	Very High

The results indicate that teachers perceived substantial reflective benefits from using the diagnostic system. The highest mean score was reported for increased awareness of students' non-cognitive learning needs (M = 4.27), followed by support for reflective teaching practices (M = 4.19).

Qualitative findings further illustrate how the system encouraged teachers to reconsider their assumptions about student learning behavior. As Ibu Sri Mu'awana explained: *"The system is like a mirror. Sometimes I thought I was already differentiating my instruction, but the profiles showed that I was still treating everyone the same."* The school principal also observed changes in teachers' reflective practices across the school: *"We encourage teachers to use the data for reflection rather than evaluation. The important thing is that teachers have a pedagogical reason behind every instructional decision."*

Document analysis of lesson plans and student progress notes provided additional evidence of this reflective process. Among the 28 student progress notes analyzed, 19 (68%) contained explicit references to non-cognitive behavioral indicators such as persistence, engagement, and willingness to attempt challenging tasks. These records suggest that diagnostic categories had gradually become part of teachers' everyday language when describing student development. Overall, these findings indicate that the machine learning–based diagnostic system supported teachers not only in making instructional decisions but also in developing deeper reflective awareness of student learning behavior.

The integration of quantitative and qualitative findings across the four major themes is summarized in Table 7. The table illustrates how statistical results are supported and enriched by qualitative evidence from teacher interviews and document analysis. This integration demonstrates the consistency between teachers' reported perceptions of the diagnostic system and their actual experiences in classroom practice.

Table 7.
Integration of qualitative and quantitative evidence across four themes.

Theme	Qualitative Evidence	Quantitative Support
Validation of intuition	Teachers described the system as making vague feelings concrete	Alignment with observations: 4.15; Increased awareness: 4.27
Selective autonomy	Teachers overrode system recommendations based on personal knowledge	Confidence in understanding: 4.02; Confidence in decisions: 4.05
Responsive differentiation	Teachers used profiles to tailor instruction flexibly	Differentiation usefulness: 4.18; 83% used for differentiation
Reflective awareness	Teachers questioned their assumptions and tracked non-cognitive growth	Increased awareness: 4.27; Reflective support: 4.19

4. Discussion

This study was guided not by the question of whether machine-learning diagnostics can accurately classify student behavior but by a more fundamental inquiry: whether and how teachers can use such classifications to improve their pedagogical decisions. Grounded in the specific context of MI Nurul Islam Kota Lumajang, our findings offer several insights addressing this core question.

3.5 Diagnostic congruence and the construction of trust

A recurring pattern in our data was the importance of congruence between system outputs and teachers' own observations. When diagnostic profiles aligned with what teachers had already sensed, trust in the system increased. Ibu Sri Mu'awana's account of her student Abrisam illustrates this dynamic: the system did not reveal an entirely unknown problem but crystallized a diffuse impression into a recognizable pattern. This congruence was not merely confirmation; teachers valued the system precisely because it helped them articulate, organize, and track what they had previously held as vague intuitions.

This finding resonates with earlier work on teacher data use. Bolhuis, Voogt, and Schildkamp (2019) found that trust in data tools depends heavily on perceived interpretability and contextual fit. Our study extends that insight to the realm of AI-based behavioral diagnostics while also highlighting a critical boundary condition: when outputs did not align with teachers' contextual knowledge, teachers did not automatically defer to the system. Instead, they engaged in critical filtering, sometimes disregarding recommendations altogether. Ibu Sri Mu'awana's decision not to simplify tasks for Danendra exemplifies this selective engagement. Far from being a flaw, this behavior demonstrates that professional judgment remained paramount throughout the implementation process.

3.6 Beyond prediction: Machine learning as decision support

Our results strongly support conceptualizing machine learning in education as decision support rather than decision automation. Teachers consistently described the system as a source of organized evidence that informed—but did not dictate—their instructional choices. They used it alongside other sources of knowledge: curricular goals, social-emotional history, parent conversations, and moment-to-moment classroom observations. This approach aligns with Ouyang and Jiao's (2021) call for AI that is "human-instructable" and with Luckin and Colleagues (2016) argument that AI should augment rather than replace teacher expertise.

Importantly, the system's pedagogical suggestions were framed as possibilities, not imperatives. This design choice, rooted in human-centered principles, appeared to protect teachers' sense of agency. No teacher reported feeling that the system undermined their professional identity. To the contrary, many expressed that it enhanced their confidence in decision-making because they could now reference systematic evidence alongside their professional intuition.

3.7 Differentiation and responsive teaching

This study provides concrete examples of how non-cognitive diagnostic information can inform differentiated instruction. Teachers did not simply sort students into fixed categories; rather, they used the profiles to ask, "What kind of support does this particular student need right now?" This approach is consistent with Tomlinson's (2022) conception of differentiation as responsive teaching rather than mere administrative sorting. By highlighting variability in motivation, persistence, and self-regulation, the system helped teachers recognize that even students with similar academic abilities may require fundamentally different instructional approaches.

This finding carries practical implications. It suggests that AI-based diagnostics can support the implementation of differentiated instruction in ordinary classrooms—not by supplanting teacher judgment, but by enriching it with systematic, longitudinal data. When teachers have access to such information, they can tailor task complexity, feedback frequency, grouping arrangements, and scaffolding strategies with far greater precision than intuition alone would allow.

3.8 Teacher reflection and professional learning

Another important contribution of this study is its illustration of how diagnostic tools can foster teacher reflection and professional growth. Ibu Sri Mu'awana's metaphor of the system as a "mirror" captures this dynamic eloquently. Several teachers described moments when the data challenged their initial assumptions—for example, realizing that a student's disengagement was related to task design rather than lack of motivation. Such moments of cognitive dissonance are powerful triggers for teacher learning (Korthagen & Nuijten, 2022).

Importantly, reflection did not occur automatically; it was supported by the system's visualizations—trend lines and radar charts—that made behavioral patterns visible over time. This finding aligns with research on data literacy, which emphasizes that teachers need not only access to data but also representational formats that facilitate sense-making (Fjørtoft & Lai, 2021). Recent studies on teacher dashboards confirm that visualization literacy plays a critical role in how deeply teachers can interpret student data, with well-designed visual representations lowering the interpretive burden and enabling more meaningful reflection (Pozdniakov et al., 2025) When visualizations are intuitive and focused on longitudinal change, they reduce the cognitive load on teachers, allowing them to engage in deeper reflection about their practice rather than getting lost in complex data displays (Karademir et al., 2024). This is consistent with findings that teachers value seeing their data in graphical form, as it helps them reflect on their teaching and encourages them to consider changing their instructional approaches (Pozdniakov et al., 2025).

3.9 Practical implications for elementary schools

The findings of this study offer several practical implications for elementary schools that are considering the adoption of machine learning–based diagnostic systems. One important consideration is the need to prioritize interpretability over algorithmic complexity. Teachers in this study consistently valued outputs that were easy to understand and directly connected to observable classroom behaviors. Visualizations such as simple profiles or trend indicators helped teachers quickly identify patterns in student motivation, engagement, and persistence. When the diagnostic information was transparent and behaviorally meaningful, teachers were more confident in using it to inform instructional decisions. Conversely, systems that rely heavily on complex or opaque algorithms may reduce trust and discourage sustained use because teachers cannot easily verify how conclusions are generated.

Another important principle is framing the system as a decision-support tool rather than an evaluative mechanism. Teachers were more willing to engage with the system when they perceived it as supporting their professional judgment rather than assessing their performance or labeling students. In this study, diagnostic outputs were interpreted as prompts that encouraged reflection and experimentation with instructional strategies. Maintaining teachers' sense of professional autonomy, therefore, becomes a critical design principle. Schools adopting similar technologies should communicate clearly that analytic outputs are advisory and that teachers remain the primary decision-makers who interpret and contextualize the data.

The study also highlights the value of collaborative interpretation. At MI Nurul Islam, teachers often discussed diagnostic profiles informally during grade-level interactions, using them as a starting point for reflecting on classroom experiences. Schools can strengthen this practice by allocating dedicated time for collective data reflection, peer coaching, or lesson study sessions focused on interpreting student learning patterns. Leadership initiatives, such as the monthly "data dialogue" sessions organized by the school administration, illustrate how structured collaboration can help teachers move beyond individual interpretation toward shared pedagogical learning.

Another practical insight concerns the importance of combining algorithmic outputs with qualitative classroom documentation. The most meaningful instructional decisions occurred when teachers integrated system-generated profiles with their own anecdotal notes, observations, and samples of student work. This combination allowed teachers to validate patterns suggested by the system and to interpret them within the broader context of each student's learning experience. In this sense, technology functions most effectively when it complements—rather than replaces—the rich qualitative knowledge that teachers develop through daily interaction with students.

Finally, schools should recognize that successful implementation requires ongoing professional development. Initial training may introduce the technical aspects of the system, but teachers also need continuous opportunities to develop what can be described as pedagogical data literacy. This includes the ability to question analytic outputs, connect them with classroom realities, and translate them into meaningful instructional responses. Sustained professional learning environments, such as collaborative reflection sessions or practice-sharing forums, can help teachers gradually build confidence in using diagnostic data as part of their pedagogical decision-making.

3.10 Limitations and Future Research Direction

Despite these practical contributions, several limitations of the study should be acknowledged, and they also point to directions for future research. The research was conducted in a single school with a relatively small number of participants, which limits the generalizability of the findings. MI Nurul Islam represents a particular institutional context with its own leadership style, culture, and technological readiness. Future studies should replicate this research in a wider range of educational settings, including public schools, larger urban schools, and institutions with different levels of digital infrastructure, in order to determine whether similar patterns emerge across contexts.

Another limitation relates to the system's reliance on manual teacher input. While this approach intentionally preserves teacher agency, it may also introduce subjectivity or inconsistencies in behavioral recording. Future systems could explore integrating complementary data sources, such as learning platform activity logs or other digital traces, while still maintaining teachers' interpretive role in the diagnostic process. Comparative studies could then examine how different data-input models influence teacher trust, data quality, and instructional decision-making.

In addition, the study did not directly measure student learning outcomes. Although teachers reported that the diagnostic information supported their instructional planning, the research design does not allow causal conclusions regarding improvements in academic achievement or non-cognitive development. Future research could employ longitudinal or quasi-experimental approaches to investigate whether teaching informed by diagnostic analytics leads to measurable improvements in student learning.

Another methodological limitation is the absence of systematic classroom observation. The study relied primarily on teacher self-reports and document analysis to understand how diagnostic information was interpreted and used. Direct classroom observation would provide a richer perspective on how these insights are translated into real-time teaching practices and how teachers adapt analytic suggestions to the dynamic realities of the classroom.

Finally, the six-month period of system use represents a relatively short implementation window. Longer-term studies are necessary to examine how teachers' engagement with such tools evolves over time. Future research could investigate whether initial enthusiasm is sustained, whether the technology becomes routinized or ignored, and how long-term use shapes teachers' professional identity and instructional practices.

5. Conclusion

This study examined how elementary school teachers at MI Nurul Islam Lumajang used a machine–learning–based behavioral diagnostic system to support pedagogical decision-making regarding students' non-cognitive competencies. The findings indicate that teachers were able to interpret the diagnostic outputs clearly and considered them relevant to everyday classroom practice. Both quantitative and qualitative evidence suggest that the system was widely perceived as useful for identifying students' learning needs and informing instructional planning. Teachers reported using the diagnostic information to differentiate learning activities, adjust instructional strategies, and reflect more systematically on students' motivation, engagement, self-regulation, and perseverance. At the same time, the study shows that teachers did not rely on the system in a deterministic manner. Instead, they maintained professional autonomy by critically interpreting and contextualizing the system's recommendations based on their knowledge of students and classroom dynamics. In this sense, the diagnostic system functioned primarily as a decision-support tool that complemented teachers' professional judgment and helped validate their pedagogical intuition.

These findings contribute to the emerging literature on human-centered artificial intelligence in education by illustrating how teachers interpret and integrate algorithmic diagnostics within real classroom contexts. The results suggest that the educational value of such systems lies not only in their technical sophistication but also in their capacity to align with teachers' existing interpretative frameworks and to make patterns of student behavior more visible. Particularly in elementary education, where foundational learning habits are formed, integrating non-cognitive diagnostics into everyday pedagogy can support a more holistic understanding of student development that includes both academic and behavioral dimensions. Ultimately, the challenge is not merely to develop increasingly sophisticated algorithms, but to strengthen teachers' capacities to interpret analytic insights, reflect critically on student learning, and exercise contextual judgment in using technology to support responsive and informed instructional decision-making.

References

Akintolu, M. (2025). Data-Driven Decision-Making: Utilising AI-Powered Learning Analytics to Make Informed Primary Educators' Decisions. *Journal of Educators Online*, *22*(3). https://doi.org/10.9743/JEO.2025.22.3.1

Benjamin Balraj, M. (2025). Toward A Theory Of Human-AI Pedagogical Synergy: A Conceptual Framework For Reimagining The Teacher's Role In The Age Of Generative AI. *Scholarly Research Journal for Humanity Science & English Language*. https://doi.org/https://doi.org/10.5281/zenodo.16932848

Bolhuis, E., Voogt, J., & Schildkamp, K. (2019). The development of data use, data skills, and a positive attitude towards data use in a data team intervention for teacher educators. *Studies in Educational Evaluation*, *60*, 99–108. https://doi.org/https://doi.org/10.1016/j.stueduc.2018.12.002

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE.

Cabral, L., Pinto, R., & Gonçalves, G. (2025). AI-powered learning analytics dashboards: a systematic review of applications, techniques, and research gaps. *Discover Education*, *4*(1), 525. https://doi.org/10.1007/s44217-025-00964-y

Crompton, H., & Burke, D. (2018). The use of mobile learning in higher education: A systematic review. *Computers & Education*, *123*. https://doi.org/10.1016/j.compedu.2018.04.007

Cukurova, M., Suraworachet, W., Zhou, Q., & Bulathwela, S. (2025). *Towards Synergistic Teacher-AI Interactions with Generative Artificial Intelligence*. https://doi.org/10.48550/arXiv.2511.19580

Datnow, A., & Park, V. (2021). *Data-driven leadership*. Jossey-Bass.

Efendi, Z., Hanim, M. A. F., & Santoso, A. (2025). Kecerdasan Buatan (AI) dalam Pendidikan: Tinjauan Literatur Sistematis tentang Peluang, Masalah Etika, dan Implikasi Pedagogis. *Jurnal Pendidikan, Kebudayaan Dan Keislaman*, *4*(3), 134–152. https://doi.org/10.24260/jpkk.v4i3.5052

Fjørtoft, H., & Lai, M. K. (2021). Affordances of narrative and numerical data: A social-semiotic approach to data use. *Studies in Educational Evaluation*, *69*, 100846. https://doi.org/https://doi.org/10.1016/j.stueduc.2020.100846

Heikkinen, S., Saqr, M., Malmberg, J., & Tedre, M. (2023). Supporting self-regulated learning with learning analytics interventions – a systematic literature review. *Education and Information Technologies*, *28*(3), 3059–3088. https://doi.org/10.1007/s10639-022-11281-4

Herppich, S., Praetorius, A. K., & Grünkorn, J. (2023). Teachers' diagnostic judgment: A review of processes, accuracy, and pedagogical relevance. *Educational Psychology Review*, 35 (2), Article 49. https://doi.org/10.1007/s10648-023-09762-0

Holmes, W., Bialik, M., & Fadel, C. (2022). Artificial intelligence in education: Promise and implications for teaching and learning. *Center for Curriculum Redesign*. https://www.researchgate.net/publication/332180327

Ibad, T. N., Wahidah, F., Ansori, M., Rofi'ah, S. H., & Hanisy, A. (2024). Digital Literacy Management Based On Open Artificial Intelligence (Open AI) In PGMI Student Research Development. *Proceedings of the 2nd Lawang Sewu International Symposium on Humanities and Social Sciences 2023 (LEWIS HUSO 2023)*, 316–328. https://doi.org/10.2991/978-2-38476-267-5_25

Karademir, O., Di Mitri, D., Schneider, J., Jivet, I., Allmang, J., Gombert, S., Kubsch, M., Neumann, K., & Drachsler, H. (2024). I don't have time! But keep me in the loop: Co-designingrequirements for a learning analytics cockpit with teachers. *Journal of Computer Assisted Learning*, *40*(6), 2681–2699. https://doi.org/10.1111/jcal.12997

Khosravi, H., Shum, S. B., & Kay, J. (2023). Explainable learning analytics in education: A systematic review. *Computers & Education: Artificial Intelligence*, 5, 100193. https://doi.org/10.1016/j.caeai.2023.100193

Korthagen, F., & Nuijten, E. (2022). The Power of Reflection in Teacher Education and Professional Development: Strategies for In-Depth Teacher Learning. In *The Power of Reflection in Teacher Education and Professional Development: Strategies for In-Depth Teacher Learning*. Taylor and Francis. https://doi.org/10.4324/9781003221470

OECD. (2021). Beyond academic learning: First results from the survey of social and emotional skills . *OECD Publishing*. https://doi.org/https://doi.org/10.1787/92a11084-en

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, 2. https://doi.org/10.1016/j.caeai.2021.100020

Pozdniakov, S., Martinez-Maldonado, R., Tsai, Y. S., Echeverria, V., Swiecki, Z., & Gašević, D. (2025). Investigating the Effect of Visualization Literacy and Guidance on Teachers' Dashboard Interpretation. *Journal of Learning Analytics*, *12*(1), 367–390. https://doi.org/10.18608/jla.2024.8471

Revuelta-Domínguez, F., Guerra-Antequera, J., Gonzalez Perez, A., Pedrera-Rodríguez, M., & González-Fernández, A. (2022). Digital Teaching Competence: A Systematic Review. *Sustainability*, *14*, 6428. https://doi.org/10.3390/su14116428

Shneiderman, B. (2022). *Human-Centered AI*. Oxford University Press. https://doi.org/10.1093/oso/9780192845290.001.0001

Shulman, L. (1987). Knowledge and Teaching:Foundations of the New Reform. *Harvard Educational Review*, *57*(1), 1. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Tomlinson, C. A. (2022). *How to differentiate instruction in academically diverse classrooms (3rd ed.)*. ASCD.

van der Lans, R. M., van de Grift, W. J. C. M., & van Veen, K. (2022). Developing teachers' ability to interpret classroom situations: The role of diagnostic judgment. *Teaching and Teacher Education*, 110, 103598. https://doi.org/10.1016/j.tate.2021.103598

Viberg, O., Mutimukwe, C., Hrastinski, S., Cerratto-Pargman, T., & Lilliesköld, J. (2024). Exploring teachers' (future) digital assessment practices in higher education: Instrument and model development. *British Journal of Educational Technology*, 55 (6), 2597–2616. https://doi.org/10.1111/bjet.13462

Wise, A. F., & Jung, Y. (2019). Teaching with Analytics: Towards a Situated Model of Instructional Decision-Making. *Journal of Learning Analytics*, *6*(2), 53–69. https://doi.org/10.18608/jla.2019.62.4

Zhou, M., & Brown, D. (2023). Non-cognitive skills and student learning: A review of recent empirical evidence. *Educational Research Review*, 38, 100487. https://doi.org/10.1016/j.edurev.2022.100487