Beyond Accuracy and Speed: A Learning Analytics Approach to Performance Modeling in Digital Phonological Games for Brazilian Children

Tiago José Benedito Eugenio[1], Ana Lucia Hennemann[2], Denise De Micheli[3]

[1] Federal University of São Paulo, Brazil
[2]Universidad Internacional Iberoamericana, United States
[3] Federal University of São Paulo, Brazil

Abstract

The increasing use of digital technologies in education has expanded the possibilities for assessing cognitive processes related to learning. However, performance in digital tasks is still frequently analyzed using isolated metrics, such as accuracy or response time, which may fail to capture the multidimensional nature of cognitive performance. This study proposes a composite indicator, named the Game Progress Index (GPI), designed to integrate accuracy, speed, and efficiency into a standardized measure applied to digital phonological awareness games. A total of 339 Brazilian children in the early stages of literacy development participated in the study and completed seven digital game-based tasks. Different weighting models were tested, and the model showing the strongest association with performance indicators was selected. A regression-based approach was then employed to estimate expected performance, allowing the calculation of standardized residuals and the identification of discrepancies between observed and predicted outcomes. Results demonstrated a strong association between the GPI and performance rate (r = .90), supporting the coherence and sensitivity of the proposed index. The combined analysis of percentile scores and residuals revealed heterogeneous performance profiles, including children with high overall performance but lower-than-expected outcomes, as well as participants with low overall performance but outcomes above model predictions. Findings suggest that traditional normative metrics alone may be insufficient to capture individual variability in digital learning contexts. The proposed approach offers a more sensitive framework for modeling performance in educational games, with potential implications for educational assessment, learning analytics, and game-based cognitive screening.

Keywords: *learning analytics; digital assessment; serious games; cognitive performance; phonological awareness*

1. Introduction

The increasing use of digital technologies has profoundly transformed teaching, learning, and assessment processes. In contrast to traditional paper-and-pencil instruments, digital environments enable the continuous collection of behavioral data generated during user interaction with tasks, including the number of correct responses, response times, interaction patterns, and performance trajectories [1–3]. These records, often referred to as behavioral trace data, expand opportunities to investigate the cognitive processes involved in learning and create new possibilities for the development of more sensitive and informative assessment systems [4–6].

Within this context, Learning Analytics has emerged as an important field of research dedicated to the collection, analysis, and interpretation of educational data to better understand and optimize learning processes [1,7]. Unlike traditional assessment approaches, which focus primarily on final performance outcomes, contemporary Learning Analytics perspectives seek to understand how individuals perform tasks by examining information related to cognitive processing dynamics and learner behavior throughout task execution [2].

Despite these advances, a substantial proportion of digital assessments still rely on isolated indicators to represent participant performance. Measures such as the number of correct responses, percentage of correct answers, or total completion time are frequently analyzed separately, which may limit the understanding of the multidimensional nature of cognitive performance [3]. From a theoretical perspective, performance in cognitive tasks emerges from the interaction of multiple components, including accuracy, processing speed, and problem-solving efficiency [8,9]. Consequently, individuals with similar levels of accuracy may exhibit substantially different cognitive profiles when temporal and process-related aspects are taken into account.

This issue is particularly relevant in the context of literacy acquisition. Among the various predictors of reading development, phonological awareness is widely recognized as one of the most important skills for acquiring an alphabetic writing system and has consistently been associated with reading performance across different languages and educational contexts [10,11]. In recent years, digital tasks designed to assess these skills have been increasingly adopted, facilitating automated data collection and the generation of more precise performance measures [12].

However, the availability of large volumes of data alone does not necessarily lead to advances in understanding learning processes. One of the central challenges lies in developing analytical models capable of integrating multiple dimensions of performance while producing indicators that are meaningful and interpretable for researchers and educators [5,13]. In this regard, composite metrics have been proposed as a promising alternative for representing performance more comprehensively by combining information related to accuracy, speed, and efficiency into a single synthetic indicator [3,14].

At the same time, predictive modeling approaches have gained increasing attention within the field of Learning Analytics because they enable the estimation of expected performance based on behavioral variables collected during task execution [5,15]. Comparing observed and expected performance makes it possible to identify individual discrepancies that may not be captured by traditional normative metrics. Within this framework, residual analysis represents a particularly valuable strategy, as it allows the identification of participants whose performance falls above or below the expectations established by predictive models [16,17].

Against this background, the present study proposes a Learning Analytics–based approach for modeling performance in digital phonological awareness games. To this end, the Game Progress Index (GPI), a composite metric integrating indicators of accuracy, speed, and efficiency derived from participants' interactions with digital tasks, was employed. Subsequently, a quadratic model based on Accuracy per Minute (APM) was used to estimate participants' expected performance, enabling the calculation of standardized residuals and the identification of discrepancies between observed and predicted outcomes [5,15].

Accordingly, the aim of this study was to investigate the relationship between the Game Progress Index (GPI) and Accuracy per Minute (APM), as well as to examine the potential of residual analysis for identifying performance profiles that are not adequately represented by traditional normative indicators. By integrating composite metrics, predictive modeling, and the analysis of individual discrepancies, this study seeks to contribute to the development of more sensitive approaches for game-based cognitive assessment in educational contexts.

2. Method

2.1. Study Design

This was a quantitative, observational, cross-sectional study conducted to investigate performance modeling in Portuguese-language digital phonological awareness games using a Learning Analytics–based approach. Behavioral data collected during participants' interactions with digital tasks were analyzed, including indicators of accuracy, completion time, and efficiency.

2.2. Participants

A total of 369 children in the early stages of literacy development initially participated in the study. Of these, 26 were excluded from the analyses because they presented a diagnosis or indication of neurodevelopmental disorders according to the study eligibility criteria, and 4 did not complete the assessment battery. The final sample therefore consisted of 339 children. Participants included in the analyses had a mean age of 6.45 years (SD = 0.84), ranging from 5.10 to 8.03 years. The sample comprised 170 boys (50.1%) and 169 girls (49.9%). Regarding educational level, 129 participants (38.1%) were enrolled in Early Childhood Education, 121 (35.7%) attended Grade 1, and 89 (26.3%) attended Grade 2. All participants were recruited from a private school located in São Bernardo do Campo, Brazil.

2.3. Instruments

Data were collected using seven digital games designed to assess phonological awareness skills. The games were developed by the authors and implemented within a Brazilian gamified cognitive assessment platform (https://clickneurons.com.br). Tasks were organized in a modular structure, with each game containing either 10 or 20 items, resulting in a total of 100 items across the assessment battery. This distribution was defined according to the complexity of the skills being assessed, aiming to balance cognitive load across tasks while ensuring adequate representation of different components of phonological processing [10,11]. As shown in Table 1, the games assessed skills such as phoneme–grapheme correspondence, syllable segmentation, identification of initial and final syllables, and rhyme recognition. Tasks were presented through an interactive digital interface, with automated recording of correct responses and response times for each item, enabling the calculation of accuracy, speed, and efficiency indices.

Table 1. Organization of the digital games included in the phonological awareness assessment battery

Game *Portuguese / English name*	**Primary Skill**	**Task Type**	**No. of Items**
Sopa de Letrinhas *Letter Soup*	Phoneme–Grapheme Correspondence	Letter selection following an auditory stimulus	10
Toca o Som *Touch the Sound*	Auditory Discrimination	Selection of the grapheme corresponding to the presented sound	20
Mexe a Boca *Move Your Mouth*	Syllabic Awareness	Identification of the number of syllables	10
Correio das Sílabas *Syllable Mail*	Initial Syllables	Matching words sharing the same initial syllable	10
Tesouro das Sílabas *Syllable Treasure*	Initial Syllables	Matching words sharing the same initial syllable	20
Oficina das Sílabas *Syllable Garage*	Final Syllables	Identification of the final syllable	20
O Poeta Falou *The Poet Said*	Rhyme Awareness	Identification of rhyming words	10

During task execution, immediate feedback was provided to participants to maintain engagement and support understanding of task requirements. From a measurement perspective, the digital environment enabled the collection of detailed behavioral data, including: (a) number of correct responses (accuracy); (b) total number of items presented; (c) response time per game; and (d) total task completion time.

To increase the precision of temporal measures, the execution time used in the analyses excluded intervals not directly related to the cognitive processing demands of the task, such as transitions between stimuli, visual and auditory feedback periods, and delays inherent to interface operation. This procedure was intended to ensure that timing measures primarily reflected participants' cognitive processing and response execution, thereby reducing the influence of extraneous factors unrelated to task performance [6,14,17].

The resulting interaction data served as the basis for calculating the GPI, a composite metric developed to integrate indicators of accuracy, speed, and efficiency derived from participants' interactions with the digital games. The index ranges from 0 to 100, with higher scores indicating better performance.

2.4. Data Collection Procedures

Data collection was conducted in a school setting through individual assessment sessions administered by previously trained researchers. All participants completed the tasks using the same device model, an 11th-generation iPad with an 11-inch screen, ensuring standardized presentation of visual stimuli.

Auditory stimuli were delivered through noise-canceling headphones to minimize environmental distractions and ensure uniform testing conditions across participants. During the assessment, children completed a battery of seven digital phonological awareness games designed to evaluate skills associated with early literacy development.

Performance data were automatically recorded by the digital platform, including information related to response accuracy and task completion time. All records were subsequently exported for statistical analysis.

The study was approved by the Research Ethics Committee of the Federal University of São Paulo (Approval No. 8.132.619; CAAE No. 94688725.9.0000.5505) and was conducted in accordance with national ethical guidelines for research involving human participants. Participation was authorized by the children's legal guardians prior to data collection.

3. Results

Table 2 presents the descriptive statistics for the main variables investigated. The GPI showed a mean score of 75.01 (SD = 15.07), whereas APM presented a mean of 7.35 correct responses per minute (SD = 4.12). Performance estimated by the quadratic model showed a mean similar to that observed for the GPI (M = 75.01; SD = 14.45), suggesting good overall correspondence between observed and predicted values.

Table 2. Descriptive statistics of the main study variables (N = 339).

Variable	Mean	SD	Minimum	Maximum
GPI	75.01	15.07	40.02	99.36
APM	7.35	4.12	1.04	19.86
Predicted GPI	75.01	14.45	44.22	96.21
Quadratic Residual	0.00	4.28	-25.14	11.23

Note: GPI = Game Progress Index; APM = Accuracy per Minute; SD = Standard Deviation.

3.1. Association Between Overall Performance and Temporal Efficiency

A strong positive association was observed between the GPI and APM, $r = .913$, $p < .001$. This result indicates that participants with greater temporal efficiency tended to exhibit higher levels of overall performance in the digital tasks. The coefficient of determination ($r^2 = .834$) indicated that approximately 83.4% of the variability observed in GPI was associated with variability in APM.

3.2. Modeling Expected Performance

To estimate participants’ expected performance, a second-order polynomial regression model was fitted using APM as the predictor variable. The model demonstrated an excellent fit to the data ($R^2 = .919$; adjusted $R^2 = .919$), explaining approximately 91.9% of the variance observed in the GPI (Table 3). Analysis of variance indicated that the model was statistically significant, $F(2,336) = 1916.95$, $p < .001$. The resulting equation was:

$$GPI = 36.730 + 7.475(APM) - 0.235(APM^2).$$

Table 3. Polynomial Regression Model Predicting GPI from APM

Predictor	B	95% CI	p
Constant	36.730	—	<0.001
APM	7.475	[7.03, 7.92]	<0.001
APM²	-0.235	[-0.259, -0.210]	<0.001

Both the linear APM term (B = 7.475, 95% CI [7.03, 7.92], $p < .001$) and the quadratic term (B = −0.235, 95% CI [−0.259, −0.210], $p < .001$) contributed significantly to the model. The inclusion of the quadratic term increased the model’s explanatory power relative to the simple linear association reported previously, providing evidence of a non-linear relationship between APM and GPI. Figure 1 illustrates the distribution of participants and the fitted quadratic curve.

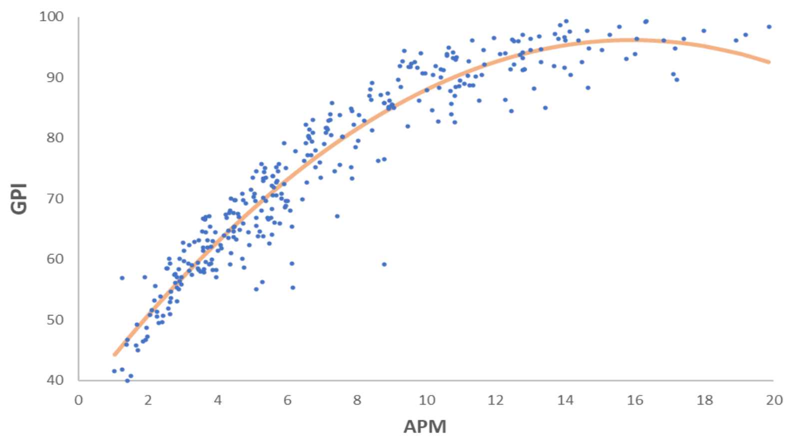

Fig. 1. Relationship between APM and GPI with quadratic model fit ($R^2 = 0.919$).

3.3. Performance Quadrant Analysis

Based on the predictive model, individual residuals were calculated as the difference between observed and expected performance. Residuals were subsequently standardized and analyzed jointly with GPI percentile ranks. Participants were classified using the 50th percentile of the GPI distribution and a standardized residual value of zero as cutoff points. The combination of these criteria allowed each participant to be positioned within a two-dimensional space defined by normative performance and discrepancy from expected performance.

The integration of percentile rank and standardized residual scores enabled the identification of four performance profiles: **Expected High**, **Unexpected High**, **Expected Low**, and **Unexpected Low**. These profiles represent distinct combinations of normative performance level and deviation from model-based expectations (Figure 2).

Table 4. Distribution of Participants Across Analytical Quadrants

Profile	**n**	**%**
Expected High	114	33.6
Unexpected High	56	16.5
Unexpected Low	74	21.8
Expected Low	95	28.0
Total	339	100.0

Note: Expected High = participants with GPI percentile rank ≥ 50 and standardized residual ≥ 0; Unexpected High = GPI percentile rank < 50 and standardized residual ≥ 0; Unexpected Low = GPI percentile rank ≥ 50 and standardized residual < 0; Expected Low = GPI percentile rank < 50 and standardized residual < 0.

Participants were distributed across all four quadrants. The discrepant groups (Unexpected High and Unexpected Low) jointly represented 38.3% of the sample, indicating that a substantial proportion of participants exhibited performance patterns that differed from those predicted by the model.

3.4. Association Between Performance Quadrants and School Grade

The distribution of participants across performance quadrants varied significantly as a function of school grade, $\chi^2(6) = 176.89$, $p < .001$. Cramer's V indicated a large association ($V = .511$). As shown in Figure 2, children enrolled in Early Childhood Education were predominantly concentrated on the left side of the analytical space, corresponding to the Expected Low and Unexpected High profiles, whereas students in Grades 1 and 2 were more frequently represented on the right side, corresponding to the Expected High and Unexpected Low profiles.

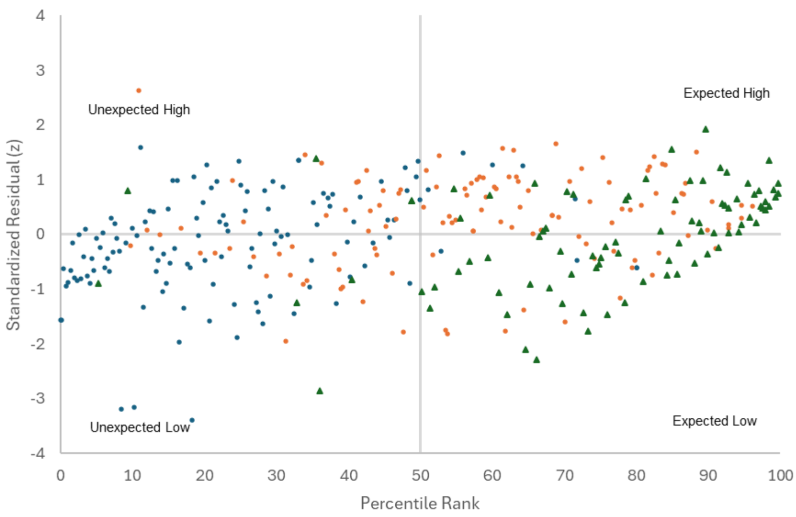

Fig. 2. Distribution of participants according to percentile rank and standardized residual scores. Blue circles = Early Childhood Education; orange circles = Grade 1; green triangles = Grade 2. Reference lines indicate the 50th percentile and z = 0 cutoffs.

4. Discussion

The present study investigated a Learning Analytics–based approach for modeling children's performance in Portuguese-language digital phonological awareness games. Results revealed a strong association between the Game Progress Index (GPI) and Accuracy per Minute (APM), as evidenced by the observed correlation ($r = .913$, $p < .001$) and the high coefficient of determination of the quadratic model ($R^2 = .919$). Taken together, these findings indicate that the integration of accuracy and temporal efficiency provides a robust representation of participants' performance in digital tasks related to early literacy development.

Traditionally, digital educational assessments rely on isolated indicators, such as the number of correct responses or total completion time. Although these measures provide valuable information, they do not always capture the multidimensional nature of cognitive performance. The findings of the present study reinforce the importance of simultaneously considering multiple dimensions of learner behavior, a perspective consistent with contemporary approaches in Learning Analytics and digital assessment [1–3]. In this context, the GPI was designed to integrate accuracy, speed, and efficiency into a single metric, generating a measure that was strongly associated with the behavioral data produced during interaction with the games.

Another important finding was the identification of a non-linear relationship between APM and GPI. The significance of the quadratic term indicates that performance gains associated with increased temporal efficiency do not occur uniformly across the distribution. Performance improved more rapidly at lower levels of APM, followed by a progressive deceleration at higher levels. This pattern is consistent with findings from cognitive performance research, in which additional gains in speed tend to yield progressively smaller improvements in overall performance [8,9].

The most relevant contribution of this study, however, concerns the analysis of discrepancies between observed and expected performance. The integration of percentile ranks and standardized residuals enabled the identification of distinct performance profiles, revealing that a substantial proportion of the sample (38.3%) exhibited patterns that diverged from model-based expectations. This finding suggests that traditional normative indicators may be insufficient to fully characterize individual differences in digital learning environments. Participants occupying similar positions within the normative distribution may nevertheless display markedly different performance profiles when efficiency patterns captured by predictive modeling are taken into account. This interpretation is consistent with recent Learning Analytics research advocating the integration of multiple sources of evidence to better understand individual differences in student performance [5,14,15].

The results also demonstrated a significant association between performance quadrants and school grade. Children enrolled in Early Childhood Education were predominantly concentrated in the lower-performance regions of the analytical space, whereas students in Grades 1 and 2 were more frequently represented in the higher-performance regions. This pattern suggests that the proposed approach is sensitive to developmental differences expected throughout the literacy acquisition process. At the same time, the presence of participants distributed across different quadrants within the same grade level indicates that the model may also capture individual variability that is not fully explained by schooling alone.

Taken together, the findings suggest that combining composite performance indicators, predictive modeling, and residual analysis can expand the interpretative potential of data generated in digital learning environments, providing information that complements the metrics traditionally used in educational assessment.

Several limitations should be considered when interpreting the results. First, the sample consisted of Brazilian children in the early stages of literacy development, which limits the generalizability of the findings to other age groups and educational contexts. Second, the data were obtained from a specific battery of Portuguese-language digital phonological awareness games, and it cannot be assumed that similar patterns would emerge in other cognitive domains or types of digital tasks. In addition, the cross-sectional design precludes inferences regarding individual changes over time or causal relationships among the variables investigated. Finally, although the model demonstrated substantial explanatory power, further research is needed to evaluate its stability in independent samples and across different educational settings.

5. Conclusion

This study presented a Learning Analytics–based approach for modeling children's performance in Portuguese-language digital phonological awareness games. The results revealed a strong association between the GPI and APM, as well as an excellent fit of a quadratic model capable of explaining approximately 91.9% of the observed variance in performance.

The combined analysis of percentile ranks and standardized residuals enabled the identification of performance profiles that would not be readily detected using conventional normative indicators, highlighting the potential of predictive modeling to complement digital assessment practices. Furthermore, the association observed between performance quadrants and school grade suggests that the proposed approach is sensitive to developmental differences occurring during literacy acquisition.

Overall, the findings indicate that the integration of behavioral metrics, statistical modeling, and Learning Analytics can contribute to the development of more informative digital assessment systems, providing a more nuanced understanding of individual differences in educational contexts. Future studies should examine the applicability of this approach across other cognitive domains, age groups, and educational settings, further expanding the evidence base for its use in assessment and learning monitoring.

REFERENCES

[1] Ifenthaler D., Yau J. Y.-K. Utilising Learning Analytics for Study Success: Reflections on Current Empirical Findings. Research and Practice in Technology Enhanced Learning, 15(1), 2020.
[2] Buckingham Shum S. Luckin R., Learning Analytics and AI: Politics, Pedagogy and Practices. British Journal of Educational Technology, 50(6), 2019, pp. 2785–2793.
[3] Shute V. J., Rahimi S. Review of Computer-Based Assessment for Learning in Elementary and Secondary Education. Journal of Computer Assisted Learning, 37(1), 2021, pp. 19–32.
[4] Greiff S., Niepel C., Scherer R., Martin R. Understanding Students' Performance in a Computer-Based Assessment of Complex Problem Solving: An Analysis of Behavioral Process Data. Computers in Human Behavior, 61, 2016, pp. 36–46.
[5] Baker R. S., Inventado P. S. Educational Data Mining and Learning Analytics In: Larusson J. A., White B. (Eds.), Learning Analytics: From Research to Practice, New York, Springer, 2014, pp. 61–75.
[6] Mislevy R. J., Behrens J. T., DiCerbo K. E., Levy R. Design and Discovery in Educational Assessment: Evidence-Centered Design, Psychometrics, and Educational Data Mining. Journal of Educational Data Mining, 4(1), 2012, pp. 11–48.
[7] Viberg O., Hatakka M., Bälter O., Mavroudi A. The Current Landscape of Learning Analytics in Higher Education. Computers in Human Behavior, 89, 2018, pp. 98–110.

[8] Heitz R. P. The Speed–Accuracy Tradeoff: History, Physiology, Methodology, and Behavior. Frontiers in Neuroscience, 8, 2014, Article 150.
[9] Ratcliff R., Smith P. L., Brown S. D., McKoon G. Diffusion Decision Model: Current Issues and History. Trends in Cognitive Sciences, 20(4), 2016, pp. 260–281.
[10] Snowling M. J., Hulme C. The Science of Reading: A Handbook, Oxford, Blackwell Publishing, 2021.
[11] Melby-Lervåg M., Lyster S.-A. H., Hulme C. Phonological Skills and Their Role in Learning to Read: A Meta-Analytic Review. Psychological Bulletin, 138(2), 2012, pp. 322–352.
[12] McTigue E. M., Solheim O. J., Zimmer W. K., Uppstad P. H. Critically Reviewing Early Literacy Assessment Tools in a Digital Age. Reading Research Quarterly, 55(S1), 2020, pp. S111–S127.
[13] Ferguson R. Learning Analytics: Drivers, Developments and Challenges. International Journal of Technology Enhanced Learning, 4(5–6), 2012, pp. 304–317.
[14] Greiff S., Wüstenberg S., Avvisati F. Computer-Generated Log-File Analyses as a Window into Students' Minds? A Showcase Study Based on PISA 2012. Computers & Education, 98, 2016, pp. 92–105.
[15] Tempelaar D. T., Rienties B., Nguyen Q. Towards Actionable Learning Analytics Using Dispositions. IEEE Transactions on Learning Technologies, 13(3), 2020, pp. 449–460.
[16] Cohen J., Cohen P., West S. G., Aiken L. S., Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 3rd ed., New York, Routledge, 2013.
[17] Goldhammer F., Martens T., Lüdtke O. Measuring Cognitive Processes in Computer-Based Assessments: Evidence from Process Data. Educational Measurement: Issues and Practice, 36(4), 2017, pp. 15–25.