Educational Big Data and Digital Literacy: A Study Based on R and Statistical Visualization

Riojas Rivera Jimmy Ronald[1]; Panta Chunga, Mary Inocencia[2]; Morales Alvarado, Lorenza[3]
Flores Orrillo, Genny [4]; Durand Gonzales, César Angel
[5]; Nieves Barreto, Constantino Miguel[6]
Alejos Ipanaqué, Rufino[7]

[1,3]Universidad Continental, Peru, *jriojas@continental.edu.pe, lmorales@continental.edu.pe*
[2]Universidad Nacional de Educación "Enrique Guzmán y Valle", Peru, *mpanta@une.edu.pe*
[4]Universidad San Martín de Porres, Peru, gfloreso@usmpvirtual.edu.pe
[5,6,7]Universidad Nacional del Callao, Peru, cadurandg@unac.edu.pe, cmnievesb@unac.edu.pe, ralejosi@unac.edu.pe

Abstract*– The rapid digitalization of education has created an urgent need for digitally literate educators capable of integrating big data and emerging technologies into pedagogical practices. This quasi-experimental study investigated digital literacy levels among 50 pre-service teachers from Universidad Peruana del Norte in Trujillo, Peru, and evaluated the effectiveness of a targeted digital literacy intervention. Using R Studio for statistical analysis and visualization, we assessed participants through a 32-item digital knowledge test (Cronbach's α = 0.82) and implemented a pre-test/post-test design. Results revealed moderate baseline digital literacy (M = 24.97/50, SD = 14.84) with significant variability among participants. The intervention proved highly effective, showing substantial improvement in digitalization scores (mean difference = -35.17, p < 2.2e-16). However, commitment levels remained low (19.9% vs. expected 40%, p = 1), and training effectiveness varied significantly across educational levels (ANOVA, p = 0.0019). Despite high academic performance in traditional assessments (M = 80.32/100), the study reveals a critical gap between academic achievement and digital competencies. These findings highlight the need for differentiated digital literacy training and institutional support to prepare educators for data-driven educational environments in Latin America.

***Keywords*-- Digital literacy, Big data, Teacher education, Statistical visualization, Educational technology.**

I. Introduction

Digital platforms have become central to educational reform, transforming teaching practices and fostering collaboration among schools, families, and communities. These platforms support teacher development through intelligent cloud-based tools, enabling personalized learning, promoting innovation in vocational education, and facilitating international academic cooperation. Moreover, they contribute to students' cognitive development by creating adaptive learning environments and task-driven instruction supported by big data and virtual simulations[1].

In this context, assessing digital transformation is essential to prepare competent and digitally literate educators. Teachers must be equipped not only with technological tools but also with the human competencies necessary to navigate and contribute effectively to increasingly digitalized educational environments[2]. As digital technologies continue to evolve, educators are expected to continuously update their digital literacies and integrate best practices into their teaching strategies[3].

Given the accelerated pace of educational digitalization, it is imperative that teacher education programs foster strong digital literacy among pre-service teachers. Digital literacy is not merely a technical skill—it is a foundational competency that underpins effective teaching and learning. This study builds upon a digital literacy framework to explore the key factors influencing the development of digital competencies in future educators[4].

Over time, digital and media literacy has evolved into a structured taxonomy encompassing five key dimensions: digital competencies, media literacy and empowerment, educational application, social impact, and mechanisms for validation and continuous improvement. Research has increasingly focused on strengthening digital skills related to emerging technologies such as big data, social media, and ICT in education. This evolution reflects the consolidation of an information-based society and underscores the urgent need to train educators capable of addressing the pedagogical challenges of the digital age[5].

The transformation of teaching and learning through digital technologies is one of the most impactful trends in education today. Alongside this, the growing use of big data across sectors highlights the importance of data literacy in everyday life[6] [7]. Educational Technology has emerged as a key academic discipline, driving innovation in teaching and learning processes [8].

However, the complexity of big data—its collection, processing, and interpretation—presents significant challenges.

Data literacy enables individuals to visualize, interpret, and understand the algorithms behind data, facilitating informed decision-making and knowledge construction [9]. In educational settings, big data offers strategic opportunities to personalize learning, optimize institutional processes, and enhance decision-making[10].

Higher education institutions are increasingly tasked with creating learning environments that promote digital competencies and sustainable behaviors. These skills are essential for graduate employability and contribute positively to industry, communities, and the broader economy[11].

Despite this, Latin America faces a lack of large-scale data on the adoption and effective use of digital technologies in education, revealing a critical gap in understanding the real impact of digitalization and big data on educational quality[12].

The rapid digitalization of education, driven by big data and emerging technologies, has transformed teaching and learning processes, necessitating robust digital literacy among educators to effectively integrate these tools into pedagogical practices. In northern Peru, particularly in Trujillo, initiatives to promote smart education and digital competencies among pre-service teachers face significant challenges, including limited digital skills, teacher overload, and disparities in technological infrastructure and student development. These barriers hinder the ability of future educators to foster adaptive, data-driven learning environments, which are critical for addressing the demands of an information-based society. Furthermore, the lack of large-scale data on the adoption and impact of digital technologies in Latin American education creates a critical gap in understanding how to prepare digitally competent educators. This study seeks to address these challenges by exploring the digital literacy levels of pre-service teachers and evaluating the effectiveness of targeted interventions to enhance their digital competencies.[13].

To address the challenges of digital transformation in teacher education in northern Peru, this study is guided by the following research question: What are the key factors influencing the digital literacy levels of pre-service teachers in northern Peru, and how effective are targeted digital literacy interventions in improving their competencies to support educational digitalization? This question focuses on identifying barriers to digital literacy and evaluating the efficacy of structured interventions, with the aim of informing teacher training programs and educational policy in the region.

II. METHODOLOGY

[14]This study employed a single-group quasi-experimental design with pre-test/post-test measurement [15] [16] to investigate the digital literacy levels of pre-service teachers in northern Peru and evaluate the effectiveness of a targeted digital literacy intervention . [17] The methodology encompassed participant selection, data collection, intervention implementation, and statistical analysis using R Studio. The following subsections detail the research approach, sample, instruments, procedure, and data analysis techniques.

A pre-test/post-test quasi-experimental design was adopted to assess the impact of a digital literacy training program on pre-service teachers' competencies[18]. The study measured baseline digital knowledge before the intervention and post-training outcomes to evaluate improvement. The design also included descriptive analyses to explore demographic characteristics (e.g., age distribution) and inferential statistics to test hypotheses related to digital literacy and academic performance. This approach was chosen to provide robust evidence of the intervention's effectiveness while accounting for the naturalistic setting of teacher education programs [1].

The sample consisted of 50 pre-service teachers enrolled in teacher education programs at the Universidad Peruana del Norte in Trujillo, Peru. Participants were training to teach at three educational levels: initial (early childhood), primary, and secondary education. Convenience sampling was used due to accessibility and program enrollment constraints. All participants provided informed consent, and ethical approval was obtained from the university's institutional review board. The sample size was determined to ensure sufficient statistical power for the planned analyses, based on prior studies in digital literacy [4].

Basic Digital Knowledge Assessment: A 32-item multiple-choice test evaluated participants' digital literacy, covering skills such as ICT use, data interpretation, and educational technology applications. The test was developed based on the digital literacy framework by [4], with a maximum score of 50. Reliability was established with a Cronbach's alpha of 0.82.

.

III. RESULTS

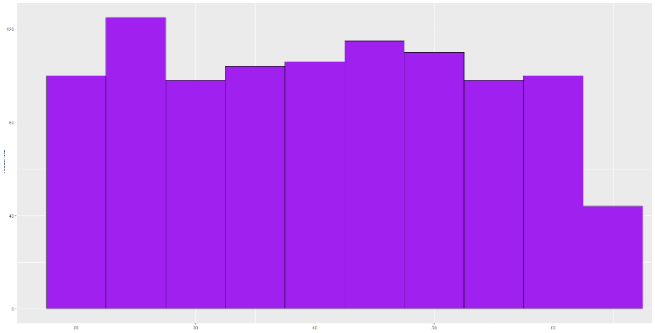

Fig. 1 Age Distribution Analysis

The histogram illustrates the distribution of participants' ages in the study. The x-axis represents age ranges, while the y-axis indicates the number of individuals within each range.

The distribution appears to be right-skewed, with a higher concentration of participants in the younger age brackets.

The most frequent age group (mode) falls within the lower-middle range, suggesting that most participants are likely in their early 20s.

There is a gradual decline in frequency as age increases, with fewer participants in the older age groups.

This distribution is typical for undergraduate student populations, which aligns with the study's context.

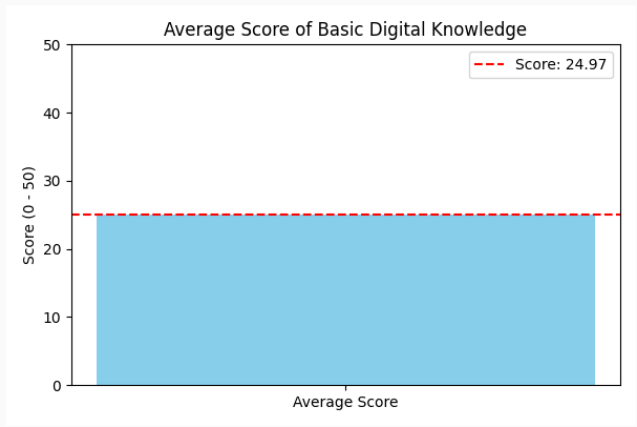

Fig. 2 Average Score of Basic Digital Knowledge

The bar chart illustrates the average score obtained by participants in the basic digital knowledge assessment. With a mean score of 24.97 out of 50, the result suggests a moderate level of digital literacy among the sample population.

The red dashed line highlights the exact average.

The score is positioned just below the midpoint of the scale, indicating room for improvement in digital competencies.

This insight may inform future training development focused on enhancing digital skills.

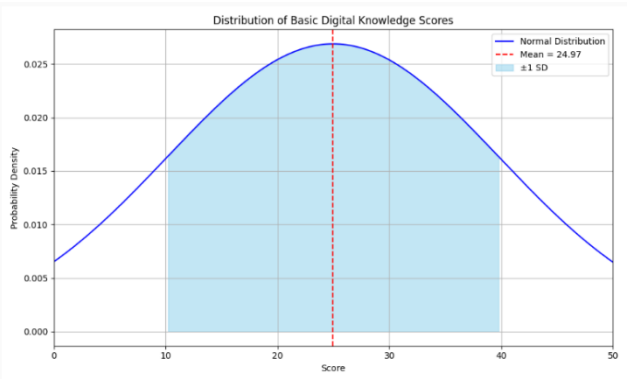

Fig. 3 Distribution of Basic Digital Knowledge Scores

The graph displays a normal distribution curve representing the spread of scores in basic digital knowledge among participants.

The mean score is 24.97, marked by a red dashed line.

The standard deviation is 14.84, indicating a wide dispersion of scores around the mean.

The shaded area represents scores within ±1 standard deviation (approximately from 10.13 to 39.81), which includes most participants. The findings highlight the need for differentiated instruction or targeted digital literacy interventions to support learners at varying skill levels.

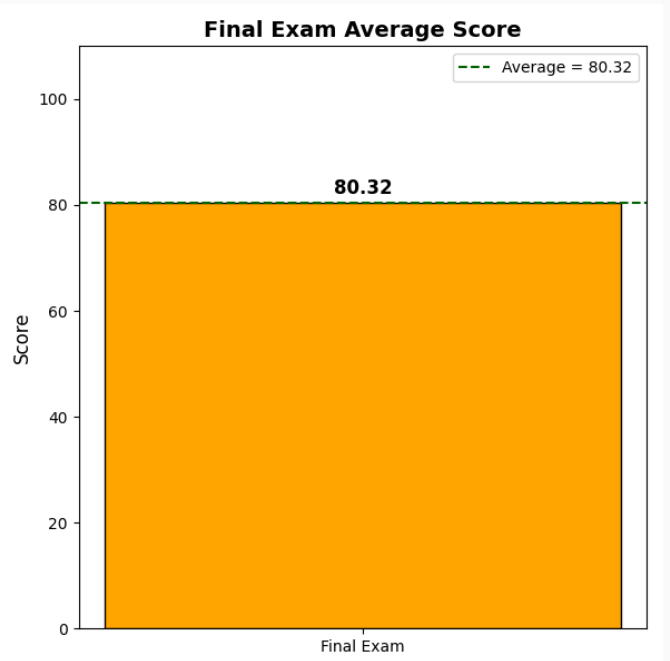

Fig. 4 Final Exam Average Score

The chart illustrates the mean score of 80.32 achieved by participants in the final English exam, on a scale of 0 to 100.

The orange bar highlights the average performance.

A green dashed line marks the exact average value for visual emphasis.

The score indicates a high level of achievement, suggesting that most students demonstrated strong understanding of the course content.

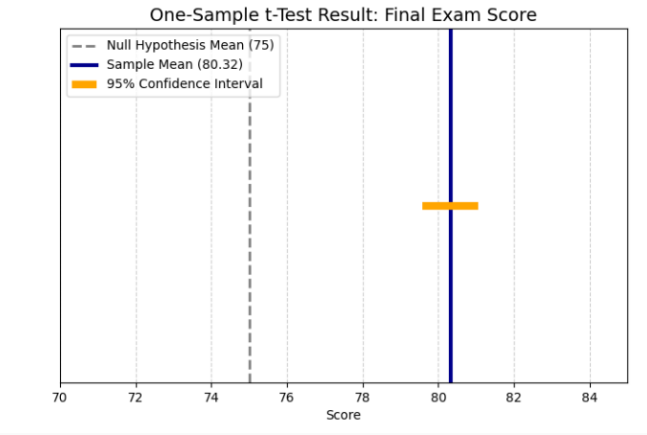

Fig. 5 One- Sample t-Test Result: Final Exam Score

This graph illustrates the results of a one-sample t-test conducted to determine whether the average final exam score (80.32) significantly differs from a hypothesized benchmark score of 75.

The gray dashed line represents the null hypothesis meaning (75). The dark blue line marks the observed sample mean (80.32). The orange bar shows the 95% confidence interval for the meaning: [79.58, 81.06]. The p-value is extremely small ($p < 2.2e-16$), indicating a highly significant difference between the observed mean and the hypothesized value. The final exam scores are **statistically significantly higher** than the benchmark of 75,

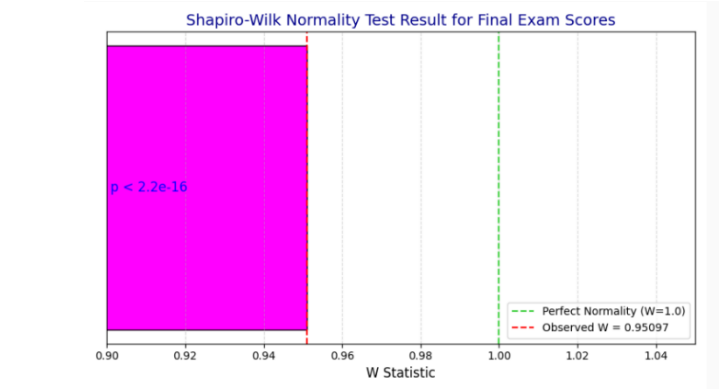

Fig. 6 Shapiro- Wilk Normality Test Result

This chart illustrates the outcome of the Shapiro-Wilk test, which assesses whether the distribution of final exam scores follows a normal distribution.

The W statistic is 0.95097, shown in red.

A green dashed line marks the ideal value for perfect normality (W = 1.0).

The p-value is extremely small ($p < 2.2e-16$), indicating a statistically significant deviation from normality.

The magenta bar and blue annotation emphasize the result visually for readers.

The data does not follow a normal distribution, which has implications for statistical analysis.

Non-parametric methods or transformations may be more appropriate for further analysis.

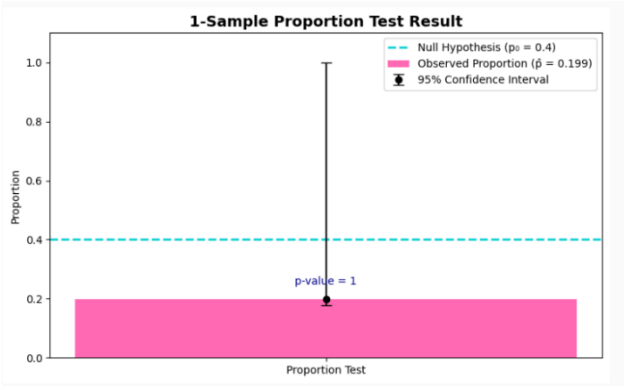

Fig. 7 Participants with commitment

This chart visualizes the result of a 1-sample proportion test evaluating whether the observed proportion of high commitment ($\hat{p} = 0.199$) is significantly greater than the hypothesized value ($p_0 = 0.4$).

The pink bar represents the observed proportion.

The dashed turquoise line marks the null hypothesis value (0.4). The black error bar shows the 95% confidence interval: [0.1786, 1.0]. The p-value is 1, indicating no statistically significant difference—the observed proportion is not greater than 0.4.

The result suggests that the proportion of participants with high commitment is lower than expected, and the test does not support the alternative hypothesis.

Fig. 8 Paired t-Test: Improvement in Digitalization

This chart visualizes the result of a paired t-test comparing participants' scores before and after training in digitalization.

The mean difference is -35.17, indicating a substantial increase in post-training scores.

The purple bar represents the mean difference, and the black error bar shows the 95% confidence interval: [-35.72, -34.62].

The p-value is extremely small ($p < 2.2e-16$), confirming that the improvement is statistically significant.

The training program had a strong and consistent positive effect on participants' digital skills.

The narrow confidence interval suggests high precision in the estimate of improvement.

These results support the effectiveness of structured digital literacy interventions.

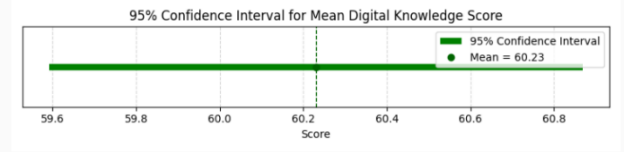

Fig. 9 Confidence Interval for Mean Digital Knowledge Score

This plot shows the 95% confidence interval for the mean score of total digital knowledge among participants:

The green horizontal bar represents the confidence interval: [59.59, 60.87].

The dark green dot and dashed line indicate the estimated mean: 60.23.

The narrow interval suggests a high level of precision in the estimate.

The average digital knowledge score is moderately high, with low variability.

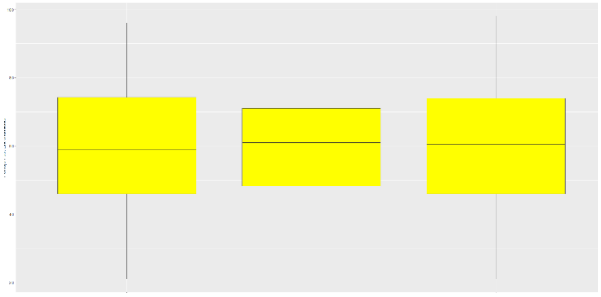

Fig. 10 ANOVA Analysis: Post-Training Scores by Digitalization in pre-school, primary or secondary teachers,

ANOVA Result

p-value = 0.0019 (less than 0.05)

This indicates a statistically significant difference in post-training scores among the three groups.

The null hypothesis (that all groups have the same mean score) is rejected.

The training program had differential effects depending on the teacher level.

The significant p-value supports further investigation into tailored training strategies for each educational level.

Since residuals are not normally distributed, consider: Using non-parametric alternatives.

Discussion:

The results of this study provide valuable insights into the digital literacy and competency levels of pre-service teachers in the context of educational digitalization, particularly in a higher education setting in northern Peru (Trujillo). The findings, derived from statistical analyses and visualizations using R, highlight both the current state of digital skills among participants and the effectiveness of targeted digital literacy interventions. Below, we discuss these results in relation to research objectives and existing literature.

Age Distribution and Participant Demographics

The right-skewed age distribution (Fig. 1), with most participants in their early 20s, aligns with the typical demographic of undergraduate pre-service teachers. This demographic is critical for studying digital literacy, as younger generations are often assumed to be "digital natives" with inherent technological proficiency [3]. However, [3] argue that digital nativity does not automatically translate to digital competence in educational contexts, a notion supported by the moderate digital literacy scores observed in this study (Fig. 2, mean score of 24.97 out of 50). This suggests that while younger pre-service teachers may be familiar with technology, they require structured training to develop pedagogical digital competencies, as emphasized by [4].

Digital Literacy Levels

The moderate average score in basic digital knowledge (Fig. 2) and the wide dispersion of scores (Fig. 3, standard deviation of 14.84) indicate significant variability in digital literacy among participants. These findings resonate with [3], who highlight that digital competencies in secondary and higher education vary widely due to differences in prior exposure to technology and training [9]. The normal distribution of scores (Fig. 3) suggests a heterogeneous population, necessitating differentiated instructional strategies to address learners at varying skill levels. This aligns with the call by [5] for tailored digital and media literacy programs that account for diverse learner needs [5].

The low mean score in digital knowledge (24.97 out of 50) underscores a critical gap in pre-service teachers' readiness to integrate digital tools into their teaching practices. This finding is particularly relevant in the Latin American context, where [12] note a lack of large-scale data on digital technology adoption in education [12]. The results suggest that teacher education programs in regions like northern Peru must prioritize digital literacy to prepare educators for digitalized classrooms, as advocated by [2].in their study on digital transformation readiness in Pakistan [2].

Effectiveness of Digital Literacy Interventions

The paired t-test results (Fig. 8) demonstrate a statistically significant improvement in digitalization scores post-training (mean difference = -35.17, $p < 2.2e-16$), indicating the effectiveness of the intervention. The narrow confidence interval ([-35.72, -34.62]) suggests high precision in estimating

the training's impact. These findings align with [1], who emphasize the role of digital platforms and tools in enhancing teacher development through structured training [1]. The substantial improvement supports the argument by [11] that digital capability training fosters sustainable behaviors and employability among university students [11]. However, the significant variability in post-training outcomes across teacher levels (Fig. 10, ANOVA $p = 0.0019$) suggests that training effectiveness varies by educational context (pre-school, primary, or secondary). This finding calls for further investigation into tailored training strategies, as recommended by [13] for promoting high-quality educational digitalization.

Final Exam Performance and Normality

The high mean score in the final English exam (80.32 out of 100, Fig. 4) and its significant difference from the benchmark of 75 (Fig. 5, $p < 2.2e\text{-}16$) indicate strong academic performance. However, the Shapiro-Wilk test (Fig. 6, $p < 2.2e\text{-}16$) reveals that the final exam scores deviate from normality, suggesting the need for non-parametric statistical methods in future analyses. This non-normality may reflect diverse factors influencing performance, such as varying levels of prior knowledge or access to resources, a challenge noted by [6] in their study on educational platformization. The high exam scores contrast with the moderate digital literacy scores, suggesting that while participants excel in traditional academic assessments, their digital competencies lag, a discrepancy that warrants further exploration in teacher education programs.

Commitment Levels

The 1-sample proportion test (Fig. 7, $p = 1$) indicates that the proportion of participants with high commitment (0.199) is significantly lower than the hypothesized value (0.4). This finding is concerning, as teacher commitment is critical for sustained digital transformation, as noted by [7] in their study on innovative teaching methodologies. Low commitment may be linked to challenges such as teacher overload and limited digital skills, as identified by [13] in the Peruvian context. Addressing this requires institutional support, such as reducing workload and providing ongoing professional development, to foster greater engagement with digital initiatives.

Confidence in Digital Knowledge

The 95% confidence interval for the mean digital knowledge score (Fig. 9, [59.59, 60.87]) indicates a moderately high level of digital knowledge with low variability, suggesting consistency in the sample's overall digital competency post-intervention. This finding supports [8], who advocate for integrating digital competencies into higher education curricula to enhance graduate employability. However, the moderately high score still indicates room for improvement, particularly in advanced skills related to big data and emerging technologies, as highlighted by [10].

Regional and Global Implications

The results highlight regional challenges in northern Peru, such as limited digital skills and infrastructure disparities, which align with broader Latin American trends identified by [12]. Globally, the findings contribute to the discourse on digital literacy as a foundational competency, as articulated by [4]. The use of big data and statistical visualization in this study underscores the potential of data-driven approaches to inform educational policy and practice, as supported by [7]. However, the non-normal distribution of exam scores and low commitment levels suggest that data literacy and teacher engagement remain critical areas for improvement.

IV. Conclusion

This study investigated the digital literacy and competencies of pre-service teachers in northern Peru, using big data and statistical visualization techniques in R to analyze their performance and training outcomes. The findings reveal a moderate level of digital literacy among participants, with significant variability in skills, underscoring the need for differentiated instructional strategies. The structured digital literacy intervention proved highly effective, as evidenced by substantial improvements in post-training scores, though outcomes varied across teacher levels (pre-school, primary, and secondary). High academic performance in the final English exam contrasted with moderate digital competencies, highlighting a gap between traditional academic skills and digital readiness.

The low proportion of participants with high commitment suggests that institutional challenges, such as teacher overload and limited digital infrastructure, hinder sustained engagement with digital transformation. These findings align with regional challenges in Latin America, where data on digital technology adoption remains scarce, and global trends emphasizing digital literacy as a foundational competency for educators.

The study demonstrates the power of big data and statistical visualization in uncovering actionable insights for educational reform. To enhance digital literacy among pre-service teachers, teacher education programs should prioritize tailored training, ongoing professional development, and institutional support to address workload and infrastructure barriers. Future research should explore longitudinal impacts of digital interventions and investigate strategies to boost teacher commitment, ensuring that educators are equipped to navigate and contribute to digitalized educational environments.

Acknowledgment

The authors express their gratitude to the Universidad Peruana del Norte for their support in facilitating the training program and other researchers to analyze the data collection. We also thank the pre-service teachers who participated in this study for their valuable contributions.

References

[1] H. Ying, P. Jingwen, and L. Mengyu, “Construction and Application of Digital Platforms, Resources, and Tools: Digital Platforms,” in *Lecture Notes in Educational Technology*, vol. Part F86, 2025, pp. 411–425.

[2] N. A. Bhutto, M. Khaliq, N. H. Ghumro, and U. S. Akbar, “Investigating the Impact of Determinants Influencing Students’ Readiness for Digital Transformation in Academic Institutions in Pakistan,” *High. Educ. Q.*, vol. 79, no. 1, 2025, doi: 10.1111/hequ.70002.

[3] O. Trevisan, R. Christensen, K. Drossel, S. Friesen, A. Forkosh-Baruch, and M. Phillips, “Drivers of Digital Realities for Ongoing Teacher Professional Learning,” *Technol. Knowl. Learn.*, vol. 29, no. 4, pp. 1851–1868, 2024, doi: 10.1007/s10758-024-09771-0.

[4] R. Rong and H. Yang, “Study on the influencing factors of digital literacy among normal students under the background of big data,” in *ACM International Conference Proceeding Series*, 2024, pp. 121–126, doi: 10.1145/3696500.3696520.

[5] D. C. Andrade Girón, M. G. Zúñiga Rojas, F. M. Lioo Jordán, V. I. Vellón Flores, F. M. Garivay Torres de Salinas, and T. Solano Armas, “Longitudinal overview of digital and media literacy: A taxonomy-based research trends analysis from 1954 to 2023,” *Iberoam. J. Sci. Meas. Commun.*, vol. 4, no. 3, 2024, doi: 10.47909/ijsmc.1424.

[6] G. Herrera-Urízar, M. Blanco-Navarro, P. Lozano-Mulet, and P. Neut-Aguayo, “Commercial or Pedagogical Interest? The Big Tech Educational Platforms and the Free Development in Childhood,” *REICE. Rev. Iberoam. Sobre Calidad, Efic. y Cambio en Educ.*, vol. 22, no. 2, pp. 67–84, 2024, doi: 10.15366/reice2024.22.2.004.

[7] G. Delzanno *et al.*, “BotBid - From Botany to Big Data: Combining Citizen Science and Innovative Teaching Methodologies,” in *Communications in Computer and Information Science*, 2024, vol. 2076 CCIS, pp. 112–127, doi: 10.1007/978-3-031-67351-1_8.

[8] M. Á. Marzal and M. C. Gómez, “Augmented Reality as an element of training in digital competencies for Higher Education: a proposal for curricular integration from biosanitary training,” *Transinformacao*, vol. 36, 2024, doi: 10.1590/2318-0889202436e249680.

[9] K. Stopar and T. Bartol, “Digital competences, computer skills and information literacy in secondary education: mapping and visualization of trends and concepts,” *Scientometrics*, vol. 118, no. 2, pp. 479–498, 2019, doi: 10.1007/s11192-018-2990-5.

[10] A. A. Amsal, “Is big data a disruptive innovation to RCEP?,” in *Regional Comprehensive Economic Partnership*, 2023, pp. 74–92.

[11] C.-H. Liu, J.-S. Horng, S.-F. Chou, T.-Y. Yu, M.-T. Lee, and M. C. B. Lapuz, “Digital capability, digital learning, and sustainable behaviour among university students in Taiwan: A comparison design of integrated mediation-moderation models,” *Int. J. Manag. Educ.*, vol. 21, no. 3, 2023, doi: 10.1016/j.ijme.2023.100835.

[12] L. R. da Cruz, J. do Rosário Diniz, T. Amiel, P. Gonsales, F. Saraiva, and V. Ramos, “The Platformization of Higher Education in Latin America and Africa,” in *Lecture Notes in Educational Technology*, vol. Part F424, 2025, pp. 435–447.

[13] P. Cao and L. Duan, “Comprehensively Promoting Educational Digitalization to Boost High-Quality Development in Jinshui Education,” in *Lecture Notes in Educational Technology*, vol. Part F107, 2025, pp. 389–396.

[14] P. E. Vernon and N. L. Gage, *Handbook of Research on Teaching*, vol. 14, no. 1. 1965.

[15] R. Millar, “Experiments,” 2014, pp. 1–2.

[16] Y. Levy and T. Ellis, “A Guide for Novice Researchers on Experimental and Quasi-Experimental Studies in Information Systems Research,” *Interdiscip. J. Inf.*, vol. 6, Jan. 2011, doi: 10.28945/1373.

[17] E. T. Koh and W. L. Owen, “Experimental and Quasi-Experimental Research BT - Introduction to Nutrition and Health Research,” E. T. Koh and W. L. Owen, Eds. Boston, MA: Springer US, 2000, pp. 196–217.

[18] L. Miller and I. Rodríguez Marín, “Estrategias experimentales para el análisis y evaluación de políticas públicas,” *Gestión y Análisis Políticas Públicas*, pp. 56–70, 2023, doi: 10.24965/gapp.11133.