European Public & Social Innovation Review

ISSN 2529-9824

Research article

Analysis of interaction with AI-assisted writing tools

Análisis de la interacción con herramientas de escritura asistida por IA

Celia Gallardo Herrerias: Universidad de Almería, Spain.
cgh188@inlumine.ual.es

Receipt Date: 17/03/2026

Acceptance Date: 19/04/2026

Publication Date: 24/04/2026

How to cite the article

Gallardo Herrerias, C. (2026). Analysis of interaction with AI-assisted writing tools [Análisis de la interacción con herramientas de escritura asistida por IA]. *European Public & Social Innovation Review*, 11, 01-16. https://doi.org/10.31637/epsir-2026-2952

Abstract

Introduction: The educational application of AI requires a responsible leadership as well as compensation of PSA, that is, significantly contributing to the students' well-being. **Methodology:** Two types of approaches have been executed: (1) the validation of the leadership scale in 470 teachers in Spain, using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA); and (2) the analysis of PSA in 55 university students in Italy & the UK, recorded via heart rate (HR) biosensors, artificial intelligence-supported facial analysis (CNN), in addition to the PSAS. **Results:** The leadership scale led to a valid four-component solution of factors - Empowerment, Orientation, Caution, & Collaboration, with excellent fit indices, namely CFI = 0.977 & RMSEA = 0.067. K-mean analysis of the students' processed physiological information yielded differences in stress levels, in which the average HR measured 149.8 bpm in the high-stress group. **Conclusion:** The effectiveness of AI adoption is contingent upon having a visionary, whereas having ethical leaders is a prerequisite rather than having mere technological accessibility.

Keywords: artificial intelligence; leadership; oratories; mental stress; learning.

Resumen
Introducción: La aplicación educativa de la IA requiere un liderazgo responsable, así como la compensación del APS, es decir, una contribución significativa al bienestar del alumnado. Este estudio examina los componentes de dicha tarea mediante validación estructural y monitorización multimodal. **Método:** Se han aplicado dos enfoques: (1) la validación de la escala de liderazgo en 470 docentes españoles, mediante Análisis Factorial Exploratorio (AFE) y Análisis Factorial Confirmatorio (AFC); y (2) el análisis del APS en 55 estudiantes universitarios de Italia y el Reino Unido, registrado mediante biosensores de frecuencia cardíaca (FC), análisis facial basado en inteligencia artificial (ACN), además del APS. **Resultados:** La escala de liderazgo arrojó una solución válida de cuatro componentes: Empoderamiento, Orientación, Precaución y Colaboración, con excelentes índices de ajuste: CFI = 0,977 y RMSEA = 0,067. El análisis de K-medias de la información fisiológica procesada de los estudiantes reveló diferencias en los niveles de estrés, donde la frecuencia cardíaca promedio fue de 149,8 lpm en el grupo de alto estrés. **Conclusión:** La efectividad de la adopción de la IA depende de contar con un visionario, mientras que contar con líderes éticos es un prerrequisito, más que la mera accesibilidad tecnológica.

Palabras clave: lnteligencia artificial; liderazgo; oratoria; estrés mental; aprendizaje.

1. Introduction

Within the context of 21st-century education, Information and Communication Technologies (ICT) have moved from being auxiliary tools to becoming the central framework of the educational process (Martínez, 2025). This technological shift is marked by the incorporation of disruptive technologies such as the Internet of Things (IoT), Artificial Intelligence (AI), and Machine Learning (Zawacki et al., 2019).

As these technologies move to the forefront of school processes, the importance of educational leadership cannot be overstated in ensuring that their incorporation is not only technological but also responsible and ethical (Holmes et al., 2022; Gallardo et al., 2021). For children with Attention Deficit Hyperactivity Disorder (ADHD), AI-mediated environments present a rare chance to overcome the usual challenges in literacy, particularly in orthography and grammar, if there is a visionary and supportive leadership to oversee the process.

Students with ADHD often experience substantial difficulties in the mechanical and cognitive processes of writing (Atturu, 2025). This is often seen in their poor spelling (orthography) and grammar (syntax), which are not necessarily indicative of their cognitive abilities but rather their executive functioning and working memory deficits (Gallardo et al., 2022).

Based on the Three Systems Model developed by Dawadi et al. (2021), students with ADHD react to stressful academic situations, such as writing, through three interconnected pathways: the physiological pathway (heart rate increase due to stress), the cognitive pathway (perception of failure), and the behavioral pathway (observable errors and task avoidance).

In the presence of grammatical complexity, the ADHD students may suffer from a condition of writing anxiety, which is similar to the public speaking anxiety (PSA) experienced by students in higher education institutions. This anxiety is a type of situation-specific social anxiety that can adversely affect their academic self-esteem and overall mental well-being (Kooli, 2025).

Although traditional teaching approaches involve manual correction, the recent development of AI-assisted writing software such as Read&Write and Co:Writer provides personalized learning assistance that is specifically designed to meet the individual neurocognitive requirements of the student (Zdravkova et al., 2022). Nevertheless, the successful utilization of such technology is often obstructed by the absence of orientation and empowerment of the educational personnel within the institution.

An important knowledge gap in the existing literature has been found to be that there are technical assessments of assistive software available, but there is a profound gap in research that employs SEM in AMOS to test the validity of institutional factors, including leadership and digital culture, that moderate the use of these software products (Bernedo et al., 2024).

Leadership research suggests that successful adoption requires not simply technological availability but visionary, supportive, and ethical leadership in order for the adoption to be successful. Existing models, including the ISTE standards, do not currently focus on the role of artificial intelligence in assisted writing for students in special education and the role of transformational leadership in this area (Melo et al., 2025).

Moreover, despite the predicted automation of almost 40% of the non-teaching tasks, like tracking educational progress, there is no empirical research available on the allocation of this freed-up time (approximately 13 hours a week) dedicated by the administration to fulfill the socio-emotional requirements of ADHD children. Without a strong validation process of the instruments used by the administration to support these technologies, the digital revolution within special education is fragmented and lacks ethical caution.

The proposed research is based on the Transformational leadership theory, where the leader focuses on the establishment of the vision, which is communicated to the entire community. The concept of AI-assisted writing is based on the following four dimensions:

1. Empowering Leaders: Encouraging faculty to utilize such tools as Co:Writer to increase the autonomy and skills of ADHD students.

2. Orientation: Offering pedagogical advice for reworking the process of writing with AI.

3. Warning: Addressing the concern of ‘algorithmic bias,’ ‘data protection’ with respect to sensitive neurodivergent data.

4. Collaboration-Culture: Encouraging a school culture that promotes experimenting with AI without fear of making mistakes.

Through the combination of these factors with a pre-post intervention analysis (ANOVA), this study aims to find not only whether the software increases literacy but also how the structural support of the school management facilitates this increase.

1.1 Objectives and Research Questions

The main objective of this research is to test a structural model (using AMOS) that connects leadership support at institutions with the academic progress of ADHD students through AI writing tools. The objectives of this research are:

- O1: Determine the effect of Read&Write and Co:Writer tools on the orthographic and grammatical accuracy of ADHD students through ANOVA comparisons.

- O2: Test the psychometric qualities of a newly developed instrument aimed at measuring the responsible use of AI in special education settings.

- O3: Investigate the correlation between physiological signs of stress (detected through heart rate sensors) and students' perceptions of themselves when working on writing tasks aided by AI.

Therefore, the following Research Questions (RQs) will be posed:

- RQ1: Does AI-assisted writing technology make a significant difference in correcting spell and grammar mistakes among students with ADHD compared with conventional approaches?

- RQ2: In what way can the variable of 'Empowering Leadership' predict the integration of assistive technology in the classroom as measured by the structural model of four factors in AMOS?

- RQ3: Is there a significant correlation between high heart rate (HR) variation (a sign of emotional regulation) and the effective utilization of AI tools among ADHD students?.

- RQ4: In what way does the ethical caution practiced by school administrators impact the sense of security as well as academic integrity of AI-generated content in special education?

Through the engagement of these questions, this research offers a contribution towards the theoretical framework of educational leadership within AI-enhanced learning spaces, emphasizing that the shift towards digital literacy is a process that should be informed by the principles of ethical stewardship and a strong awareness of neurodiversity.

2. Method

The current study made use of a non-experimental design that is ex post facto in order to analyze natural perceptions of leadership as well as interaction with AI writing support tools without experimenting.

The research methodology for this study employed an explanatory sequential mixed methods approach wherein the primary source of data for analysis was the quantitative data while the secondary source was the qualitative data.

The main subjects involved in the first phase of the experiment included a sample of 470 Spanish in-service teachers drawn from different educational levels such as preschool, primary education, secondary education, vocational training education and so forth. The diverse samples, who were aged between 19 to 55 years, offered strength in the evaluation of the hypothesis that related to the role played by school administration teams including directors, heads of studies and secretaries in promoting and performing proper guidance in the application of artificial intelligence tools for writing purposes.

In the second phase, the composition of the samples was similar to students from various countries such as the UK and Italy to compare their stress and engagement profiles in their writing exercises. The voluntary participants signed the consent form concerning data treatment due to the enhanced protection offered under the 2024 EU AI Act.

The sample size calculation a priori for the teacher sample (N=470) was conducted considering a minimum ratio of 10-15 participants per variable to be estimated within the context of structural equation modeling. The relatively small student sample (N=55) was in line with research involving experimental and physiological methods as well since multiple measurement approaches complicate large sample sizes.

The sampling of both groups was based on criterion sampling principles. The criterion of selecting the teachers was that they (a) were active in teaching, (b) had at least one-year work experience, and (c) were involved in some digital/AI learning tools (either as instructors or users). Exclusion criteria were incomplete responses or lack of informed consent. Sampling was achieved through mailing lists, professional organizations and regional education authorities in Spain. Response rate was approximately 62%, which is acceptable for such academic voluntary surveys.

In relation to the student sample (n = 55), participants for the experiment were selected using convenience sampling from student populations enrolled in collaborating universities in Italy and the UK, following specified selection criteria for the research:

(a) participation in higher education institutions;

(b) involvement in academic tasks demanding extensive writing;

(c) absence of any cardiac illness that could hinder their capacity to provide reliable HR estimates; and

(d) absence of prior exposure to the specific forms of AI being studied.

One notable psychometric instrument used to measure the institutional readiness for the deployment of AI is a self-developed 30-item questionnaire designed to estimate the attitudes of teachers in four separate domains of readiness: Empowering Leaders, Orientation, Caution, and Collaboration-Culture. Each item in the questionnaire was rated on a 7-point Likert scale, in which 1 denoted total disagreement while 7 indicated totally agree.

The development of the scale followed psychometric techniques of item generation based on theoretical constructs, expert validation of the items (N=3 experts in education leadership and AI), and item refinement prior to piloting the scale. Content validity was established based on the expert agreement exceeding 85%.

Specifically, the Empowering Leaders dimension of the questionnaire aimed at examining how leaders within these institutions assist in using AI technologies in planning (e.g., with the help of LessonPlans.ai), multimedia authoring, as well as evaluating academic achievements (e.g., through Moodle Analytics).

In turn, the Caution dimension focused on the management's readiness for addressing the issues of algorithmic discrimination, safeguarding personal information, as well as legal issues of AI usage and academic integrity.

Finally, the Collaboration and Culture dimension was developed to assess how collaboration culture is promoted among parents, teachers, and students concerning artificial intelligence.
In the preparatory stage of administering the questionnaire, a pilot study was conducted using 32 teachers as respondents. ensured the clarity and accuracy of the questionnaire items.

If there were any ambiguities in wording among the respondents with regards to their knowledge about the use of artificial intelligence ethically, corrections were made. Afterward, the questionnaire was posted online via the Qualtrics website and was completed within one month. Several strategies were employed in order to reduce common method variance.

As part of enhancing the questionnaire instrument to obtain more accurate results, a prescreening phase was implemented using a strict selection criterion. Questions with high variation (standard deviation greater than one) and meeting univariate normality criteria within the -1 and 1 range for all variables except Caution (1.5 to -1.5), according to proper leadership research practices, were retained.

The technological strategy in observing the students' use of the Read&Write and Co:Writer software was conducted in accordance with the Three Systems Model, which comprised the Physiological, Cognitive, and Behavioral levels.

Data were collected through classroom experiments that took around one hour. Environmental variables such as light, ambient noise, and seating arrangement were kept constant throughout the experiment to rule out any effects resulting from changes in the surrounding environment. The class followed this structure:

1) briefing and signing of the informed consent form,

2) measurement of heart rate for five minutes,

3) filling out of the PSAS survey questions,

4) doing the designated writing task with the assistance of AI, and

5) debriefing of participants after completion of the writing task.

Participants were required to write an argumentative essay on any non-controversial topic ranging from 250-300 words.

Physiological level involved the use of a Polar OH1+ together with Optical Heart Rate (HR) Monitor and Polar Flow Online App. The device made it possible for students to record their heart rate where 60 heart rate measurements were recorded per minute which were later used to analyze the stress levels based on average, minimum, maximum, and standard deviation of HRs (also known as fluctuations). HR fluctuations indicated that individuals were able to regulate their emotions which is a particularly challenging task for individuals with ADHD.

Prior to all experiments the calibration process of the instruments was conducted to make sure that the measurement results would be valid. The validity of Polar OH1+ device has been proven by many physiological studies conducted before.

The Behavioral level included recording of participants during the usage of the tools through the means of video. The collected data were then analyzed through a machine learning emotion detection algorithm utilizing CNN (Convolutional Neural Networks) and OpenCV library for Python programming language.

Within this process, the use of a segmentation technique was used to detect emotions in relation to each student's facial features took place; this segmentation provided percentage scores for emotions such as happiness, neutrality, fear, sadness, anger, and surprise. In this regard, attention was given to code happiness, in stressful writing conditions for instance, nervous smile as an act to relieve social tension.

As for emotion detection, the pre-trained machine learning model was fine-tuned using a validated dataset (FER-2013) for the purpose of increasing the model's accuracy in educational environment. Machine learning model emotion recognition accuracy was over 85%.

As for cognitive measurement, before the onset of writing behavior, participants were subjected to the Public Speaking Anxiety Scale (PSAS) measuring anxiety based on self-reporting of the degree of stress, such as palpitation, nausea, or terror, among others. According to the findings of Cronbach’s Alpha, the test exhibited a high degree of internal consistency in contemporary research ($\alpha > 0.90$).

The process of synchronization was carried out for all physiological and behavioral data through matching the time stamps. Preprocessing procedures included noise removal, normalization, and the identification of outliers based on the z-score (±3). Missing data under the mean substitution technique constituted less than 3%.

Analysis of the data entailed a stepwise process and application of multiple algorithms to generate impactful results. To enhance the quality of the analysis, it was ensured that all statistical assumptions were thoroughly tested prior to estimating the models.

The first step involved determining the validity of constructs of the leadership questionnaire tool using Exploratory Factor Analysis (EFA) through principal axis factor and oblimin rotation. For the sake of ensuring that the sample acquired for data collection is valid and appropriate for EFA, Kaiser Meyer Olkin, whose value was greater than 0.8, but specifically 0.975 in the base study, was required.

Another condition that had to be satisfied before conducting the test was the Bartlett's test of sphericity whose value needed to be statistically significant. In addition, after the completion of EFA test, it was imperative to conduct Confirmatory Factor Analysis (CFA) through AMOS graphical tool since there exist a four-factor model. In determining multivariate normality, use of Mardia's coefficient ensured the value was less than 360.

Moreover, the fit indices were used to evaluate the model, ensuring that the value of relative chi-square is <3.0 among other criteria like CFI>0.95, TLI Index>0.95, and RMSEA<0.08. The reliability of the constructs also confirmed by use of Cronbach's alpha, McDonald's Omega (ω), and composite reliability (CR), which ought to be above the cutoff of 0.70-0.80.

Moreover, multigroup invariance analysis was done to asssess consistency concerning gender (male/female) and location (rural/urban) by comparing both the models (unconstrained and constrained) and proving that the difference between the CFIs was less than 0.01.

In this research, an unsupervised learning technique was used to group student interactions into clusters of engagement and stress. Physiological data collected during the process were scaled using the MinMaxScaler feature of the scikit-learn package. Silhouette Analysis was used to compare the performance of three clustering techniques: K-means, Mean Shift, and Gaussian Mixture Models. It was found that K-means was more efficient since the best number of clusters was determined using the elbow method.

The clusters were correlated with the AI-derived emotions and PSAS self-reports on different levels of stress (relaxed, low stress, high stress, and very high stress), generated at various times in the use of writing assistance software. In order to measure how effectively Read&Write and Co:Writer narrow down the orthographic and grammatical differences, an Analysis of Variance (ANOVA) test was designed and planned to be run on the pre-test/post-test differences in error rates, thereby which provided an objective measure of the narrowed gap in literacy as a result of the current AI-based strategy.

In addition to ANOVA, post hoc comparisons using Tukey HSD were run for further inference. With this holistic approach involving structural equation modeling along with biometric data collection through the Internet-of-Things (IoT), the study enabled a thorough evaluation of the influence a particular educational establishment had on the emotional and cognitive development of children with ADHD.

In addition, parallel analysis was conducted to confirm the adequacy of factor retention criteria, thus providing a more reliable method than eigenvalues. Convergent validity was evaluated based on average variance extracted (AVE > 0.50), whereas discriminant validity was assessed through the Fornell-Larcker criterion. In this connection, an unsupervised learning algorithm was used to identify engagement and stress categories. Cluster stability was achieved by applying bootstrap resampling techniques along with different random seeds.

Ethical guidelines were strictly followed regarding data minimization, secure data storage, and de-identification to avoid any risks for respondents.

3. Results

The primary phase of the analysis was on verifying the 30-item scale meant for the measure of the level of support of institutions regarding the use of AI tools. The study indicated high variability on each of the items analyzed; the SD was beyond the recommended 1.0. This indicated that the scale managed to note the perceptions appropriately. On the analysis of the items meant to undergo structural equation modeling on AMOS, univariate normality tests of S & K were considered.

Based on the criterion (-1, 1) set for dimensions A, B, and D, as well as C, many items were eliminated to avoid distortion in the results. Items to be eliminated included A1, A3 through A6, A8, and items A10 through A11, and A13, as well as B3, B10, D3, and D7, because their scores were beyond the normality range. In Dimension C (Caution), however, a relaxed criterion of (1.5, -1.5) was used, as in other studies related to applied leadership, to ensure items C1 to C8 were retained. This formed a strong statistical foundation for the next step, which is factor analysis.

The discriminant validity of the remaining items had been ascertained via corrected item-total correlation analysis. All items had excellent discriminant validity, with values above 0.79, which is considerably greater than the minimum requirement of 0.40.

Additionally, Cronbach's alpha value on a global scale had a value of 0.977; moreover, no-item deletion had contributed toward increasing scale reliability, thereby warranting retention of the final item set.

The construct validity of this instrument was determined with regard to its factor validity by exploratory factor analysis with principal axis factorization with oblimin rotation. The appropriateness of this sample has been ascertained with a value of Kaiser-Meyer-Olkin of 0.975, indicating a very suitable sample. Bartlett's Test of Sphericity was found statistically significant, indicating that the correlations between items are adequately high for factor extraction.

The result from EFA extracted a four-factor structure with eigenvalues greater than one, explaining a total of 86.20% of the variance. Factor 1 (Orientation) explained most of the variance, which is 73.05%, followed by Factor 2 (Caution, 6.26%), Factor 3 (Collaboration and Culture, 4.18%), and Factor 4 (Empowering leaders, 2.71%).

All factor loadings were well above the threshold of 0.40. Several items in the Orientation and Caution dimensions also gave exceptionally strong loadings; for example, B2 loading was at .973, and C3 loading was at 1.015. Moreover, Oblimin rotation was further supported by significant correlations found between latent factors, such as the .793 correlation between Orientation and Collaboration.

Several CFAs were estimated in AMOS to validate the latent factors identified in the EFA. Because the first model did not reach the recommended fit indices, a second refined model was estimated. During this phase, 11 items that showed excessive covariance were removed to enhance the parsimony and fit of the instrument without loss of reliability: B6, B9, B12, B13, B5, B11, C8, C7, C5, D5, and D9.

The final 18-item model had an excellent absolute, incremental, and comparative fit index. The relative chi-square value (χ / df) was 3.122, and Root Mean Square Error of Approximation (RMSEA) was 0.067, which indicated that there was an excellent fitness level for this model too.

Comparative fitness index values were equally excellent, which was 0.977 for Comparative Fit Index (CFI) and 0.973 for Tucker-Lewis Index (TLI), exceeding .95, which makes it an excellent measure for fitness analysis. In Multivariate normality, Mardia's co Index (123.963) was below 360, which makes this measure valid for an 18-item scale.

In the final or best-fit model, the standardized regression weights for all indicators were ranging between .83 and .96, which can be observed in the structural graph below.

To address the first objective related to the improvement in spelling, grammar, and their measurement, an Analysis of Variance (ANOVA) was employed to determine the difference in the scores of students with ADHD before and after the intervention, while using AI assistance in writing.

There was a positive change in the number of orthographic mistakes made by students ($p < 0.05$) since the mean before (12.38) was lower compared to the mean after (18.3). There was also an improvement in the grammatical mistakes ($p < 0.05$) made by students, with the mean before (7.12) being lower than the mean after (11.27). Students with ADHD utilizing the word prediction, text-to-speech functions.

Moreover, the data suggested that the Orientation dimension within the leadership measure played a significant moderating role in these findings. In institutions with a greater level of pedagogical support among the teaching staff regarding the use of AI, students with ADHD perceived significantly greater gains within academic self-esteem and writing fluencies than institutions with less support.

The analysis of stress levels among students when performing writing assignments employed the Three Systems Model of writing with a focus on physiological, cognitive, and behavioral systems. The HR data collected from the Polar OH1+ devices was normalized with a normalization technique dubbed MinMaxScaler. The K-means clustering model with a Silhouette value of 0.55 accuracy made it suitable and selected four groups of students according to their conditions of heart rate HR conditions:

1. Relaxed Cluster: Average HR of 85.4 bpm, which is within the standard resting range, along with a low average change in HR (SD of 8.2). These subjects employed AI systems as personalized learning assistance without showing considerable anxiety.

2. Low Stress Cluster: Presented an average HR of 99.8 bpm along with high variation (SD = 15.4), indicating effective management of emotions.

3. High Stress Cluster: Exhibited mean HR of 116.4 bpm with high SD of 17.2. They reported the experience of heart pounding during the activity of writing. This corresponds with their PSAS survey results.

Highly displayed extreme values for HR, averaging 149.8 bpm. Interestingly, 33.3% of behavioral data for this group was coded as happy by the AI emotional parser. This was taken as indicative of nervous smiling: a means of dealing with a stressful situation.

Self-awareness analysis produced some fascinating results. For the Middlesex students, the cognitive awareness measured through the PSAS questionnaire (for example, Q13: My heart pounds) demonstrated high correlation with actual heart rate. But for the Italian case (Unimc), students belonging to the very high stress category (heart rate of 149.8 bpm) responded relatively casually to stress, with only 20% agreeing that the task is terrifying. This implied that some degree of habituation to the oral exam may be reducing their perceived risk despite prominent physiological indicators of extreme stress.

The instrument's internal consistencies were validated for each of the four dimensions, with Cronbach alpha and McDonald's Omega measures ranging from 0.94 to 0.98. Moreover, multigroup invariance tests were performed in order to validate that each model was essentially identical across different groups of individuals.

The findings supported that measurement invariance across gender and territory was valid, with a delta CFI value between unconstrained and constrained models smaller than 0.01. This further ascertains that this instrument and its predictive results remain essentially identical across educational settings, regardless of whether they are situated in an urban or a rural setting.

4. Discussion

Based on the findings of this study, there is empirical evidence of the strong link between transformational leadership, AI-based support for literacy, and biometric monitoring of neurodivergent students (Iannone et al., 2023). By juxtaposing the findings of this study with existing studies, it is possible to establish the extent to which institutional support and biometric data validate the effectiveness of technology-based interventions for students with ADHD.

The marked reduction in errors in both orthography and grammar indicated by the ANOVA analysis attests to the effectiveness of AI-based software such as Co:Writer/Read&Write as a useful individualized learning tool. However, as indicated in this study's results, this educational success can be greatly moderated by a particular aspect of institutional leadership, namely Orientation.

In this case, this conclusion matches that drawn by Guisasola (2024), who argue that in order to ensure successful implementation of disruptive technology, one requires not only technological infrastructure but visionary leaders. The marked correlation between teaching guidance and academic success in this data set verifies the source's claim that a balance between strategic guidance and actual teaching by leaders can ensure that technology integration can be both valuable and a long-term success (Bernedo et al., 2024).

Finally, in this case, the observation that Orientation and Collaboration-Culture are highly interrelated in this AMOS model verifies the hypothesis that a successful leader must have a shared vision with his or her staff in which innovation such as AI can be adopted without fear of failure or repercussions (Checa et al., 2025).

The application of the K-means algorithm in the identification of the group with higher stress levels and their average heart rate of 149.8 bpm gives the Three Systems Model credence as proposed by Dawadi et al. (2021). Just as public speaking anxiety in Guisasola (2024), the writing tasks for the ADHD students seem to cause situation-specific social anxiety, which occurs in the form of physiological responses.

The large value of the standard deviation of the HR in the Low Stress group shows the regulation of one's emotions, which, as indicated in the literature, relates to the reduced levels of anxiety and the ability to manage the task. It, therefore, means that tools such as the Polar OH1+ are essential in measuring the hidden levels of stress in the ADHD students, which would not be measured using observation (Zdravkova et al., 2022).

One of the most important aspects of this discussion is the discrepancy between AI facial expression recognition and actual physiological response. In our very high stress category, the AI-based emotion recognition software labeled 33.3% of facial expressions as being happy in situations where HR values are high. As cited in literature, this discrepancy has long been referred to as nervous smiling, a coping mechanism adopted by students to manage social stress in a high-pressure or problematic social setting.

The aforementioned observation illustrates the fallibility of AI-based emotion recognition software and re-emphasizes the need to incorporate ethical caution and critical oversight as essential components of educational administration (Holmes et al., 2022).

As indicated in literature, even recognizing and measuring social-emotive conditions in a social setting is a rather complicated task, where any measurement done purely on observation might lead to a misconception about a neurodivergent individual's actual emotional response to a setting (Catania & Garzotto, 2023; Pino et al., 2021).

Finally, the high loadings on the caution dimension in our model provide further evidence to support the argument that school leaders must demonstrate ethical stewardship. The application of AI technology among students with ADHD requires an active strategy to overcome algorithmic bias and data security to ensure equality in education.

Our results confirm that leaders’ emphasis on academic integrity and copyright (Dimension C) promotes trust-building among the community (Dimension D), and this is the only feasible way to successfully implement AI. This supports the argument that AI must facilitate, not dictate, education, and this is only possible through ethical transformational leadership.

5. Conclusions

The first prominent outcome of the conducted research is the success of the validation of the designed instrument relative to the measurement of transformational and managerial leadership in the context of AI and GenAI. It has been found to be true that the readiness of an institution for digital transformation takes multiple forms and has four distinct yet interrelated aspects such as Empowering Leaders, Orientation, Caution, and Collaboration Culture.

The structure is empirically certain through Confirmatory Factor Analysis in AMOS. The fit indices-the 0.977 CFI, 0.977 IFI, and 0.067 RMSEA-all indicate that the model has an excellent fit, passing even the most conservative scientific thresholds.

Besides, the confirmation of multivariate normality by Mardia’s coefficient of 123.963 and the evidence on measurement invariance across gender and geographical territory, urban vs. rural, make this framework a stable and generalizable tool for the diagnosis of leadership readiness in diverse educational settings. Therefore, the implication of this analysis is that unless schools are ready to engage in more than the technical procurement of AI, their leadership should develop special behavioral expressions capable of inspiring faculties toward the implementation of ethical oversight.

The key conclusion from the Empowering Leaders dimension is that the effective adoption of AI depends not only on access to technology but also on farsighted and supportive leadership. In this perspective, transformational leadership is about building, sharing, and sustaining a collective vision that will catalyze significant digital changes.

It can be inferred from the data that teachers appreciate their leaders for being able to promote professional agency and intrinsic motivation in the team. To this effect, by promoting autonomy enhancement practices, school administrators can in turn encourage teachers to feel comfortable in an experimenting environment without fear of error.

Such an approach helps teachers shift from being mere adopters of technology to being innovators in the classroom who can use AI for the purposes of rejuvenating learning by means of AI-induced redesigning in the educational process.

It is observed from the research study that the need for leadership beyond visionary rhetoric includes guidance as a part of facilitation. The Orientation component that accounted for a significant amount of variance (73.05%) in the exploratory analysis shows the need for leaders to offer certain guidelines among other things.

Teachers should be guided on how to use the technical aspects of tools such as ChatGPT or Moodle Analytics and also on the non-technical aspects of integrating these in the class. When this comes from the management teams of the organization, it boosts the teacher's esteem, ensuring that the aspect of integrating technologies in class is of added value. This means that Orientation is the link that ensures the success of the technological strategy in class.

Among the most important findings of this study is the need for the ethical caution that is necessary in the role of educational administration in the modern educational setting in the light of the rapid implementation of AI in educational settings at this stage.

The study concludes that school administrators must be ethical guardians who emphasize ethics such as equity, transparency, and responsibility. Through the integration of ethical watchfulness as part of its essential operations, administrators can prevent unintended harm while promoting the integration of AI systems without undermining educational equity. This is even more significant when working with sensitive data related to emotion recognition or biometrics, as it reinforces the importance of balancing technological strength with ethical regulation.

Moving on further to the level of the student, the implications of the findings include that for public speaking anxiety (PSA) engagement, it is crucial that the concept of PSA engagement adopt a multi-modal paradigm that incorporates the Three Systems Model (physiological, cognitive, and behavioral). It can lead to MISrepresentation of pupil perceptions of items within their actual self when only one source of data, for instance, self-report data for assessment within a college course, is considered.

Integration of wearable IoT technology (Polar OH1+) and AI-powered emotion recognition using CNN analysis demonstrated that physiological responses, such as a sudden increase in heart rate (HR), may be indicative of stress, which is either repressed or not acknowledged by students. Analysis of the study showed that a large fluctuation of HR can be considered a reliable indicator of regulated emotional responses, which would lead to a correct diagnosis of student stress related to oral presentations.

K-means unsupervised machine learning demonstrates that the research can indeed identify discrete student profiles related to levels of stress. The identified clusters, relaxed, low stress, high stress, and very high stress, provide a predictive framework for determination of students at risk from disengagement or academic failure due to anxiety.

Indeed, it was derived from these findings that students in the very high stress cluster reached average heart rates of 149.8 bpm, indicating states of extreme physiological arousal. Thus, it can be concluded that AI/wearables can help teachers in the design of personalized training pathways by pointing out those students in need of specific counseling or coaching to overcome their fears.

A striking implication for the limitations of AI-based emotion recognition emerged: behavioral data can be misleading. the ai models often labeled students in the cluster very high stress as happy, with 33.3% of their expressions classified that way.

It concludes that this is often a case of nervous smiling, a social strategy used to manage tension in uncomfortable situations. That creates a very important point: recognizing and measuring emotions in social contexts remains extremely challenging and complex. Therefore, AI-driven behavioral analysis should always be cross-checked with physiological data, to prevent misinterpretation and ensure that a student's distress is not mistaken for positive engagement.

A significant gap in student self-awareness was revealed in the comparison between the results of PSAS and sensor data - answering RQ2. Some cohorts showed good consistency; others - especially the Italian sample-reported that they were not terrified whereas they demonstrated very high physiological stress.

This therefore suggests that students may tend to mask their weakness or fragility, or habituate themselves to stressful assessment situations, such as oral examinations. For this reason, the availability of physical, cognitive, and behavioral information concerning students can be useful to raise the awareness of individual weaknesses and strengths, hence a more transparent and more effective learning process.

The research recognizes that AI is likely to automate about 40% of non-teaching tasks, such as tracking student progress. This would free up the equivalent of about 13 hours a week for teachers.

The basic inference here is that such freed-up time should be channeled into the human aspects of education, namely, inspiring in students a quest to learn through the vehicle of personal relationships, along with socio-emotional support. AI should be regarded merely as an instrument of optimization for decision-making and classroom management; the human dimension of adopting technology remains the most critical factor for sustainable educational transformation.

Finally, the longitudinal process for schools' digital transformation and cross-cultural validation of findings lead to the final conclusion. The systematic design methodology fills a significant methodological gap in the area through the use of a validated leadership instrument combined with a multi-modal monitoring protocol.

- Replication of the validation of leadership models in varied cultural and organizational contexts.

- To study how perceptions of leadership change as AI becomes embedded in school culture.

- Exploiting the tools in real diagnosis on a real educational scenario to capture meaningful differences across leadership practices.

- Designing AI-powered interventions to help students surmount Public Speaking Anxiety based on their clustering profiles.

6. References

Atturu, H. (2025). Effectiveness of an artificial intelligence–based platform in supporting children with autism spectrum disorder. *JMIR Neurodiversity*, *2*(1), 70-89. https://doi.org/10.2196/70589

Bernedo, D. H., Loayza, Y. T., Gonzales, J. N., Saavedra, J. K., Barrientos, A. R. y Romero, R. (2024). The role of artificial intelligence in the adaptation of students to virtual educational environments. *Metaverse Basic and Applied Research*, 3, 124. https://doi.org/10.56294/mr2024124

Catania, F. y Garzotto, F. (2023). A conversational agent for emotion expression stimulation in persons with neurodevelopmental disorders. *Multimedia Tools and Applications*, 82, 12797-12828. https://doi.org/10.1007/s11042-022-14135-w

Checa, L., García, I., Gallardo, C. P. y Cambil, L. (2025). Alfabetización en ciencias computacionales de futuros docentes para atender a la diversidad. *Edutec*, 92, 55-70. https://doi.org/10.21556/edutec.2025.92.3687

Dawadi, S., Shrestha, S. y Giri, R. A. (2021). Mixed-methods research: A discussion on its types, challenges, and criticisms. *Journal of Practical Studies in Education*, *2*(2), 25-36. https://doi.org/10.46809/jpse.v2i2.20

Gallardo, C. d. P., Caurcel, M. J., Crisol, E. y Jarque, S. (2021). Assessment of apps aimed at developing basic instrumental skills in autistic children and teenagers. *Mathematics*, *9*(9), 1032-1043. https://doi.org/10.3390/math9091032

Gallardo, C. d. P., Rodríguez, A., Caurcel, M. J. y Capperucci, D. (2022). Functionality of apps for people with autism. *International Journal of Environmental Research and Public Health*, *19*(12), 7019-7034. https://doi.org/10.3390/ijerph19127019

Guisasola, J. (2024). La investigación basada en el diseño: algunos desafíos y perspectivas. *Revista Eureka*, *21*(2). https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2024.v21.i2.2801

Harkins, A., Carling, L. y Peloff, D. C. (2025). Artificial intelligence in special education. *Encyclopedia*, *5*(1), 11-28. https://doi.org/10.3390/encyclopedia5010011

Holmes, W., Porayska, K. y Holstein, K. (2022). Ethics and artificial intelligence in education. *British Journal of Educational Technology*, *53*(6), 1559-1578. https://doi.org/10.1007/s40593-021-00239-1

Iannone, A., Martínez, L. y Gómez, P. (2023). Breaking barriers: The intersection of AI and assistive technologies in autism. *Journal of Assistive Technology*, *17*(2), 107-125. https://doi.org/10.3390/jpm14010041

Kooli, C. (2025). AI-driven assistive technologies in inclusive education. *International Journal of Education and Development Using ICT*, *21*(1), 1-17. https://doi.org/10.1016/j.sftr.2025.101042

Kotsi, S., Handrinou, S., Iatraki, G. y Soulis, S. (2025). A review of artificial intelligence interventions for students with Autism Spectrum Disorder. *Disabilities*, *5*(1), 7-23. https://doi.org/10.3390/disabilities5010007

Martínez, M. A. (2025). Inteligencia artificial y educación. *Revista Tecnológica-Educativa Docentes 2.0*, *18*(1), 245-257. https://doi.org/10.37843/rted.v18i1.614

Melo, V., Basantes, A., Gudiño, C. y Hernández, E. (2025). The impact of artificial intelligence on inclusive education. *Education Sciences*, *15*(5), 539-557. https://doi.org/10.3390/educsci15050539

Mohammed, M. A., Alyahya, S., Mukhlif, A. A., Abdulkareem, K. H., Hamouda, H. y Lakhan, A. (2024). Smart Autism Spectrum Disorder learning system based on remote edge healthcare clinics and internet of medical things. *Sensors*, *24*(23), 7488. https://doi.org/10.3390/s24237488

Pino, M. C., Vagnetti, R., Valenti, M. y Mazza, M. (2021). Comparing virtual vs real faces expressing emotions in children with autism: An eye-tracking study. *Education and Information Technologies*, *26*(5), 5717-5732. https://doi.org/10.1007/s10639-021-10552-w

Zawacki, O., Marín, V., Bond, M. y Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, *16*(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zdravkova, K., Krasniqi, V., Dalipi, F. y Ferati, M. (2022). Cutting-edge communication and learning assistive technologies for disabled children: An artificial intelligence perspective. *Frontiers in Artificial Intelligence*, 5, 970430. https://doi.org/10.3389/frai.2022.970430

Financiación: Esta investigación no recibió financiamiento externo.

AUTORA:

Celia Gallardo Herrerias
Universidad de Almería, Spain.
cgh188@inlumine.ual.es

Orcid ID: https://orcid.org/0000-0001-5515-1269
Google Scholar: https://acortar.link/3AlQ3V