Communication Technologies in Education

Proceedings of ICICTE 2025

Editors

Evangeline (Litsa) Marlos Varonis

& Anastasia (Nancy) Pyrini

Communication Technologies in Proceedings of ICICTE 2025

Editors

Evangeline (Litsa) Marlos Varonis
and Anastasia (Nancy) Pyrini

Communication Technologies in Education: Proceedings of ICICTE 2025
ICICTE 2025, Chania, Greece, July 8 to 10, 2025

Conference Director: Nancy Pyrini
Chair of the Scientific Committee: Ġorġ Mallia
Chair of the Communications Committee: Marcia Håkansson Lindqvist
Editors of the Proceedings: Evangeline (Litsa) Marlos Varonis and Anastasia (Nancy) Pyrini
Graphic Design: Ġorġ Mallia

Conference Organiser: PARAGON-eduTech

Citation of individual papers in APA Style:
Lastname, FirstInitial. (2025). Title of paper. In E. Varonis and A. Pyrini (Eds.), *The International Conference on Information Communication Technologies in Education (ICICTE 2025) Proceedings*, pp. n1-n2. ICICTE.

Editorial Board

CONTENTS

Preface to the ICICTE 2025 Proceedings: At the Intersection of Emerging Technologies and Education

Evangeline (Litsa) Marlos Varonis
Co-editor, ICICTE 2025 Proceedings

The 2025 ICICTE Conference brings together an exciting and diverse collection of research at the intersection of emerging technologies and education, with a particular emphasis on generative artificial intelligence (GenAI), immersive tools, and digital literacy. As digital transformation accelerates across the educational landscape, these proceedings capture the global dialogue on how educators, researchers, and learners are experimenting with—and adapting to—the evolving role of AI and interactive technologies in teaching, learning, and leadership.

Several contributions examine **the practical and ethical integration of GenAI into educational environments**, including studies on ChatGPT's use in formative assessment, its accuracy across languages, and student and teacher attitudes toward its implementation. Together, these papers explore the balance between AI's promise to support learning and the challenges it poses to academic integrity, trust, and equitable access.

Other authors investigate the role of **pedagogical design** in harnessing emerging technologies—from storytelling and comics as tools for AI education to immersive virtual reality aligned with higher education learners' emotional and cognitive needs. These contributions emphasize not only the power of novel tools but also the importance of thoughtful, student-centered design.

A strong thread of inquiry also runs through these proceedings around **digital citizenship, critical thinking, and media literacy**, with game-based and narrative-based tools proposed to combat misinformation and engage students in complex societal issues. Parallel to this, school leaders and head teachers from around the world share insights into the infrastructural, professional development, and leadership challenges of embedding technology into everyday practice.

Taken together, this volume offers a rich, international perspective on the future of digital education. It showcases a shared commitment to exploring the affordances of technology—while remaining critically attuned to the values, structures, and human connections that must underpin its use.

Each of these exciting studies is summarized in one sentence below, in order of their presentation at the conference.

Visual Narratives that Teach: The Role of Print and Digital Comics as Educators, *Ġorġ Mallia, University of Malta, Malta,* explores the educational power of comics as hybrid visual-textual narratives that foster engagement and deep learning across a variety of subjects and student age groups.

Aligning the Affordances of Immersive Virtual Reality with Educational Objectives through the IVRPM, *Daniela Rocha Bicalho and João Piedade (Universidade de Lisboa, Portugal),* examines how immersive virtual reality (IVR) can enhance learning in higher education by aligning its unique affordances—such as immersion, embodiment, and interactivity—with students' cognitive and emotional needs.

AI Education Through Storytelling, *George Gadanidis (Western University) and Janette M. Hughes (Ontario Tech University)*, presents an innovative approach to AI education through storytelling, using graphic narratives to engage students with the ethical, societal, and technical dimensions of AI in accessible and meaningful ways.

From the Board to the Mind: The Role of Modern Board Games in Fostering Computational Thinking in Primary Education, *João Piedade and Fábio Machuqueiro (UIDEF, Instituto de Educação, Universidade de Lisboa, Portugal)* demonstrates that modern board games can significantly enhance primary students' computational thinking by engaging them in strategic, rule-based gameplay that promotes abstraction, logic, and problem-solving.

Communication and Engagement Are Key: Academics' Views of Opportunities and Challenges with Fully Online Asynchronous Teaching, *Christine Armatas and Vikki Pollard (Australian Catholic University, Australia),* explores academics' experiences teaching asynchronous, fully online university courses, revealing that while professional development supports were helpful, challenges around student engagement, communication, and workload significantly impacted their teaching effectiveness.

Promoting Critical Thinking through the Newspiracy Project: Insights from Greece, *Konstantinos Karampelas (University of the Aegean, Greece), Anastasia Pyrini, and Georgios Sarrigeorgiou (PARAGON-eduTech*, Greece), *Konstantinos Tsolakidis (University of the Aegean, Greece) and Despina Sarrigeorgiou (PARAGON-eduTech*, Greece), evaluates the effectiveness of "Newspiracy," a digital game-based learning tool designed to enhance students' critical thinking and media literacy regarding misinformation and conspiracy theories by engaging students in a controlled, reflective manner.

Attitudes of Head Teachers toward the Use of Digital Educational Materials in Greece*, Konstantinos Karampelas, Nikolaos Raptis, and Maria Kouroutsidou (University of the Aegean, Greece),* explores Greek head teachers' attitudes toward digital educational materials, finding mostly positive perceptions but also highlighting barriers such as inadequate training and infrastructure.

Evaluating GPT-4's Proficiency on Norwegian Exams and Tests—And Exploring the Broader Implications for Educational Practice, *Rune Krumsvik and Lise Jones (University of Bergen, Norway),* evaluates GPT-4's performance on a wide range of Norwegian-language exams, finding a 94.3% average accuracy and highlighting the model's strong potential as a multilingual, multimodal educational tool for formative and summative assessment.

Evaluating ChatGPT's Effectiveness in Formative Assessment Practices for CS Education: A SWOT Analysis, *Jacqui Chetty, University of Birmingham, United Kingdom,* uses a SWOT analysis to evaluate ChatGPT's role in formative assessment for computer science education, finding it a useful supplemental tool for learning and self-reflection, though limited by accuracy concerns and ethical risks.

Towards Effective Automated Grading in CS1: A Comparison Between GenAI and the In-House Grading Tool, *Pieter Joubert, Wendy Yanez-Pazmino, and Jacqui Chetty (University of Birmingham, United Kingdom),* compares a GenAI grading tool using OpenAI's API with a traditional automated grading system in an introductory programming course, finding that GenAI can grade formative assessments with comparable accuracy and consistency—especially when provided with detailed prompts—while offering potential benefits for scalability and efficiency.

Student Experiences with ChatGPT in Higher Education: Insights from a Two-year Global Study, *Aleksander Aristovnik and Matej Ravšelj (University of Ljubljana, Slovenia)*, presents the results of a large-scale, two-year global study that reveals a significant rise in student adoption of ChatGPT for academic tasks, with increasing satisfaction and integration into daily study routines, reflecting its normalization in higher education.

Generative Artificial Intelligence in Primary and Secondary Education in Portugal: Acceptance and Use by Students *(Nuno Dorotea and Célia Ribeiras (University of Lisbon, Portugal),* investigates the factors influencing the acceptance and use of Generative AI among Portuguese primary and secondary students using the UTAUT2 model, revealing that habit, performance expectancy, hedonic motivation, and personal innovation significantly shape students' intention and frequency of GenAI use.

Artificial Intelligence (AI) and School Leadership—School Leaders' Reflections on Professional Use of AI, *Marcia Håkansson Lindqvist (Mid Sweden University, Sweden) and Fanny Pettersson (Umeå University, Sweden),* examines how Swedish school leaders perceive the integration of AI in education, emphasizing the need for redefined professional digital competence to ensure ethical and effective use in order to benefit the school as an organization.

Development of a Theoretical Framework for Self-evaluation of Adaptive Digital Learning Platforms Based on Artificial Intelligence: A Systematic Review, *Ana Pedro, Nuno Dorotea, and Bárbara Acevedo (University of Lisbon, Portugal),* proposes a theoretical framework for the self-evaluation of AI-driven adaptive learning platforms, highlighting the importance of real-time personalization, self-assessment, and ethical implementation in improving educational effectiveness.

Innovative Technologies in Ethical Expertise: Architecture and Functional Capabilities of the System, *Gulmira Bekmanova, Assel Omarbekova, Dinara Kabdylova, Aiganym Bessembayeva, Alina Yerbolatova, Altanbek Zulkhazhav, and Lena Zhetkenbay (L.N. Gumilyov Eurasian National University, Kazakhstan),* presents the development of an AI-powered system for ethical expertise and researcher training, featuring automated video lectures, adaptive testing, expert evaluation, and a scalable microservice architecture to enhance transparency and personalization in research ethics education.

The Challenges That Artificial Intelligence Brings to Australian Transnational Programs, *Kathy Michael (Victoria University, Australia)*, investigates how AI technologies affect academic integrity and pedagogy in Australian transnational programs in China, offering strategies to uphold standards without disrupting cross-border partnerships.

VISUAL NARRATIVES THAT TEACH: THE ROLE OF PRINT AND DIGITAL COMICS AS EDUCATORS

Ġorġ Mallia
University of Malta
MALTA

Abstract

This short paper looks at the nature of comics, how they are natural inducers of affinity in children, and how, as a result, they can be used for teaching and learning. Comics are a hybrid medium made up of visual sequential illustration and a text-based narrative that complements it. In the unique, symbiotic nature of the two interlaced genres, enhanced by other add-ons, comics become a very powerful method of communicating concepts and ideas. This paper examines the nature of comics and how they can be used as pedagogical tools that insidiously communicate.

Introduction

There has never been a satisfactory definition for comics. And the reason is quite simple. Because of the uniqueness of the comics genre and the incredible stylistic versatility of the multitude of comic artists that have enriched the medium with their talents, it makes it very difficult to wholly define it in ways that encompass all that it is. I suppose, for want of something better, we will need to go with one of the first definitions, and one written by the first person to put comics on the academic podium, Will Eisner, the ground-breaking artist who was one of the first to realise just how much could be done with the genre.
In a seminal book on the origins of comics analysis, Eisner (1985) wrote

> 'Comics' deal with two major communicating devices, words and images. Admittedly, this is an arbitrary separation. But since in the modern world of communication they are treated as independent disciplines, it seems valid. Actually, they are derivatives of a single origin, and in the skillful employment of words and images lies the expressive potential of the medium. (p. 13).

Essentially, the harnessing of two distinct art forms in a symbiotic relationship with each other, and utilising a number of conventions, each of which adds to the form, has created a powerful communicating instrument that is, simultaneously, a means of entertainment. As such, comics can also be a popular educational tool. Not simply as a facile way of passing on information, but essentially, and much more productively, as a way for teachers and students to convey what they think and feel, expressing themselves with this double whammy of communicative methods.

Research by Sousanis (2020), among others, has shown that giving students, even those who did not normally draw, comics-creating exercises, was greatly beneficial. "Through these simple comics-making exercises we'd been doing together, they had been teaching themselves and gained a tremendous wealth of understanding that, I believe, far exceeded what they would've attained at this point from readings" (Sousanis, 2020, p. 93).

This is not a one-off, either. The genre lends itself in various ways to pedagogical usage. The most common way is through experimentation with using comics and graphic novels as a way of creating insight into various academic areas. Lan Dong (2013), for example, used them to teach global awareness, and Daniel Ian Rubin (2013), dystopian literature.

This brief paper looks at comics and methods and approaches to the use of comics creation as an entertaining pedagogical tool.

Of Comics and Graphic Novels

Comics tell stories.

That is what they do best. They tell stories of fantastic beings who walk like gods among men, and of little boys who imagine stuffed tigers to be best friends. They speak of war, though hardly ever of peace. They have remapped the Wild West and re-imagined every classic novel that one is likely to find in the nether recesses of any library. They are fantasy, and they are reality. They can be full of humour, or full of the grimmest possible philosophy. Their settings can be a jungle at the beginning of the twentieth century, or an unnamed cityscape in the far-flung future, full of men riding pterodactyls.

Comics tell stories about everything, everyone, everywhere and every time.

I use the word "Comics" on purpose. Think of it as a generic term rather than what it might signify semantically as a word. Because there is nothing comic in Miller's *The Dark Knight Returns* (2002) and one would be hard-pressed to laugh at the funny-looking animals that populate Spiegelman's *Maus* (2003). Though they are etymologically accurate in being referred to as Comics because of their origins, the word became a hold-all that describes the unique genre that has developed in leaps and bounds over the last century and a half.

In any case, each permutation of the name means different things each time. The Graphic Novel is not, in its essential nature, a Comic in format, though it uses the same basic language. Nor is the Comic Strip. Nor, come to that, is the Comic Book. Generic differences become even stronger when the output of different countries is considered, because each alternative approach makes for a unique ensemble of the same elements. And since we are talking about an art form here, then we must also take into consideration the uniqueness of style that each artist brings into the fold, making labelling a difficult, and at times totally unnecessary, task. It is, in a sense, transmedial narratology (Stein & Thon, 2013), but there are times when it goes even beyond that.

In fact, Comics narrative has as many definitions as it has styles, as indicated above. The very language of Comics permits an enormous number of variations to the way a story is told, though there are a number of clear parameters within which writers and artists must work.

The literature provides us with a few descriptions of the genre that might actually present an idea of what those parameters might be. Here are a few.

Comics and cartoons encyclopaedist Maurice Horne based his own description of the genre on Coulton Waugh's 1947 definition: "The Comics are a form necessarily including the following elements: a narrative told by way of a sequence of pictures, a continuing cast of characters from one sequence to the next, and the inclusion of dialogue and/or text within the picture" (Horne, 1976, p. 47).

And by the father of sequential art, Will Eisner, "The format of the Comic Book presents a montage of both word and image, and the reader is thus required to exercise both visual and verbal interpretive skills. The regimens of art (e.g. Perspective, symmetry, brush stroke) and the regimens of literature (e.g. Grammar, plot, syntax) become superimposed upon each other. The reading of the Comic Book is an act of both aesthetic perception and intellectual pursuit" (Eisner, 1985, p. 8).

Finally, a complex one provided by that most diffused of books about Comics, "Juxtaposed pictorial and other images in deliberate sequence, intended to convey information and/or to produce an aesthetic response in the viewer." (McCloud, 1994, p. 9).

Taking the suggestions of all three definitions, we are left with a number of musts: 1) a narrative; 2) pictures/images/art; 3) words/text/literature; 4) sequence; and 5) continuing cast of characters.

In fact, I would add one more must to the list. I agree totally with Wright (2003), who holds "the assumption that most Comic Books succeed or fail on the merits of their storytelling" (p. xvii). So my number 6) would be storytelling.

The very nature of graphic storytelling permits diversity. In much the same way that an artist has a personal aesthetic, and a writer can adopt a particular style, the image-text combo that is a graphic narrative takes both on board. The resultant uniqueness produced by each creator, or creator-team, has paved the multi-textured and vari-coloured quilt that are Comic Strips, Comic Books, and Graphic Novels since the inception of the genre.

The language of Comics permits this. It actually has a lot of similarities with the language we use every day with its functional and content words (Saraceni, 2003), tying in with the Comics' functional and content components. They are, of course, fused together in ways that make the analysis of one separate from the other quite difficult, and the separation for analytical purposes has been a major red herring to those who have tried, because the critical terminology used is taken from the component genres, ignoring the fact that the compound creates a demand for a dedicated critical vocabulary.

This does not mean there cannot be a dissection for critical analysis. It means that the dissected elements need to be examined from what is totally a graphic narrative-based critical stance. "The graphic elements of Comics art are woven together to create the warp and woof of the medium's visual nature" (Harvey, 1996, p. 9). Harvey does go on to distinguish between four distinct graphic threads: (1) Narrative breakdown; (2) composition; (3) layout; and (4) style. Saying that, in the end, "the visual tapestry always emerges whole" (p. 9).

But in the end, there has to be fusion. As Carrier (2000) writes, referring to Winsor McCay's work, in reply to his own question as to how do sequences constitute a narrative sequence and not just a sequence of images:

> What is required is the self-evident presentation of the images as connected, as forming a causal sequence. Difficult as it may be to specify necessary and sufficient conditions for success in this synthesis, everyone is aware of what happens when such narratives "work" or fail. (p. 56)

The connection is in the main implicative. "Comics tell a story both on different levels of the pictures and different levels of time. Hence, something also happens between the pictures, and the consumer is called upon to discover the development of actions within the individual picture." (Silbermann, 1986, p. 21). Pacing and time are essential characteristics of the narrative, both as portrayed in the panels, and as implied by the gutters that link in almost gestalt fashion and provide the full continuity for the reader.

Comics, Children, Teaching and Learning

So how do comics tell stories to children, so that in turn, children can tell stories with comics? Comics were not originally intended for children.

"The idea that comics were for children was really an idea that took hold in the middle of the twentieth century. Before that,using words and pictures together to tell a story sequentially was a good way to reach just about anyone who could read" (Sanders, 2016). It was at that time, however, that the very nature of the medium, with images helping slow readers, and good readers getting the full package, seeing illustration endorse the words they read, that comics developed into what became primarily a children's market. This was to change some time in the last decade of the nineties, when comics were reclaimed by adults … or, to explain it better, those adults that grew out of the comics-loving children.

Of course, this varies across countries, and I do believe that even before comics were specifically made with children in mind, comics must have been read by youngsters. Still, in the States, when the modern comic book format was developed by Max Gaines in 1933, with his Funnies on Parade (Rhoades, 2008), there seems to be consensus that that is when children became the main readers of the genre. In the UK, the Dundee-based company DC Thomson had its stable of juvenile comics, including what was to become their flagship title, *The Dandy*, which was first published in 1937. They did not invent the British children's comic, but "it is surely no exaggeration to say that Dandy and Beano revolutionized the world of British children's comics, including full-colour covers and anthropomorphic characters" (Chapman, 2011, p.31). Preceding all of this, and in a way hinting at the road ahead, the Italian *Il Giornalino della Domenica*, that from the start ran comics-like stories, started publication in 1906 (Gallo, 2008). *Il Corriere dei Piccoli*, which began publication in 1908, followed suit (Castaldi, 2017).

In spite of a drop in sales of comics aimed at children, explainable because of the multiple types of entertainment available to youngsters, there can be no doubt that children have an affinity with the genre, both in its printed and its fledgling online form.

Philip Nel (2020) makes an excellent case as to why comics are so important within children's maturing and educational processes:

> Comics for young readers merit our attention. As works read by people who are still very much in the process of becoming, children's comics have the potential to be among the most influential books in a young person's life. These comics give children some of their earliest aesthetic experiences. They introduce small humans to art, language, graphic design, and ideas. Finally, even though comics have found their way into educational

> institutions, they are still as likely to be found in children's bedrooms as in classrooms; the advantage is that comics are still books that children can choose to read themselves." (p. 135)

With that as the base on which to build the reasoning why comics should be integrated into the curriculum, both, potentially, for mainstream teaching (with books like Elder's *Reading with Pictures* (2014), and others like it making life easier) and for special projects, like the MIRACLE project's utilisation of enhanced comics as a bridge to knowledge about climate change (coMics and IllustRations Augmented to tackle CLimate change in primary Education (MIRACLE), 2025). Comic creation was also at the core of the CLIMATOPIA project, in which a fantasy story gave rise to diverse discussions on climate awareness (Varonis et al., 2024). All that is left is the question "how" – because innovation definitely needs to be at play in cases like these.

Comics scholar Charles Hatfield (2005) pointed at the 1970s as the years in which there was a turning point in comics being accepted as instructional tools in American elementary and secondary schools. Up to today, there are programmes throughout the States that use comics in different areas of teaching and learning (Tilley & Weiner, 2017).

Studies have clearly shown how useful comics are for this purpose. As far back as 2007, Mallia proved that comics are not just enticers to motivation. That is to say, they are not just instruments for the affective domain, but can be considered to affect cognition, in much the same way that reading does. (Mallia, 2007).

Although comics have not drawn an enormous amount of attention from academia, there have been a number of academic analysts, or at least serious thinkers, who have striven to define the mechanics of the genre. Duncan and Smith (2009) list a total of twenty-eight milestones in the development of comic art studies, beginning with Frederick Coulton Waugh's The Comics from 1947, and ending with the University of Florida's launching of the first online refereed academic journal about comics, *ImageText: Interdisciplinary Comics Studies* in 2004. This does not mean that there were no other publications other than those mentioned by Duncan and Smith during the fifty-seven years in the compendium, but the fact that only twenty-eight instances of scholarship are mentioned as standing out is indicative of the dearth of in-depth analyses of the genre. The fact that many of the works mentioned by Duncan and Smith are actually histories of the development of comics and not works that actually try to understand the workings of comics continues to emphasise this.

Still, research is ongoing, even in the universities. According to Tilley & Weiner, "Recent dissertations have studied diverse issues such as comics as literacy tools for community college students (Burke 2012), the suitability of comics as texts in high school social studies classrooms (Boerman-Cornell 2012), the impact of comics on middle-grade students' reading motivation (Edwards 2008), and the pedagogy of comics production with young children (Stoermer 2009)."

Conclusions

Comics are a medium that is diverse, rich, and amenable to educational practice. The multifaceted nature of the genre is conducive to a large number of permutations when it comes to both curriculum and project inclusion. The fact that it is essentially an entertainment medium, helping create a motivational reason for its presence (in and out of the classroom), is value added and can be described as a way to get to the end result – i.e. a non-classroom route to classroom inclusion.

Comics can be used as pedagogical tools that can inform and be created by students to inform and analyse personal understanding of a theme. They can be both print and online; in the second case, also open to enhancement, helping create a game-like context that helps student interactivity and motivational thrust.

But, in essence, it is the very nature of the genre that entices, allures, and draws students to it, making them open to the understanding of what is being communicated, or, in the case of their own creation of comics, what and how they would like to communicate it.

References

Boerman-Cornell, W. (2012). *Learning to See History: A Content Analysis of the Affordances of Graphic Novels for High School Teaching* [Unpublished doctoral dissertation]. University of Illinois at Chicago.

Burke, B. P. (2012). *Using comic books and graphic novels to improve and facilitate community college students' literacy development* [Unpublished doctoral dissertation]. Indiana University of Pennsylvania.

Carrier, D. (2000). *The aesthetics of comics.* The Pennsylvania State University Press.

Castaldi, S. (2017). A brief history of comics in Italy and Spain. In F. Bramlett, R. T. Cook, & A. Meskin (Eds.), *The Routledge companion to comics* (pp. 79-87). Routledge.

Chapman, J. (2011). *British comics: A cultural history.* Reaktion Books.

Dong, L. (2013). Beyond borders: Teaching global awareness through the graphic novel. In C. K. Syma & R. G. Weiner (Eds.), *Graphic novels and comics in the classroom (*pp. 220-232). McFarland & Company, Inc.

Duncan, R., & Smith, M. J. (2009). *The power of comics: History, form & culture.* Continuum.

Edwards, B. (2008). *Motivation and middle school readers: Graphic novels, comic Books, and free voluntary reading time* [Unpublished doctoral dissertation]. University of Oklahoma.

Eisner, W. (1985). *Comics & sequential art* (21st printing, 2001). Poorhouse Press.

Elder, J. (2013). *Reading with pictures.* Andrews McMeel Publishing.

Gallo, C. (2008). Vamba e i ragazzi del "Giornalino della Domenica" a Fiume. *In Atti della Accademia Roveretana degli Agiati. A, Classe di scienze umane, lettere ed arti, 8*(1*),* 293-320.

Harvey, R. C. (1996). *The art of the comic book: An aesthetic history.* University Press of Mississippi.

Hatfield, C. (2005) *Alternative comics: An emerging literature*. University Press of Mississippi.

Horne, M. (Ed.) (1976). *The world encyclopaedia of comics.* Avon

Mallia, G. (2007). Learning from the sequence: The use of comics in instruction. *ImageText*, *3*(3) 104-114. https://imagetextjournal.com/learning-from-the-sequence-the-use-of-comics-in-instruction/

McCloud, S. (1994). *Understanding Comics: The Invisible Art.* Harper.

Miller, F. (2002). *Batman: The dark knight returns*. DC Comics.

MIRACLE (2025). *Miracle project*. https://project-miracle.eu/index.php/en/

Nel, P. (2020). Children and comics. In C. Hatfield & B. Beaty (Eds.), *Comics studies: A guidebook (*pp. 126-137). Rutgers University Press.

Rhoades, S. (2008). *A complete history of American comic books.* Peter Lang.

Rubin, D. I. (2013). "Remember, remember the fifth of November": Using graphic novels to teach dystopian literature. In C. K. Syma & R. G. Weiner (Eds.), *Graphic novels and comics in the classroom (*pp. 84-90). McFarland & Company, Inc.

Sanders, J. S. (2016). How comics became kids' stuff. In M. A. Abate & J. S. Sanders (Eds.), *Good grief! Children and comics (*pp. 9-28). Billy Ireland Cartoon Library & Museum.

Saraceni, M. (2003). *The language of comics.* Routledge.

Silbermann, A. (1986). The way toward visual culture: Comics and comic films. In A. Silbermann & D-D. Dyroff (Eds.), *Comics and visual culture: Research studies in ten countries (*pp. 11-27). K.G. Saur.

Sousanis, N. (2020). Thinking in comics: All hands-on in the classroom. In S. E. Kirtley, A Garcia, & P. E. Carlson (Eds.), *With great power comes great pedagogy: Teaching, learning, and comics* (pp. 92-116). University Press of Mississippi.

Spiegelman, A. (2003). The complete maus: A survivor's tale. Pantheon Books.

Stein, D., & Thon, J.-N., eds. (2013). *From comic strips to graphic novels: Contributions to the theory and history of graphic narrative.* De Gruyter. https://doi.org/10.1515/9783110427660

Stoermer, M. (2009) *Teaching between the frames: Making comics with seven and eight year old children, a search for craft and pedagogy* [Unpublished doctoral dissertation]. Indiana University.

Tilley, C. L., & Weiner, R. G. (2017). Teaching and learning with comics. In F. Bramlett, R. T. Cook, & A. Meskin (Eds.), *The Routledge companion to comics* (pp. 358-366). Routledge.

University of Florida Department of English. (2004). *ImageText.* https://imagetextjournal.com/

Varonis, O., Varonis, E. M., Pyrini, A., & Ioannides, C. Climatopia: Self-directed learning to improve quality of life. In E. M. Varonis and A. Pyrini, (Eds.), *Communication technologies in education: Proceedings of ICICTE 2023* (pp. 77-94). http://www.icicte.org/assets/08-icicte2023%e2%80%93proceedings%e2%80%93varonis.pdf

Waugh, C. (1947). *The comics.* Macmillan.

Wright, B. W. (2001). *Comic book nation: The transformation of youth culture in America.* John Hopkins University Press.

Author Details

Ġorġ Mallia
University of Malta
Malta
Gorg.mallia@um.edu.mt

ALIGNING THE AFFORDANCES OF IMMERSIVE VIRTUAL REALITY WITH EDUCATIONAL OBJECTIVES THROUGH THE IVRPM

Daniela Rocha Bicalho and João Piedade
UIDEF, Instituto de Educação, Universidade de Lisboa
PORTUGAL

Abstract

This study presents the Immersive Virtual Reality Pedagogical Model (iVRPM), a framework that aims to support the planning of educational experiences in immersive virtual environments. The model proposes the alignment between the affordances of Immersive Virtual Reality (iVR)—such as immersion, interactivity, and embodiment—and specific educational objectives. It also proposes the alignment between task typology and the technological features of immersive environments, linking interactivity levels to Bloom's revised taxonomy. These levels guide educators in designing activities that are consistent with educational objectives, task complexity, motor skill development, and students' familiarity with immersive technologies.

Introduction

The integration of emerging technologies into education has created new opportunities to enhance teaching and learning processes. Among these innovations, Immersive Virtual Reality (iVR) has stood out for its potential to provide interactive experiences through digital environments that simulate real-world contexts or even transport users to hypothetical, imagined, or historically reconstructed scenarios. As Wu et al. (2020) state, this technology has been widely used to create realistic, situated learning contexts that students would not usually have access to.

iVR technology enables users to interact naturally within three-dimensional environments using devices such as head-mounted displays (HMDs), motion sensors, handheld controllers, and advanced graphics processors. These tools allow for real-time motion tracking and continuous stereoscopic image generation, resulting in highly interactive and embodied environments (Won et al., 2023). With advancements such as hand tracking and more intuitive controllers, both fidelity and freedom of movement have significantly improved, enhancing the user experience. In this context, it becomes essential to understand how content is encoded, experienced, and retained in these environments (Johnson-Glenberg, 2018).

Despite its potential, the pedagogical integration of iVR still presents a challenge, particularly due to the lack of conceptual models that guide its application in alignment with teaching and learning demands. The so-called "immersion principle" asserts that immersive virtual environments lead to better learning outcomes when they incorporate effective instructional design principles (Makransky, 2021). In other words, the author argues that immersion alone does not guarantee improved learning performance, but the use of well-designed instructional methods within these environments can enhance cognitive processes such as selection, organization, and integration of information.

This study presents the Immersive Virtual Reality Pedagogical Model (iVRPM), a framework designed to support the planning of immersive educational experiences by aligning the interactive features of the technology with Bloom's revised taxonomy (Anderson et al., 2001). Developed through a Design-Based Research (DBR) approach, the model connects levels of interactivity to proposed tasks and the corresponding dimensions of knowledge, enabling the progressive development of psychomotor skills required for the use of the technology, as well as advancement through the cognitive levels of learning.

Affordances of Immersive Virtual Reality

In educational practice, iVR has emerged as a promising technology capable of creating interactive simulations that replicate real-world processes and situations (Merchant et al., 2014). These simulations provide a personalized experience, allowing students to engage more deeply with the content (Wu et al., 2020).

As Makransky (2021) highlights, traditional multimodal learning—based on words and images—has evolved into more dynamic forms through the use of emerging technologies, such as animations and, more recently, immersive virtual environments. Virtual reality enables students to interact with pedagogical agents in three-dimensional contexts that would not be possible in the physical world and are significantly more realistic than videos or computer-based simulations.

As Zilles Borba (2023) noted, immersive devices also profoundly transform the ways in which digital content is produced and consumed. According to the author, iVR constitutes an advanced human-computer interaction interface, structured by transparency mechanisms that eliminate the perception of technological mediation, creating the illusion of presence within the digital environment. This sense of immersion and presence, experienced in 360° scenarios, allows individuals to see, hear, and interact with digital content through natural sensations, characterizing a communicational process distinct from that provided by conventional screens (Zilles Borba, 2023).

The combination of real-time motion tracking and stereoscopic image generation has enabled the creation of more immersive, interactive, and embodied digital environments. These spaces provide students with opportunities to interact with the content, allowing them to engage actively—not merely as observers, but as participants in the experience.

Among the affordances of immersive virtual reality (iVR), key features include the ability to promote immersion, the spatial manipulation of three-dimensional objects, the experimentation with situations that go beyond physical reality, and the active involvement of the student's body in the experience (Johnson-Glenberg, 2018; Makransky & Petersen, 2021; Won et al., 2023).

As observed by Won et al. (2023), the technological capabilities of iVR influence the creation of authentic sensory stimuli capable of inducing the sensation that virtual objects and environments are real, thereby enhancing sensory and representational fidelity. The authors also emphasize that these technologies allow users to act naturally and intuitively in virtual environments, through coherent and fluid actions.

Two central affordances of iVR, Immersion and Interaction, operationalized through aspects such as representational fidelity and immediacy of control — refer to how well simulations replicate real-world environments not only in appearance but also in the emotional, cognitive, and behavioral responses they evoke in users (Harris et al., 2021; Petersen et al., 2022). They also include the range of learner interaction modes (Dalgarno & Lee, 2010), supporting active involvement.

Supported by other authors, Makransky & Petersen (2021) indicated that iVR is characterized as a complex media system capable of providing sensory immersion and sophisticated content representation. Immersion is directly tied to the degree of vividness offered by a system, which is an objective measure of its ability to exclude the outside world. This vividness depends on factors such as the number of senses activated and the quality of the hardware used (Cummings & Bailenson, 2016). Dalgarno and Lee (2010) suggested that immersion should not be treated as a standalone property but as dependent on other aspects present in an immersive learning experience. In this sense, immersion can be described as "a psychological state characterized by the perception of being enveloped, included, and interacting with an environment that offers a continuous flow of stimuli and experiences" (Agrawal et al., 2020, p. 277).

Zilles Borba (2023), building on the work of previous authors, proposed a dynamic structured around three pillars — realism, interactivity, and involvement — which together form the concept of believability, associated with the perception of reality in iVR experiences. In this framework, involvement refers to the quality of the narrative and the ability of the storyline to capture attention and evoke emotions (Zilles Borba, 2023, p.78). According to the author, these pillars are closely related to the concept of plausibility, originally defined by Slater et al. (2009) as the user's acceptance of the virtual environment as credible, which enables realistic behaviors even in simulated contexts, depending on narrative coherence and meaningful responses to events. Additionally, involvement may also be influenced by individual factors such as attention, motivation, prior experiences, and immersive potential, a term proposed by Agrawal et al. (2020) to describe the user's subjective predisposition to engage in virtual environments.

According to the author, these pillars are closely related to the concept of plausibility, originally defined by Slater (2009) as the user's acceptance of the virtual environment as credible, which enables realistic behaviors even in simulated contexts, depending on narrative coherence and meaningful responses to events.

Technological factors can influence embodiment, which may have a positive effect on learning outcomes (Klingenberg et al., 2024). Embodiment, as highlighted by Johnson-Glenberg (2018), refers to the idea that learning is enhanced when students actively engage with the content using their bodies through gestures, physical movements, or object manipulation in virtual environments. In virtual reality, this sensation typically occurs when the participant looks down from a first-person perspective and sees a virtual body replacing their own — especially when that virtual body is programmed to move synchronously with the participant's real movements (Klingenberg et al., 2024). In addition, it is important to consider that the development of psychomotor skills, such as the dexterity and coordination required to navigate and interact within the virtual environment, also plays a crucial role in the learning process. The learner's ability to operate the technology can influence both how they interact with the experience and their capacity to engage immersively in the activities (De Freitas et al., 2010).

Methodology

This study adopts the methodological approach of Design-Based Research (DBR), which is structured in three phases. DBR is characterized by iterative cycles of design, implementation, analysis, and redesign, aiming both at the development of practical solutions and the advancement of theoretical understanding in real educational contexts (Zheng, 2015). As highlighted by Tinoca et al. (2022), this approach is recognized for its capacity to promote innovation and transformative interventions, particularly through authentic and integrated teaching and learning practices. A more detailed account of the iterative process followed in this research is available elsewhere (Bicalho et al., 2023). The phases of this study are described below:

(1) A systematic literature review (Bicalho et al., 2024), which identified gaps and opportunities in the use of immersive technologies in

education, offering insights for the construction of pedagogical frameworks.

(2) The development and validation of the iVRPM (Bicalho et al., 2025), which aligns key affordances of immersive environments with educational objectives.

(3) The planned implementation and validation of the framework through an educational prototype with higher education students.

This article focuses on Phase 2, specifically on demonstrating how the iVRPM framework enables the alignment between iVR affordances and educational activities. It presents a mapping strategy that links different levels of interactivity to specific learning objectives and task types, grounded in Bloom's revised taxonomy (Anderson et al., 2001) and supported by recent literature on immersive learning.

Developed within the scope of a DBR methodology, the iVRPM framework integrates three main theoretical references: (i) the Cognitive-Affective Model of Immersive Learning (CAMIL) (Makransky & Petersen, 2021); (ii) the XR ABC Framework (Lion-Bailey et al., 2019); (iii) Bloom's Revised Taxonomy (Anderson et al., 2001).

The following section details the iVRPM, emphasizing its underlying logic and how it organizes the articulation between immersive virtual reality affordances and educational objectives.

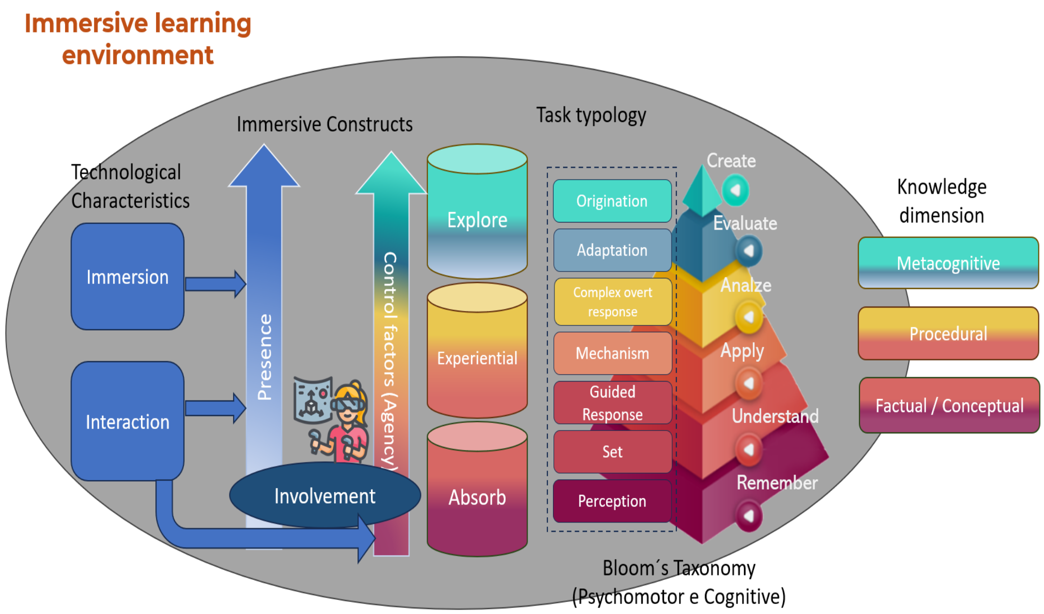
iVRPM - Immersive Virtual Reality Pedagogical Model

The iVRPM was developed to guide pedagogical planning in iVR environments, addressing the theoretical gap identified in the literature regarding the lack of structured models that integrate their technological aspects of iVR with educational goals (Bicalho et al., 2024; Radianti et al., 2020).

The proposed framework establishes a relationship between technological features of virtual environments, such as Immersion and Interaction, and educational objectives, using the revised Bloom's Taxonomy as its foundation (Anderson et al., 2001). Figure 1 illustrates its core components.

Figure 1

Pedagogical Framework - iVRPM: Immersive Virtual Reality Pedagogical Model (Bicalho et al., 2025)



The iVRPM establishes a positive relationship between the technological characteristics of immersion and interaction, highlighting how the combination of sensory fidelity (realism) and the interactive capabilities of the environment contribute to the construction of the sense of Presence. In addition, the framework suggests that the degree of alignment between the user's movements and the virtual environment's response enhance the sense of Control (Agency). These dimensions are strongly associated with student involvement, understood as a psychological state in which focus and energy are directed toward specific stimuli or activities (Shadiev et al., 2021). The framework proposes that involvement increases as Presence and Agency are intensified, and in turn, reinforces these sensations, generating a cycle that amplifies the immersive experience. This cycle is directly aligned with the concept of believability, as proposed by Zilles Borba (2023), which synthesizes the perception of reality as the result of the interplay between the sensory, interactive, and subjective aspects of the experience.

Based on this foundation, the iVRPM suggests that educational activities in iVR can be structured into different levels, depending on the affordances offered by the environment, with an emphasis on interaction. This approach takes into account recent technological advancements, such as increased visual fidelity, hand tracking, more intuitive controllers, and greater freedom of movement, that make immersive environments more responsive and sensitive to user actions. This stratification is justified by the limitations of the traditional binary division between low and high immersion, commonly applied to immersive technologies, which does not fully account for the variety of possible experiences within highly immersive environments. Although "high immersion is typically associated with the exclusion of the physical world and multisensory stimulation" (Cummings & Bailenson, 2016), different combinations of Immersion and Interaction can result in qualitatively distinct educational experiences.

To this end, the iVRPM proposes a stratification of immersive experiences into three levels based on the user's ability to control, interact with, and influence the virtual environment. These levels are:

- **Absorb**: Passive participation, limited interaction, minimal feedback, and no significant bodily involvement.
- **Experiential**: Active participation, object manipulation with congruent gestures, immediate feedback, and partial bodily involvement.
- **Explore**: Autonomous participation, creation and modification of the environment, full bodily engagement, and immediate, responsive feedback.

This proposal is further enriched by the sensory and action-related elements described by Won et al. (2023), which include aspects such as visual and auditory fidelity and responsiveness to user actions.

The dimensions of *Participation*, *Feedback/Assessment*, and *Engagement Strategy*, as proposed by Mystakidis and Lympouridis (2024), are used as a supportive reference to enrich the description of the levels in the iVRPM. These dimensions help articulate how users engage with the simulation (Participation), how the system responds to their actions (Feedback), and how the experience maintains user motivation (Engagement). *Participation* (Dimension (Dim.) 12) ranges from passive observation to active content creation. *Feedback / Assessment* (Dim. 13) includes various levels of system responsiveness, from no feedback to automated evaluations and post hoc expert reviews. *Engagement Strategy* (Dim. 14) encompasses methods such as task-only execution, gamified elements, storytelling, and world-building. To operationalize this alignment, each iVRPM level can be described through these dimensions:

- **Absorb –** Observation, None or Basic Mediation, Task only
- **Experiential –** Guided/Unguided Practice, Automated Feedback, Gamification/Narration
- **Explore –** Content Authoring, Immediate Feedback/Debriefing, World Building

The revised Bloom's taxonomy is structured as a bidimensional matrix, combining cognitive processes (such as remembering, applying, or creating) with types of knowledge (factual, conceptual, procedural, and metacognitive). In the context of the iVRPM, this structure is preserved by associating each interaction level with both a cognitive level and a knowledge dimension. Rather than establishing rigid classifications, the framework identifies predominant combinations that tend to emerge from the affordances activated at each level. This approach allows educators to align immersive tasks with pedagogical intentions more precisely, ensuring that interaction design supports the type of learning expected in each activity.

Building on this association, the level of interactivity in immersive activities can help educators align specific learning objectives with Bloom's revised taxonomy and the corresponding types of knowledge involved. At the **Absorb** level, tasks are designed to support the *Remember* and *Understand* stages of the cognitive domain, involving simple navigation and content interpretation, and are associated with factual and conceptual knowledge. The **Experiential** level corresponds to *Apply* and *Analyze*, offering moderate interactivity through object manipulation and simulations that promote conceptual and procedural knowledge. Finally, the **Explore** level aligns with the *Evaluate* and *Create* stages of Bloom's taxonomy, supporting high interactivity, learner autonomy, and active involvement through the construction and modification of content within the virtual environment. This progression also highlights the psychomotor demands of immersive environments, requiring learners to engage cognitively while navigating and interacting through embodied action.

In addition to aligning interactivity levels with cognitive objectives, the iVRPM also considers the activation of psychomotor processes, based on Simpson's taxonomy (Simpson, 1972). Each proposed level entails increasing demands in terms of motor skills and the degree of embodiment required from the learner. In the initial stages, perceptual involvement and guided gestures are sufficient, whereas more advanced levels require greater coordination, spatial awareness, and full-body involvement. Moreover, the effective execution of immersive tasks depends on the learner's familiarity with the devices, which directly influence agency and performance.

Table 1 summarizes how the psychomotor levels proposed by Simpson (1972) can be interpreted within the context of the iVRPM, illustrating the progression of motor and embodied demands across the three interaction levels.

Table 1

Interpretation of Simpson's Psychomotor Levels in the Context of iVRPM

	Simpson's Psychomotor Levels *(1972)*	**Description**
Absorb	1. Perception	Recognition of simple sensory stimuli in the virtual environment, with minimal bodily interaction.
	2. Set	Initial readiness for action with basic tasks and visual guidance.
	3. Guided Response	Execution of basic actions with support, including trial and error.
Experiential	4. Mechanism	Confident execution of actions, with greater motor integration and precision.
	5. Complex Response	Advanced motor skills with increased control and coordination.
Explore	6. Adaptation	Autonomous adjustment of movements in response to the complexity of the environment.
	7. Origination	Creation of novel motor solutions and fully integrated actions with complete freedom.

Thus, the iVRPM proposes a structure that relates levels of interactivity to educational goals, taking into account technological affordances, task complexity, and the skills required for the use of iVR. The framework aims to support the planning of educational experiences in immersive environments by providing criteria to organize activities according to learning objectives.

Conclusion

Designed as a theoretical contribution, the iVRPM offers a conceptual structure to support the alignment between the affordances of immersive virtual reality and educational objectives. By organizing immersive learning experiences into three interactivity levels (Absorb, Experiential, and Explore), the framework helps to map tasks to different stages of Bloom's Revised Taxonomy (Anderson et al., 2001), ranging from recognizing and understanding content to applying, analyzing, evaluating, and creating within virtual environments.

This alignment is reinforced by the concept of believability (Zilles Borba, 2023), which frames the virtual experience as the result of the integration of realism, interactivity, and involvement. These dimensions contribute to the perception of reality by combining sensory fidelity with the immersive environment's ability to focus and sustain learner attention, thereby influencing presence and agency.

Within the iVRPM, these dimensions are reflected in the definition of interactivity levels, which are associated with both cognitive processes and dominant knowledge types (factual, conceptual, procedural, and metacognitive). By helping educators to balance technological affordances with pedagogical intent, the framework supports the design of meaningful and coherent immersive learning experiences, taking into account task complexity and learner capabilities. In addition, the framework considers the psychomotor aspects involved in immersive learning, incorporating the development of embodied skills. By proposing this articulation, it offers a theoretical contribution aimed at guiding pedagogical design in immersive environments.

Future research should investigate its practical application in real educational contexts, exploring its effects on student performance and the effectiveness of the planned immersive experiences.

Notes

This work was supported by national funds through the Portuguese Foundation for Science and Technology, I.P. (FCT), under the PhD studentship reference 2022.11515.BD (https://doi.org/10.54499/2022.11515.BD), and within the scope of the Research Unit UIDEF – Unidade de Investigação e Desenvolvimento em Educação e Formação (UIDB/04107/2020, https://doi.org/10.54499/UIDB/04107/2020).

References

Agrawal, S., Simon, A., Bech, S., Bæntsen, K., & Forchhammer, S. (2020). Defining immersion: Literature review and implications for research on audiovisual experiences. Journal of the Audio Engineering Society, 68(6), 404–417. https://doi.org/10.17743/jaes.2020.0039

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. Longman. http://books.google.com/books?id=JPkXAQAAMAAJ&pgis=1

Bicalho, D. R., Piedade, J., & Matos, J. F. (2023). As potencialidades da realidade virtual como ferramenta de apoio em práticas educativas. *PRATICA - Revista Multimédia de Investigação em Inovação Pedagógica e Práticas de e-Learning, 6*(3), 36–42. https://doi.org/10.34630/pel.v6i3.5353

Bicalho, D. R., Piedade, J. M. N., & Matos, J. F. de L. M. (2024). Systematic review of educational practices with immersive virtual reality in higher education. *Medi@ções, 12*(1), 95–111. https://doi.org/10.60546/mo.v12i1.410

Bicalho, D. R., Piedade, J., & Matos, J. F. (2025). iVRPM: Conceptual proposal of an immersive virtual reality pedagogical model. *Applied Sciences, 15*(4), 2162. https://doi.org/10.3390/app15042162

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Media Psychology, 19(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? British Journal of Educational Technology, 41(1), 10–32. https://doi.org/10.1111/j.1467-8535.2009.01038.x

De Freitas, S., Rebolledo-Mendez, G., Liarokapis, F., Magoulas, G., & Poulovassilis, A. (2010). Learning as immersive experiences: Using the four-dimensional framework for designing and evaluating immersive learning experiences in a virtual world. British Journal of Educational Technology, 41(1), 69–85. https://doi.org/10.1111/j.1467-8535.2009.01024.x

Harris, D. J., Buckingham, G., Wilson, M. R., Brookes, J., Mushtaq, F., Mon-Williams, M. & Vine, S. J. (2021). Exploring sensorimotor performance and user experience within a virtual reality golf putting simulator. Virtual Reality, 25, 647–654. https://doi.org/10.1007/s10055-020-00480-4

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. Frontiers in Robotics and AI, 5(81). https://doi.org/10.3389/frobt.2018.00081

Klingenberg, S., Bosse, R., Mayer, R. E., & Makransky, G. (2024). Does embodiment in virtual reality boost learning transfer? Testing an immersion-interactivity framework. Educational Psychology Review, 36, Article 116. https://doi.org/10.1007/s10648-024-09956-0

Lion-Bailey, C., Lubinsky, J., & Shippee, M. (2019, October 20). A learning framework for augmented and virtual reality. *TeachThought*. https://www.teachthought.com/technology/augmented-virtual-reality/

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): a Theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 33,* 937–958. https://doi.org/10.1007/s10648-020-09586-2

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. Computers and Education, 70, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

Mystakidis, S., & Lympouridis, V. (2024). Designing simulations in the metaverse: A bBlueprint for experiential immersive learning experiences. In V. Geroimenko (Ed.), *Augmented and virtual reality in the metaverse* (pp. 65–79). Springer Nature. https://doi.org/10.1007/978-3-031-57746-8_4

Petersen, G. B., Petkakis, G., & Makransky, G. (2022). A study of how immersion and interactivity drive VR learning. Computers & Education, 179, 104429. https://doi.org/10.1016/j.compedu.2021.104429

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Shadiev, R., Wang, X., & Huang, Y.-M. (2021). Cross-cultural learning in virtual reality environment: Facilitating cross-cultural understanding, trait emotional intelligence, and sense of presence. *Educational Technology Research and Development, 69*, 2917–2936. https://doi.org/10.1007/s11423-021-10044-1

Simpson, E. J. (1972). *The classification of educational objectives in the psychomotor domain*. In M. T. Rainer (Ed.), *Contributions of behavioral science to instructional technology: The psychomotor domain* (Vol. 3, pp. 43–56). Englewood Cliffs, NJ: Gryphon Press, Prentice Hall. Retrieved from https://files.eric.ed.gov/fulltext/ED084870.pdf

Slater, M., Lotto, B., Arnold, M. M., & Sanchez-Vives, M. V. (2009). How we experience immersive virtual environments: The concept of presence and its measurement. *Anuario de Psicología*, *40*(2), 193–210. https://doi.org/10.1344/105.000001253

Tinoca, L., Piedade, J., Santos, S., Pedro, A., & Gomes, S. (2022). Design-based research in the educational field: A systematic literature review. *Education Sciences, 12*(6), 410. https://doi.org/10.3390/educsci12060410

Won, M., Ungu, D. A. K., Matovu, H., Treagust, D. F., Tsai, C.-C., Park, J., Mocerino, M., & Tasker, R. (2023). Diverse approaches to learning with immersive Virtual Reality identified from a systematic review. Computers and Education, 195, 104701. https://doi.org/10.1016/j.compedu.2022.104701

Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. British Journal of Educational Technology, 51(6), 1991–2005. https://doi.org/10.1111/bjet.13023

Zheng, L. (2015). A systematic literature review of design-based research from 2004 to 2013. Journal of Computers in Education, 2, 399–420. https://doi.org/10.1007/s40692-015-0036-z

Zilles Borba, E. (2023). Plataformização da publicidade em ambientes imersivos da realidade virtual: metaversos, marcas e acreditabilidade [Advertising platformization in immersive virtual reality environments: metaverses, brands and credibility]. Comunicação Mídia e Consumo, 20(57). https://doi.org/10.18568/cmc.v20i57.2733

Author Details

Daniela Rocha Bicalho
UIDEF, Instituto de Educação, Universidade de Lisboa
PORTUGAL
danielabicalho@edu.ulisboa.pt

João Piedade
UIDEF, Instituto de Educação, Universidade de Lisboa
PORTUGAL
jmpiedade@ie.ulisboa.pt

AI EDUCATION THROUGH STORYTELLING

George Gadanidis
Western University

Janette M. Hughes
Ontario Tech University

CANADA

Abstract

Primary and secondary school curricula in many countries are integrating AI education across subject areas, with a focus on AI applications, societal implications, and ethics. However, this is a recent phenomenon, so there is a gap in available resources to support such curricula. In this paper, we share our approach to designing story-based resources for teaching and learning about AI.

Storytelling

Storytelling is not a frill. It is not simply entertainment. Boyd (2009) notes that our appeal and necessity for storytelling developed through evolutionary adaptation. Bruner (1986) identified two distinct (yet complementary) modes of thought: the *narrative*, concerned with meaning-making through storytelling, and the *paradigmatic*, concerned with truth through logic. Story makes us human and adds humanity to teaching and learning. Boyd (2001) adds that good storytelling involves solving *artistic* puzzles of how to create situations where the audience experiences the pleasure of surprise and insight. Anthropologist Ellen Dissanayake (1992) wrote of the human biological necessity to experience, share and learn from surprising events and stories.

The artistic puzzles we endeavoured to solve involved creating graphic stories that: (1) immerse students in the application and impact of AI; and (2) help students understand how AI works. Below we discuss a sample of these graphic stories and their potential for AI education.

Application and Impact of AI

The graphic story *Meehaneeto* (Hughes and Gadanidis, 2021), shown in Figure 1, explores the potential social, economic and environmental consequences of AI's unchecked development and uncritical use, like social isolation and behaviour manipulation. A video reading of this story is available at https://eduapps.ca/community/#AI. The story engages students to consider which of today's technologies that may keep or abandon, if they had a choice. This choice is made possible by the story plot, which starts with a society that centuries ago abandoned all its technology. As chance would have it, a young girl discovers the old technology and starts bringing some of it back, including Panopteeto, which is a fictional technology analogous to the AI driven social media in our society.

Figure 1

Meehaneeto Story

One study (Butler-Ulrich and Hughes, 2025), which used *Meehaneeto* with grades 6-8 students in an AI camp setting, notes:

> The graphic novel *Meehaneeto* played a central role in shaping students' thought processes, serving as a framework for examining their lives and society. Participants reflected on how technology and AI influenced the characters' lives, drawing parallels to their own experiences with AI. *Meehaneeto* acted as a thematic anchor, engaging participants and distilling complex ethical AI concepts into an accessible narrative, enhancing critical and complex thinking through the process of meaning-making. Narrative-based learning, recognized for fostering meaning-making in students (Pantaleo, 2016), added significant value by promoting deeper understanding and critical thinking. The findings are organized to provide a clear and contextualized discussion of key outcomes related to each research question. (p. 101).

Another story that we plan to use for the same purpose is *AI Farm* (Gadanidis and Hughes, 2021), shown in Figure 2, which is a retelling of George Orwell's (1944) *Animal Farm*. This story presents a dystopia that results once agbots (agricultural robots) gain intelligence and dominate humans.

Figure 2

AI Farm Story

How AI Works

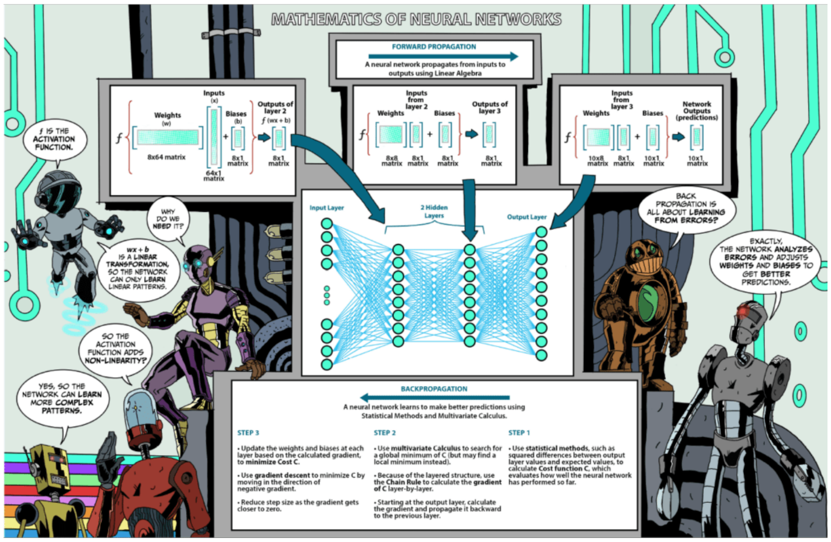
Jurisdictions focusing on educating students on how AI works and how AI is developed are realizing that a lot of AI development – a lot of the "intelligence" of AI – relies heavily on mathematical algorithms (United Nations Educational, Scientific and Cultural Organization (UNESCO), 2022). Through an AI education outreach project through Western University (see https://ai-ed.ca) we developed a set of resources to help students conceptualize neural networks and even engage with some of the underlying mathematics.

The infographic shown in Figure 3 (available at https://ai-ed.ca/nn-math) shows how a neural network uses matrices to store and linear algebra to manipulate data (forward propagation) and statistical methods and multivariate calculus to refine its predictions (back propagation).

We have also developed graphic stories to help students as young as grade 4 engage with matrices in familiar contexts.

Figure 3

Neural Network Mathematics

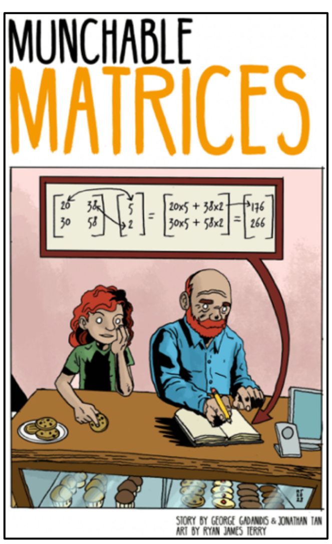


Data Management with Matrices

Around grade 4, students learn about collecting and storing data in frequency tables. Below are some tasks and dialogues from the children's story *Munchable Matrices* (Gadanidis & Tan, 2024) (Figure 4), which is freely available at https://ai-ed.ca/matrices-1. Data from the story helps prepare students to make sense of matrix operations in the forward propagation processes of neural networks (Gadanidis et al., 2024).

Figure 4

Munchable Matrices Story Cover



The context for the story is a small pastry shop, open only on Fridays and Saturdays, which is introducing cookies and muffins to its product line. It uses frequency tables to keep track of sales, as displayed in Tables 1 and 2.

Table 1

Week 1 Cookie and Muffin Sales

WEEK 1 Sales	6 Cookie Bundle ($5)	1 Muffin ($2)
Friday	8	15
Saturday	12	22

Table 2

Week 2 Cookie and Muffin Sales

WEEK 2 Sales	6 Cookie Bundle ($5)	1 Muffin ($2)
Friday	12	23
Saturday	18	36

"Nice frequency tables, Dad," says his daughter as she opens a package of cookies.

"Why don't you use matrices?" she adds.

"Huh?" he replies.

"Like this, Dad," she smiles. "Then we can add them together."

$$\begin{bmatrix} 8 & 15 \\ 12 & 22 \end{bmatrix} + \begin{bmatrix} 12 & 23 \\ 18 & 36 \end{bmatrix} = \begin{bmatrix} 20 & 38 \\ 30 & 58 \end{bmatrix}$$

"You can even multiply the result with the price matrix!"

$$\begin{bmatrix} 20 & 38 \\ 30 & 58 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 20x5 + 38x2 \\ 30x5 + 58x2 \end{bmatrix} = \begin{bmatrix} 176 \\ 266 \end{bmatrix}$$

The way father and daughter work through the matrices is depicted in Figure 5 and Table 3.

Figure 5

Munchable Matrices Story Images and Excerpt

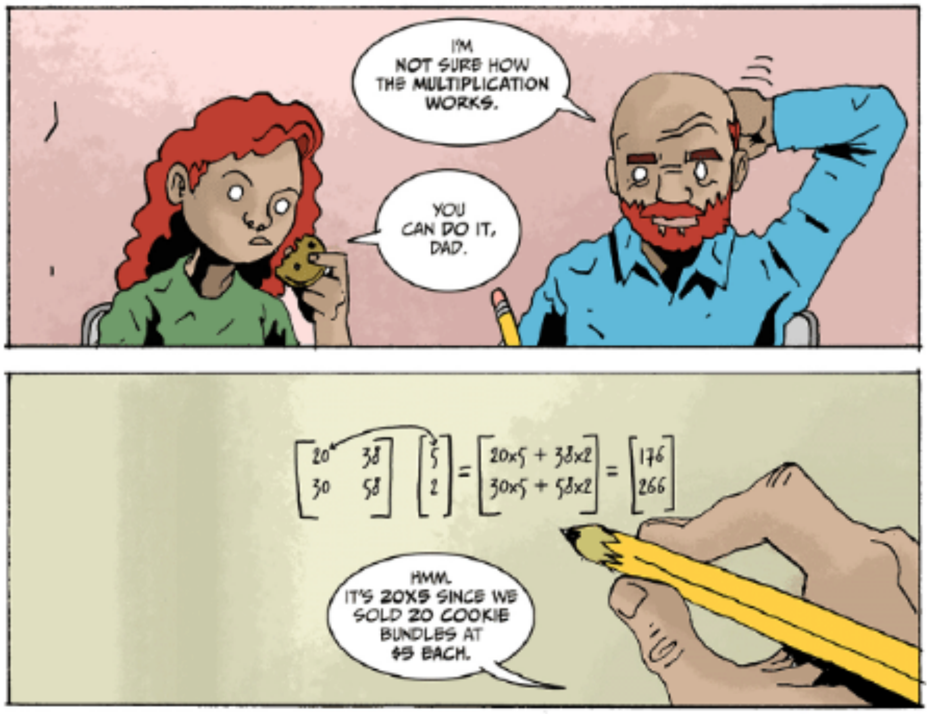


Table 3

Week 3 Cookie and Muffin Sales

WEEK 3 Sales	6 Cookie Bundle ($5)	1 Muffin ($2)
Friday	25	45
Saturday	42	65

"Perfect, Dad. Here, add the third week's sales as well. I started it for you."

$$\begin{bmatrix}20 & 38\\30 & 58\end{bmatrix} + \begin{bmatrix}25 & 45\\ & \end{bmatrix} = \begin{bmatrix} & \\ & \end{bmatrix}$$

$$\begin{bmatrix} & \\ & \end{bmatrix}\begin{bmatrix}5\\2\end{bmatrix} = \begin{bmatrix} \\ \end{bmatrix} = \begin{bmatrix} \\ \end{bmatrix}$$

Such a dialogue places students in the position of solving puzzles of making sense of how matrices are multiplied. Danesi (2020) notes that puzzles "are as intrinsic to human nature as are humor, language, art, music, and all of the other creative faculties that distinguish humanity from other species" (p. 197).

Concluding Remarks

AI is becoming pervasive in our society. At the same time, how AI works and what may be its impact are not well understood. Graphic stories may help situate and contextualize AI in ways that engage attention, anchor abstract concepts to concrete experiences, and help students develop a robust and critical understanding of AI.

References

Boyd, B. (2009). *On the Origin of Stories: Evolution, Cognition, and Fiction.* Cambridge, MA: Belknap Press of Harvard University Press.

Boyd, B. (2001). The origin of stories: Horton hears a Who. *Philosophy and Literature 25*(2), 197-214.

Bruner, J.S. (1986). *Actual minds, possible worlds.* Cambridge, MA: Harvard Press.

Buttler-Ulrick, T. and Hughes, J. (2025). Going over the wall: Supporting artificial intelligence literacy using design fiction. *Journal of Digital Life and Learning 4*(1), 86-105.

Danesi, M. (2020). *An anthropology of puzzles: The role of puzzles in the origins and evolution of mind and culture*. London: Routledge, Taylor & Francis Group.

Dissanayake, E. (1992). *Homo Aestheticus: Where art comes from and why.* New York, NY: Free Press.

Gadanidis, G., & Hughes, J. N. (2021). *AI farm.* Whitby, ON: LearnX Publications.

Gadanidis, G., Li, L. & Tan, J. (2024). Mathematics & artificial intelligence: Intersections & educational implications. *Journal of Digital Life & Learning 4*(1), 1-24.

Gadanidis, G. & Tan, J. (2024). *Munchable matrices.* Western University.

Hughes, J.N., & Gadanidis, G. (2021). *Meehaneeto*. Whitby, ON: LearnX Publications.

Orwell, G. (1944). *Animal farm.* Secker and Warburg Publishers.

Pantaleo, S. (2016). Critical thinking and young children's exploration of picture book artwork. *Language and Education, 32*(2), 152-168.

UNESCO. (2022). K-12 AI Curricula: A mapping of government-endorsed AI curricula. Paris, France: UNESCO.

Author Details

George Gadanidis
Western University
CANADA
gadanidis@gmail.com

Janette M. Hughes
Ontario Tech University
CANADA
janette.hughes@uoit.ca

FROM THE BOARD TO THE MIND: THE ROLE OF MODERN BOARD GAMES IN FOSTERING COMPUTATIONAL THINKING IN PRIMARY EDUCATION

João Piedade and Fábio Machuqueiro
UIDEF, Instituto de Educação, Universidade de Lisboa
PORTUGAL

Abstract

This study analyzed the role of Modern Board Games (MBG) as pedagogical tools for developing Computational Thinking (CT) in primary school students. A mixed-methods research design was employed with a sample of 40 students. Quantitative results, obtained through pre-test, midterm test, and post-test, revealed significant differences in the scores of the experimental group compared to the control group. Qualitative analysis, based on observations and field notes, revealed that the game mechanics enabled students to construct computational strategies promoting greater engagement. The study concludes that the use of Modern Board Games (MBG) can represent an innovative and effective approach for fostering CT in primary education.

Introduction

Computational Thinking (CT) first gained significant attention within the academic community in 2006. Jeannette Wing, a Computer Science professor at Columbia University in New York, highlighted CT as a fundamental skill for everyone, essential in today's world. She argued that CT should be considered a core competency, just like learning to read, write, or do arithmetic (Wing, 2006). Wing's call to action prompted an immediate response from educators, researchers, and institutions, who, over the years, have engaged in ongoing debate on the most effective ways to foster CT skills among students.

Today, a growing number of schools around the world are showing interest in integrating CT into their curricula. In most cases, this integration is closely tied to programming and computer science (Bocconi et al., 2022). Indeed, numerous initiatives aimed at promoting CT have emerged across all levels of compulsory education, including early childhood education. These efforts reflect not only the increasing recognition of CT's importance, but also the consensus that it should be cultivated from a young age through to higher education (Bers, 2023; Relkin et al., 2023; Yang et al., 2023).

Despite having emerged in 2006, CT remains an evolving and dynamic educational topic. The number of published studies on CT began to rise sharply after 2017, particularly within the field of educational research (Chen et al., 2023). Thematic phases in the field of CT could be grouped into the following categories: (1) the early work on computational education (Papert, 1993); (2) the emergence of the first CT definitions (Wing, 2006, 2008; Lee et al., 2011); (3) the identification of CT characteristics (Barr & Stephenson, 2011; Brennan & Resnick, 2012; Grover & Pea, 2017; Shute et al., 2017); and (4) the rapid growth of empirical and review studies on CT (Bers, 2020; Moreno-León et al., 2015).

Within the diverse landscape, CT has increasingly been introduced into primary education systems around the world through three main pedagogical approaches: programming, educational robotics, and Unplugged Activities (UA). Unplugged Activities were initially designed to introduce students to computer science concepts without the use of digital devices (Bell & Vahrenhold, 2018). However, they are now widely used in various contexts to support CT development (Munasinghe et al., 2023). These activities may include puzzles, magic tricks, or games, and are intended to cultivate the type of reasoning typically used by computer scientists through hands-on manipulation of physical materials such as paper, cards, or tokens. Within this scope, board games (BG) also play a prominent role (Menon et al., 2019).

Currently, board games are being increasingly explored as a resource to support diverse forms of learning, including CT (Bayeck, 2024; Machuqueiro & Piedade, 2022; Sousa, 2023). Still, despite their growing acceptance, the use of board games as a form of unplugged activity remains limited when compared to other CT-related approaches (Machuqueiro & Piedade, 2022). Considering that, the present study aims to address the following two research questions:

- (RQ1) What is the impact of using Modern Board Games on the development of Computational Thinking in primary school students?
- (RQ2) How do the mechanics of Modern Board Games influence the development of Computational Thinking?

This research proposes the use of MBG in classroom settings to foster CT in first-cycle primary education (1st to 4th grade). The study involved structured gameplay sessions and guided exploration of a carefully selected set of games, with the goal of evaluating the effectiveness of this approach through pre- and post-tests to assess students' Computational Thinking skills.

Computational Thinking Concepts

Computational Thinking (CT) has emerged as a key 21st-century competence, increasingly recognized as fundamental to problem-solving and the understanding of both natural and artificial systems through principles of computer science (Wing, 2006; Brennan & Resnick, 2012).

Although the term "Computational Thinking" gained prominence following Wing's (2006) work, its conceptual roots can be traced back to the seminal contributions of Seymour Papert. In Mindstorms: Children, Computers, and Powerful Ideas (1980), Papert introduced a constructivist approach to learning with computers, emphasising the importance of learners actively constructing knowledge through the design and debugging of programs using the LOGO programming language. Papert (1980) saw programming not merely as a technical skill but as a way of thinking that could empower children to approach problems systematically, creatively, and reflectively.

Wing (2006) later defined CT as a human intellectual skill involving abstraction, decomposition, automation, and iterative problem-solving, arguing that CT should be taught alongside foundational skills such as reading and mathematics. Her work sparked widespread academic interest and was pivotal in framing CT as a broadly applicable cognitive tool.

Expanding upon this foundation, Brennan and Resnick (2012) proposed a three-dimensional model that remains widely cited in the literature:

- **Computational Concepts:** such as sequences, loops, parallelism, events, conditions, operators, and data;
- **Computational Practices:** including iterative development, testing and debugging, reuse and remixing, abstraction, and modelling;
- **Computational Perspectives:** encompassing students' self-expression, empowerment, and engagement with technological artefacts and their social context.

Shute et al. (2017) contributed a complementary model focusing on six core dimensions of CT applicable across subject areas: Abstraction; Decomposition; Algorithm design; Debugging; Iteration; and Generalization. This model has proven particularly valuable in the design of assessment tasks and instructional strategies aimed at evaluating and fostering CT across educational levels.

Kafai et al. (2020) later proposed a more holistic perspective, aligning CT with Simon Sinek's "Golden Circle" theory (Sinek, 2009). Their framework integrates three layers:

- **Cognitive:** aimed at developing students' understanding of concepts, practices, and perspectives necessary for future careers;
- **Contextual:** emphasising authorship, identity, and the creation of meaningful computational artefacts;
- **Critical:** rooted in critical thinking and oriented towards social justice and transformative practice.

From a curricular standpoint, Bocconi et al. (2022) offered a consolidated overview of key concepts directly related to CT development, including: abstraction, data analysis, decomposition, pattern recognition, system thinking, algorithmic thinking, simulation, modeling, and Boolean logic.

Collectively, these dimensions and practices underscore the multifaceted nature of CT, which spans technical, cognitive, and socio-cultural domains. As Grover and Pea (2017) assert, CT is not limited to programming; rather, programming serves as one expression of the deeper thinking processes involved in CT.

In this study, these theoretical perspectives support the understanding that CT can be fostered through a range of pedagogical strategies — including the use of modern board games — which provide authentic contexts for engaging in problem-solving, abstraction, and simulation activities.

Modern Board Games and Computational Thinking Development

Modern Board Games (MBGs) have emerged as powerful analog tools for supporting the development of Computational Thinking (CT), particularly in primary education settings. While traditionally overshadowed by digital approaches such as coding and robotics, MBGs provide unplugged yet cognitively rich environments in which learners engage with core CT processes through gameplay (Bayeck, 2024).

These games, governed by structured rules and complex systems, serve as fertile ground for the expression of CT dimensions, such as abstraction, algorithm design, decomposition, pattern recognition, conditional logic, and simulation (Berland & Lee, 2011; Tsarava et al., 2019; Machuqueiro & Piedade, 2023). Empirical studies reveal that the type of mechanics embedded in a board game plays a decisive role in determining which dimensions of CT are activated during gameplay.

In the CTLM-TM framework (Computational Thinking Learning Model for Tabletop Mechanics), developed by Machuqueiro and Piedade (2023), specific types of board game mechanics were mapped to CT dimensions. The following are some key associations drawn from this model and corroborated by empirical data:

- **Resource Management & Planning Mechanics** → Associated with abstraction and algorithmic thinking, as players must anticipate outcomes and optimize decisions.
- **Conditional Play (if-then rules, action resolution)** → Encourages logical reasoning and conditional logic akin to control structures in programming.
- **Simulation and scenario exploration** → Activates debugging, simulation, and iteration, as players test strategies and refine them over multiple rounds.

- **Pattern Recognition and Set Collection** → Directly linked to the CT dimension of pattern recognition and generalisation (Tseng et al., 2019).
- **Tile Placement and Spatial Strategy** (e.g., in games like Rossio) → Supports decomposition, abstraction, and systems thinking, especially as players must visualise outcomes spatially and temporally.

Materials and Methods

This study adopted an embedded, concurrent mixed-methods design (Creswell, 2009), incorporating an experimental group (EG) and a control group (CG) to evaluate the development of Computational Thinking (CT) in students from the first cycle of basic education (grades 1st to 4th). The sample consisted of 40 third-grade students, aged between 8 and 10 years, divided into two groups (experimental and control). Sixty-minute sessions were integrated into the curriculum of an Information and Communication Technologies (ICT) project and were supervised by both the researcher and the classroom teacher.

The experimental design followed the steps outlined in Table 1, for the experimental group. The control group developed a set of unplugged pedagogical activities selected by the classroom teacher.

Table 1

Methodological Design of the Study for the Experimental Group

MBGs for CT	Pre-Test	MBG sessions	Mid-Tes	MBG sessions	Post-Test	Analysis of Results
• Systematic Analysis of Modern Board Games • Final Selection of the games for the Study • *King of Dice: The Boardgame* • *Rossio* • *PreHistorias* • *Festival*	• Beginners Computational Thinking Test (BCTt) • 1 session • 60 minutes	• King of Dice: The Boardgame and Rossio • 8 sessions • 8 weeks • Direct observation of gameplay Collection of field notes	• Beginners Computational Thinking Test (BCTt) • 1 session • 60 minutes	• PreHistorias and Festival • 8 sessions • 8 weeks • Direct observation of gameplay Collection of field notes	• Beginners Computational Thinking Test (BCTt) • 1 session • 60 minutes	• Analysis of quantitative data derived from the administration of BCTt • Content Analysis, with coding of field notes based on categories

The Beginners Computational Thinking Test (BCTt) (Zapata-Cáceres, et al., 2021) was used to measure CT skills and it was administered as a pre-test, midterm test, and post-test. The BCTt consists of 25 multiple-choice questions with progressively increasing complexity, related to Computational Thinking (CT) concepts as proposed by Brennan and Resnick (2012) and Grover and Pea (2017), such as sequences, loops (including simple and nested loops), and conditionals (including simple, compound, and while conditionals). The test also partially addresses several CT practices (incremental and iterative, testing and debugging, reusing and remixing, abstracting and modularising) observed during the problem-solving process. All these aspects encompass the CT skills outlined by Dagiene et al. (2017), which, in this specific context, can be assessed through qualitative data obtained from observations conducted during the sessions with MBG.

Qualitative data were collected through direct observations and field notes during the sessions, enabling the identification of CT dimensions explored through students' interactions with the board games. The dimensions analyzed included abstraction, algorithms, distributed computing, decomposition, debugging, conditional logic, incremental thinking, pattern recognition, and simulation.

The MBG were selected based on specific criteria and according to its mechanics, using the CTLM-TM model (Machuqueiro & Piedade, 2024), which guides the selection of games capable of fostering Computational Thinking (CT) skills (Englestein & Shalev, 2022; Machuqueiro & Piedade, 2023). Four games were used (Table 2) to promote the development of CT. Statistical techniques were used to analyse quantitative data, particularly to compare the performance between the groups through the application of the Student t-test. At the same time, qualitative data were categorised according to the selected dimensions of CT observed (Dagiene et al., 2017), and content analysis techniques were used (Bardin, 2016).

Table 2

Analysis of the Selected Modern Board Games using the CTLM-TM Model

<table>
<tr><th>Board Game</th><th>CT Learning Mechanics</th><th>MBG Mechanics</th></tr>
<tr><td>King of Dice: The Board Game (KDB)</td><td rowspan="4">Logical Reasoning
Algorithmic Thinking Debugging
Simulation
Data Analysis
Pattern Recognition
Object-Oriented Programming
Abstraction
Decomposition
Evaluation
Incremental Thinking; Modeling
Conditional Logic.</td><td rowspan="4">Turn-Based
Cooperation
Worker Placement/ Resource Management
Hand Management
Simultaneous Actions
Pattern Building
Action Queues
Simulation
Modular Board/ Tile Allocation
Real-Time</td></tr>
<tr><td>Rossio (ROS)</td></tr>
<tr><td>PreHistorias (PHI)</td></tr>
<tr><td>Festival (FES)</td></tr>
</table>

Results

The experimental intervention consisted of weekly 60-minute sessions, utilising each of the selected games over a four-week period. The pre-test was administered before the start of the intervention, the intermediate test was conducted after eight weeks, and the post-test was administered at the end of the process. The following figures 1 and 2 illustrate some of the game-based sessions carried out in the classroom, organised into several groups of students.

Figure 1

Setup of the Modern Board Games King of Dice: The Board Games and Rossio

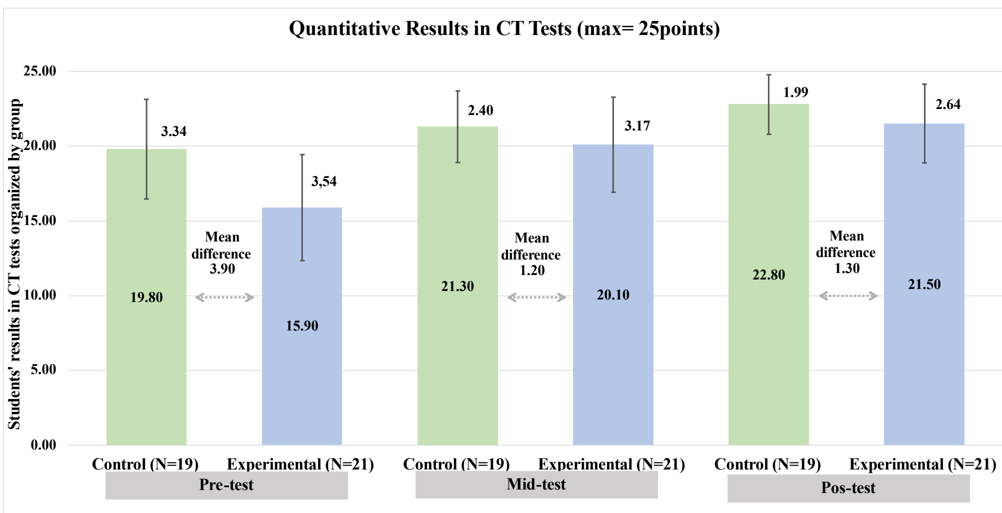
Figure 2

Setup of the Modern Board Games PreHistorias and Festival

The first research question aimed to examine the impact of using modern board games on the development of Computational Thinking in primary school students. The analysis of the results from the Computational Thinking tests indicates a positive impact on students' competencies, particularly in the experimental group (Figure 3). This group showed an increase of 5.66 points, representing a percentage gain of 35.68%, compared to the pre-test and post-test results. In contrast, the control group recorded a less significant improvement of only 2.95 points. After the administration of the pre-test, it was observed that there was no equivalence between the groups, with one group scoring, on average, 3.90 points higher than the other. The comparative analysis of means using the Independent Student's t-test indicated that this difference was statistically significant (t(38) = -3.65; p = <0.001). Given that the class groups had been pre-established by the school, the decision was made to conduct the intervention with the group that had demonstrated lower initial performance.

Figure 3

Students' results in Computational Thinking tests calculated for the experimental group and the control group. Values are presented as Mean ± Standard Deviation



In the intermediate test conducted eight weeks after the start, a significant improvement was observed in the experimental group's results (M=21.30; +5.4), which approached the performance of the control group (M=20.10; +0.3). The average difference between the groups was only 1.20 points. In other words, following the pedagogical implementation of the first two games, the groups became equivalent in terms of the Computational Thinking competencies demonstrated in the test. The comparative analysis of mean scores from the intermediate test revealed differences that were not statistically significant.

Finally, following the conclusion of the pedagogical intervention, the post-test results showed a slight improvement compared to the intermediate test (experimental group: +1.4; control group: +1.5), with a non-significant mean difference of 1.30 points between the groups ($t_{(38)}$ = -1.70; p =0.05).

To address the research question, a further comparison was conducted between each group's pre-test and post-test results using the paired samples Student's t-test, which allows for the comparison of results from the same subjects over time. As previously mentioned, the intervention enabled students in the experimental group to improve their scores by 5.66 points. The comparative analysis of means revealed that this difference was statistically significant (*t(20)=9.59; p<0.001)*, with an effect size of 2.09 calculated using Cohen's *d*. The control group showed a smaller increase compared to their initial results, with an improvement of only 2.95 points. The comparative analysis of means indicated that this difference was also statistically significant (*t(18)=3.77; p<0.001*; however, the effect size was smaller (*d=0.86)*, reflecting a lower magnitude of impact.

The evidence highlights the potential of the selected modern board games in supporting the development of Computational Thinking concepts proposed by Brennan and Resnick (2012).

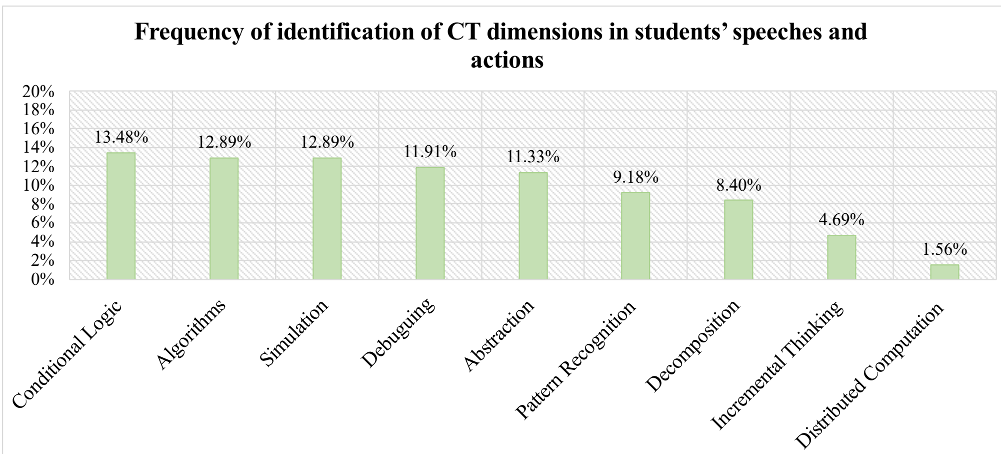
To address the second research question, "How do the mechanics of Modern Board Games influence the development of Computational Thinking?", a content analysis was conducted on field notes and the students' recorded discourse during the gameplay sessions. Based on a predefined category framework, the analysis aimed to identify evidence of the presence of different dimensions of Computational Thinking in students' discussions while solving problems and engaging with game scenarios. Figure 4 presents the frequencies with which each dimension of Computational Thinking was identified, highlighting "Conditional Logic" (13.48%), "Algorithms" (12.89%), "Simulation" (12.89%), "Debugging" (11.91%), and "Abstraction" (11.33%). The following segment presents an example of a dialogue from a group of students in which the presence of "Simulation" actions can be identified during one of the games. In this dialogue, the students/players engaged in hypothetical discussions about possible actions and their potential outcomes.

> ***Mike.LQ:*** *I can already see it coming…*
> ***Kilo.IM:*** *What?*
> ***Mike.LQ:*** *What you're going to do… you're going to take one of mine…*
> ***Kilo.IM:*** *I don't know… I could take Oscar.MA's piece — it's in the middle and close to the flames. I don't have any flames yet.*
> ***Oscar.MA:*** *But he has more points and more castles… that doesn't really make sense…*
> ***Kilo.IM:*** *Actually, it does. I won't fall behind him, and you'll be left without castles.*
> ***Oscar.MA:*** *Right… (showing some discontent).*

__Kilo.IM:__ But no, I'll place it here… I'll take this blue one and stay near a castle that's already mine. That way, I score more points. If I placed it there, I'd end up tied.
__Mike.LQ:__ Why? Over there you'd also be close to one of yours… and you'd already said it was his.
__Kilo.IM:__ But you've got more points and you're ahead. If I take Oscar.MA's piece, I tie with you. But if I take this blue one here (pointing to the tile he intended to swap), I move into first place. I could also take one from there, so you'd have fewer points… but if I manage to get an ice tile here, I get closer to a line of four in a row.

Figure 4

Frequency of identification of CT dimensions in students' speeches and actions



The main findings of the study highlight the potential of modern board games to provide meaningful challenges and strategic opportunities that foster the development of Computational Thinking. These results are consistent with several studies that have examined the use of modern board games as pedagogical tools for promoting this type of competence (Bayeck, 2024; Berland & Duncan, 2016; Tsaraya et al., 2019). According to Somma (2020), board games are computational artefacts, particularly effective in developing Computational Thinking, as they are rich in variables and function as authentic software systems.

Conclusions

These findings suggest that integrating modern board games into classroom practice can serve as an effective strategy to support the development of Computational Thinking in primary education. By engaging students in problem-solving, planning, and simulation tasks within game-based contexts, educators can create authentic learning environments that align with several Computational Thinking skills.

Future research could further explore which game mechanics are most effective in fostering specific dimensions of Computational Thinking, as well as investigate the long-term impact of sustained game-based interventions across diverse educational settings and student populations.

Notes

This work was supported by National Funds through FCT-Portuguese Foundation for Science and Technology, I.P., under the scope of UIDEF-Unidade de Investigação e Desenvolvimento em Educação e Formação, UIDB/04107/2020, https://doi.org/10.54499/UIDB/04107/2020.

References

Bardin, L. (2016). *Análise de conteúdo*. São Paulo, SP: Edições 70.

Barr, V., & Stephenson, C. (2011). Bringing computational thinking to K-12: What is involved and what is the role of the computer science education community? *ACM Inroads*, *2*(1), 48–54. https://doi.org/10.1145/1929887.1929905

Bayeck, R. Y. (2024). Understanding computational thinking in the gameplay of the African Songo board game. *British Journal of Educational Technology, 55*(1), 259-276. https://doi.org/10.1111/bjet.13353

Bell, T., & Vahrenhold, J. (2018). CS unplugged—How is it used, and does it work? In Böckenhauer, HJ., Komm, D., Unger, W. (Eds.), *Adventures between lower bounds and higher altitudes*. *Lecture Notes in Computer Science, Vol. 11011*. Springer International Publishing. https://doi.org/10.1007/978-3-319-98355-4_29

Berland, M., & Duncan, S. (2016). Computational thinking in the wild: Uncovering complex collaborative thinking through gameplay. *Educational Technology, 56*(3), 29–35.

Berland, M., & Lee, V. R. (2011). Collaborative strategic board games as a site for distributed computational thinking. *International Journal of Game-Based Learning, 1*(2), 65–81. https://doi.org/10.4018/ijgbl.2011040105

Bers, M. (2020). *Coding as a Playground*. Routledge, Taylor& Francis Group.

Bers, M. (2023). El desarrollo de Scratch-Jr: El aprendizaje de programación en primera infancia como nueva alfabetización [The development of Scratch-

Jr : Programming learning in early childhood as a new literacy]. *Virtualidad, Educación y Ciencia*, *14*(26), 43-62.

Brennan, K., & Resnick, M. (2012, April 13–17). New frameworks for studying and assessing the development of computational thinking (Vol. 1, 25 pp.). In *Proceedings of the 2012 Annual Meeting of the American Educational Research Association, Vancouver, Canada (Vol. 1)* p. 25. http://scratched.gse.harvard.edu/ct/files/AERA2012.pdf

Bocconi, S., Inamorato dos Santos, A., Chioccariello, A., Cachia, R., Kampylis, P., Giannoutsou, N., Dagienė, V., Punie, Y., Wastiau, P., Engelhardt, K., Earp, J., Horvath, M., Jasutė, E., Malagoli, C., Masiulionytė-Dagienė, V., & Stupurienė, G. (2022). Reviewing computational thinking in compulsory education: State of play and practices from computing education. In A. Inamorato dos Santos, R. Cachia, N. Giannoutsou, & Y. Punie (Eds.) *Publications Office of the European Union 2022*. https://data.europa.eu/doi/10.2760/126955

Chen, H. E., Sun, D., Hsu, T.-C., Yang, Y., & Sun, J. (2023). Visualising trends in computational thinking research from 2012 to 2021: A bibliometric analysis. *Thinking Skills and Creativity*, *47*, 101224. https://doi.org/10.1016/j.tsc.2022.101224

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Sage Publications.

Dagienė, V., Sentance, S., & Stupurienė, G. (2017). Developing a two-dimensional categorization system for educational tasks in informatics. *Informatica, 28*(1), 23–44. https://doi.org/10.15388/Informatica.2017.119

Engelstein, G., & Shalev, I. (2022). Building blocks of tabletop game design: An encyclopedia of mechanisms (2nd ed.). CRC Press. https://doi.org/10.1201/9781003179184

Grover, S., & Pea, R. (2017). Computational thinking: A competency whose time has come. In S. Sentence, E. Barendsen, & C. Schulte (Eds.), *Computer Science Education*, (pp. 19-38). https://doi.org/10.5040/9781350057142.ch-003

Kafai, Y., Proctor, C., & Lui, D. (2020). From theory bias to theory dialogue: Embracing cognitive, situated, and critical framings of computational thinking in K-12 CS Education. *ACM Inroads*, *11*(1), 44–53. https://doi.org/10.1145/3381887

Lee, I., Martin, F., Denner, J., Coulter, B., Allan, W., Erickson, J., Malyn-Smith, J., & Werner, L. (2011). Computational thinking for youth in practice. *Education Inquiry*, *2*(1), 32–37. https://doi.org/10.1145/1929887.1929902

Machuqueiro, F., & Piedade, J. (2022). Development of computational thinking using board games: A systematic literature review based on empirical studies. *Revista Prisma Social, 38*, 5–36. https://revistaprismasocial.es/article/view/4766

Machuqueiro, F., & Piedade, J. (2023). Exploring the potential of modern board games to support computational thinking. *2023 International Symposium on Computers in Education (SIIE)*, Setúbal, Portugal, 2023, (pp. 1-8). https://doi.org/10.1109/SIIE59826.2023.10423693

Machuqueiro, F., & Piedade, J. (2024). Modern board games and computational thinking: Results of a systematic analysis process. *Educational Media International, 61*(1-2), 161–183. https://doi.org/10.1080/09523987.2024.2359762

Menon, D., Romero, M., & Viéville, T. (2019). Computational thinking development and assessment through tabletop escape games. *International Journal of Serious Games*, *6*(4), 3–18. https://doi.org/10.17083/ijsg.v6i4.319

Moreno-León, J., Robles, G., & Román-González, M. (2015). Dr. Scratch: Automatic analysis of scratch projects to assess and foster computational thinking. *Revista de Educación a Distancia*, *46*, 1-23.

Munasinghe, B., Bell, T., & Robins, A. (2023). Unplugged activities as a catalyst when teaching introductory programming. *Journal of Pedagogical Research, 7*(2), 56–71. https://doi.org/10.33902/JPR.202318546

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. BasicBooks.

Papert, S. (1993). *The children's machine: Rethinking school in the age of the computer*. BasicBooks.

Relkin, E., Johnson, S., & Bers, M. (2023). A normative analysis of the TechCheck computational thinking assessment. *Educational Technology & Society, 26*(2), 118-130. https://doi.org/10.30191/ETS.202304_26(2).0009

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, *22*, 142–158. https://doi.org/10.1016/j.edurev.2017.09.003

Sinek, S. (2009). *Start with why: How great leaders inspire everyone to take action*. Portfolio.

Somma, R. (2020). *Coding in the classroom: Why you should care about teaching computer science*. No Starch Press.

Sousa, M. (2013). Mastering modern board game dDesign to build new lng experiences: The MBGTOTEACH Framework. *International Journal of*

Games and Social Impact, *1*(1), 68–93. https://doi.org/10.24140/ijgsi.v1.n1.04

Tsarava, K., Moeller, K., & Ninaus, M. (2019). Board games for training computational thinking. In M. Gentile, M. Allegra, & H. Söbke, (Eds.), *Games and Learning Alliance. GALA 2018: Lecture Notes in Computer Science, 11385* (pp. 75–85). Springer. https://doi.org/10.1007/978-3-030-11548-7_9

Tseng, C., Doll, J., & Varma, K. (2019). Exploring evidence that board games can support computational thinking. In *Proceedings of International Conference on Computational Thinking Education (pp.* 61–64).

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 140–158.

Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *366*(1881), 3717–3725. https://doi.org/10.1098/rsta.2008.0118

Yang, D., Yang, Z., & Bers, M. U. (2023). The efficacy of a computer science curriculum for early childhood: Evidence from a randomized controlled trial in K-2 classrooms. *Computer Science Education*, *35*(2), 1–21. https://doi.org/10.1080/08993408.2023.2279198

Zapata Cáceres, M., Martín-Barroso, E., & Román-González, M. (2021). BCTt: Beginners Computational Thinking Test. *In Understanding computing education: Vol. 1* (pp.46-56). Proceedings of the Raspberry Pi Foundation Research Seminar Series. rpf.io/seminar-proceedings-vol-1-zapata-caceres-et-al

Author Details

João Piedade
UIDEF, Instituto de Educação, Universidade de Lisboa
PORTUGAL
jmpiedade@ie.ulisboa.pt

Fábio Machuqueiro
UIDEF, Instituto de Educação, Universidade de Lisboa
PORTUGAL
fmachuqueiro@edu.ulisboa.pt

Communication and Engagement are Key - Academics' Views of Opportunities and Challenges with Fully Online Asynchronous Teaching

Christine Armatas & Vikki Pollard
Australian Catholic University
AUSTRALIA

Abstract

In this paper, we present findings from surveys and interviews of academics teaching in asynchronous, fully online courses at an Australian university, which highlight challenges and opportunities with this delivery mode. These include communicating and engaging with students and effective facilitation of asynchronous online discussions (Cohen & Donaldson, 2021). The usefulness of professional learning and support to help develop new, and transfer existing, skills to effectively teach in a fully online asynchronous learning mode was also acknowledged. Examples and insights from interviewees add to the research on effective online learning and are valuable for researchers and practitioners alike.

Communication and Engagement are Key - Academics' Views of Opportunities and Challenges with Fully Online Asynchronous Teaching

Student expectations about learning anywhere and anytime (Hussin, 2018) have led to an increase in asynchronous online learning (AOL). However, there is little research focusing on academic experiences of working in these courses (Earnshaw & Al-Sharif, 2024). Asynchronous, fully online delivery has the additional constraint of no live communication with students and a reliance on other communication modes, mainly discussion boards. In this paper, we present findings from surveys and interviews of academics teaching in fully online, asynchronous courses at an Australian university, which highlight challenges and opportunities with this delivery mode. Understanding how academics adapt to teaching in an AOL environment is important for assisting them to effectively teach in this delivery mode. The work reported in this paper aimed to gain insights and practical examples to add to the research on effective online learning, with a specific focus on the challenges and opportunities the AOL mode offers.

Asynchronous Online Learning

The experience of remote learning during COVID has embedded online delivery in the higher education landscape, including asynchronous online learning (AOL), defined as learning through the internet "where students engage with instructors and fellow students at a time of their convenience and do not need to be co-present online or in a physical space" (Singh & Thurman, 2019, p. 302). A common structure for these asynchronous online courses is described by O'Connor (2022) as units (or subjects) that are based around weekly learning objectives and targeted activities which scaffold toward the assessment. Learning activities can include written content, videos, and discussion boards (Reichgelt & Smith, 2024).

The asynchronous and fully online courses offered at the university where this work was conducted follow a similar format to that described above. These courses were a new venture seeking to expand the University's online footprint and involved setting up a new central team dedicated to the design and delivery of those courses. The role of the faculty was to provide content and teach. Units for these courses are developed under an in-house Online Program Manager (OPM) model (Nguyen & Gilmore, 2024), by a central team within the university. A range of learning professionals, including learning designers, learning technologists, graphic designers, editors, and multimedia experts, work with academics to develop units based around a design template. The template includes learning activities such as short videos, online quizzes, readings, etc. In all these units, students are expected to undertake activities and to also engage in weekly discussion activities moderated by online tutors. Academics do not have editing access in the units either during the design process or when the units are delivered. As a result, there is a heavy reliance on announcements and discussion boards by those teaching a unit in AOL mode to establish teacher-presence and communicate and engage with students.

With this type of templated unit design, referred to as "duet-design" (Chase, Ross & Robbie, 2017, p. 3), academics provide the curriculum, the central group does the design of the unit in the learning management system (LMS) and the academic does the final sign-off. Delivery of the units is undertaken by academics from the faculty – sometimes these academics have also worked on the unit development, but often they have not. They are also often staff on casual teaching contracts with the University. Although not the focus of this paper, OPMs and their derivatives, such as the one at our university, have been criticized for using processes that differ significantly from other forms of academic work, resulting in what has been called the "unbundled academic" (Ivancheva & Courtois, 2024), where traditional roles are broken into components, some of which are undertaken by non-academics. These OPM arrangements have also been criticized because much of the teaching is done by precarious academic workers unable to find secure university employment (Ivancheva & Courtois, 2024).

Given these concerns, and because this online venture, with its very different development and delivery model, was new (and very unfamiliar) to our university, academics involved in the delivery of these asynchronous, fully online units were supported during the first time they taught a unit. This support included a self-paced online course which modelled the unit template design and style of delivery, including learning activities, videos, readings, quizzes, and asynchronous discussions. The course introduced academics to the delivery model and gave them the experience of what it is like to be a student in one of these AOL units. In addition, during their first time teaching a unit, academics participated in a mentoring/coaching program and community of practice with other academics also new to this delivery mode. Participation in this professional learning and support, though not compulsory, was strongly encouraged and incentivized by being included in the academic's workload and/or remuneration.

Access to professional development and support is one of several factors that have significant implications for online teaching and student learning (Perrotta & Bohan, 2020). Academic staff teaching in online modes need support to develop their practices (Stone, 2017; Watson et al., 2023), the lack of which may result in academic teaching staff reproducing their practices from other modes (Cohen & Donaldson, 2021) even though these practices may not be appropriate for online learning. As such, the provision of professional development for online teaching is now an expected standard in higher education. However, provision of professional development and support does not address all the challenges that have been identified with online learning generally, or asynchronous online learning specifically, which is discussed next.

The Challenges of Asynchronous Online Delivery

Various challenges with online learning are identified in the literature which fall into several broad categories. These include accessibility and technical issues (Cahyani et al., 2021); not wanting to teach online but having to (Pomerantz & Brooks, 2017); concerns about student engagement, participation and enjoyment (Cahyani et al., 2021); doubts about the effectiveness of online learning for comprehension and topic mastery (Cahyani et al., 2021), and misconceptions about the effectiveness of online learning for students' learning (Pomerantz & Brooks, 2017). Access to professional development and support in the form of coaching or mentoring, how connected faculty feel to the campus community, and academic freedom and curriculum control also have significant implications for online teaching and student learning (Perrotta & Bohan, 2020).

All of the above challenges apply to AOL. In addition, a significant challenge with teaching in an AOL environment is connecting with students, in particular how to do this compared to face-to-face teaching and the impact this has on supporting students in their learning journey. Teacher- and social-presence in an asynchronous online course, particularly where discussion boards are the main method of interaction with students, is known to impact student satisfaction and retention (Gassell et al., 2021). Watson et al. (2023) note that, in asynchronous online courses, higher rates of perceived learning and better learning experiences are associated with stronger teaching presence. In their study of students enrolled in a capstone course in an MBA degree at a U.S. university, they found students value teaching presence in the form of recorded content lectures, detailed performance feedback, and quick responses to their queries. However, teaching in an AOL environment can be challenging as it requires adjusting to a new way of teaching, which involves a shift from teacher-centred to student-centred pedagogies (Perrotta & Bohan, 2020). Furthermore, many academics may not know how to teach online or may prefer to teach as they were taught. In particular, there are specific skills needed for effective facilitation of asynchronous online discussions (Cohen & Donaldson, 2021) which may not come naturally (Gassell et al., 2021), both of which further underscore the need for support and professional development for academics to effectively teach online.

The inclusion of discussion boards in AOL courses is typical and often the main means of communication between students and between students and teachers. In their systematic review of 35 papers relating to asynchronous online discussions in higher education, Fehrman and Watson (2021) found that, although these online discussions are ubiquitous, there is little consensus on how they should best be used and scant research on alternatives to asynchronous online discussions. From their review, they note that, in the absence of face-to-face interactions, asynchronous online discussions need to "provide community, instruction, and participation for students" (p. 203). They also note that there are benefits for students to be gained in terms of their learning, from interaction with peers and teachers. However, the quality and effectiveness of asynchronous online discussions is variable, with the need for structure being agreed upon in the literature as being critical for effectively engaging students and guiding their learning (Fehrman & Watson, 2021).

Online discussions clearly play an important role in providing opportunities for interaction with peers and teachers, which in turn can lead to enhanced learning outcomes. This importance is underscored by research that shows, for an asynchronous online course with no interaction with an instructor or with others in the course, students' performance decreased compared to that of students in the equivalent in-person experience (Jensen et al., 2022). Faulconer et al. (2022) note that tasks associated with asynchronous online discussions, such as reading and responding to peers' posts and synthesizing material from multiple sources, have high cognitive load for students. Similarly, reading, responding to and moderating discussion board posts can have a high cognitive load for teachers, as well as being quite time consuming (Fehrman & Watson, 2021). This contrasts with teaching in a face-to-face environment where student queries can be responded to in real-time, without the need to formulate a response as a discussion post or email. Given the important role asynchronous online discussions have in connecting with students in an AOL environment, and the lack of consensus relating to how discussion boards should best be used, it is likely that connecting with and engaging students via discussion boards could be the main challenge that academics face when teaching an AOL course.

Based on the above research findings, there are multiple challenges for those teaching online using an asynchronous delivery model. A significant challenge is how to engage students and promote interaction and participation, particularly when asynchronous online discussions are the main way students communicate with one another and the teacher. How to assess whether (and what) students are learning in an environment where students are not "seen" is also a challenge, as is transitioning to more student-centred teaching practices. Developing the necessary skills and being supported to teach online is also critical for academics engaged in teaching in AOL environments. Understanding the extent to which these challenges are present and how they impact on academics' experience of asynchronous online teaching is important in supporting those teaching in AOL environments, particularly for the first time.

Research Aims

Our research aims to investigate the experiences of academics at our university teaching units as part of a new AOL initiative, where many teachers experienced this mode of delivery for the first time. So, our aim was to better understand factors that help or hinder their effectiveness as teachers and how they can be supported in their roles as online facilitators, beyond the professional learning and support already provided. Survey responses and interview transcripts provided by academics teaching subjects delivered in an AOL unit were analysed to address three research questions. The first related to their perception of teaching in an AOL environment based on their experience. In particular, given most had not previously taught asynchronously online, we were interested in how they adapted their teaching practice, particularly in regard to communicating and engaging with students via announcements and discussion boards, to this new learning environment. The second concerned what factors facilitated or inhibited their experience of teaching in an AOL environment. Lastly, based on their experience, we wanted to know how well supported they felt in their role and what improvements or enhancements could be made.

Method

The original data in this paper was obtained from an ethics-approved research project. The project had two aims. The first was to understand the experiences of academics teaching units of study in AOL environments and their experiences of the Professional Development (PD) designed to support them to teach, which is the focus of this paper. The second was to understand the experiences of academics developing units for this delivery mode. There were two methods of data collection. The first was a short online survey consisting of eleven questions, several of which were focused on teaching, with two of these requiring a short answer response. Those who received the survey were also invited to undertake an interview.

Participants

The survey was sent to 108 academic staff who had undertaken teaching of at least one unit in the AOL format as part of the University's new online venture and who had also participated in the professional learning provided to academics who were teaching in the AOL delivery mode at the university for the first time. As described previously, this Professional Development included a self-paced, online course that modelled the student experience and teacher role in the asynchronous online units, together with a coaching program/community of practice where academics met regularly during their first term of teaching in the AOL mode to discuss their experience, ask question, share resources and support one another.

Academic staff who had both undertaken teaching and participated in the professional learning were selected using "purposeful sampling [whereby] the researcher specifically seeks participants who meet a set criteria" (Croxford et al., 2019, p. 4). In addition to these two criteria, academics also needed to have taught units of study delivered in the AOL mode over the past four years. We received 15 responses to the survey, with six respondents agreeing to be interviewed. Further, we randomly selected 30 academics from a list of 131 who had undertaken unit development, knowing that there would be a crossover with some who had also taught. A further three participants for interviews were recruited from this method, making nine interviews in total.

Data Collection

Data were collected using two methods: survey and interviews. The survey consisted of nine items which respondents rated using a 5-point scale (1=very dissatisfied, 2=dissatisfied, 3=neither dissatisfied nor satisfied, 4=satisfied, 5=very satisfied). Two items on the survey were relevant to teaching a unit in the AOL delivery mode, while there were three items relating to the Professional Development they received in their first term of teaching an AOL unit. There were also two open-ended items on the survey. The first asked them to describe their experience of the Professional Development (i.e., the online course and coaching), while the second asked them about their experience of teaching in the AOL delivery mode.

The interview questions were developed using research on AOL and our own experience having worked on these units as academics based in the central unit team. The interviews were undertaken on Microsoft Teams, took approximately forty minutes, and were recorded. The interviews included several questions related to the experiences of teaching, such as timelines, workload, satisfaction and autonomy. As these were open-ended interviews, much arose about the experiences of teaching outside of the scripted questions. We followed relevant themes when they arose. The interviews were transcribed, reviewed, and sent to the interviewees for clarification. The data were then analysed for recurring themes "through careful reading and re-reading of the transcribed data" (Dawadi, 2020, p. 62). These themes were reviewed and revised with the second author. For the purposes of this paper, participants were given anonymous identifying initials such as AE, MA, and CS, and the university has been de-identified.

Results

Data used to address our research questions came from two sources. The first was responses to relevant items in the survey, including responses to open-ended questions. The second source was analysis of transcripts of interviews with academics who had also completed the survey. Survey results are presented first, followed by the interview analysis.

Survey Results

A total of 15 responses were received to the survey. As the survey was designed to collect data for the ethics approved project, survey items asked about experiences with developing asynchronous online units, as well as experience teaching these units and the professional development and support received. Two items were relevant to the experience of teaching a unit in the AOL delivery mode. As shown in Table 1., survey respondents tended to agree (*M*=3.6, *SD*=0.9) that they were satisfied they understood the requirements for teaching their unit in the AOL mode, while they tended to be neutral about whether or not they were satisfied with the experience of teaching the unit (*M*=3.1, *SD*=1.0). Respondents also tended to agree that they were satisfied with the two items relating to the online course (*M*=3.8, *SD*=0.9 and *M*=3.6, *SD*=1.2 respectively) and the experience of coaching (*M*=3.7, *SD*=1.0).

Table 1

Count of Responses to Teaching-related Items on Survey

Statement	Response Count (*N*=15)					
	1	**2**	**3**	**4**	**5**	**N/A**
Understanding the requirements for teaching for asynchronous delivery	0	2	4	7	2	0
The experience of teaching the unit in AOL mode	0	5	4	5	1	0
Assistance provided when undertaking the online modules	0	1	3	6	3	0
The content of the online modules	0	3	3	4	4	0
The experience of coaching	0	2	3	5	3	0

Note. 1=very dissatisfied, 2=dissatisfied, 3=neither dissatisfied nor satisfied, 4=satisfied, 5=very satisfied

Overall, more respondents agreed they were satisfied or very satisfied with these statements than dissatisfied or very dissatisfied, although only marginally so for the item about their experience teaching the unit in AOL mode. Looking at the responses to the open-ended items, for the five respondents who indicated dissatisfaction with their experience of teaching the unit, four provided responses to the open-ended question about their teaching experience in the AOL delivery mode. The reasons for this dissatisfaction varied. Challenges with lack of interaction with students, finding it hard to motivate and support students, and the amount of time teaching in an AOL mode took up was a negative part of the experience for one respondent. Another said their experience teaching the unit "felt more like an IT role than a teaching/facilitating role" while the experience for another was described as "mixed". For one it was frustration with not being able to edit content in the unit, which was also noted by a respondent who indicated that they were satisfied (score of 4 for the item) with their experience of teaching the unit in AOL mode. The responses provided by respondents who were satisfied or very satisfied with the teaching experience included that it was an enjoyable learning experience and that their unit was well structured and easy to follow.

A similar pattern of results was seen for the items relating to the professional learning and support they experienced, with more respondents being satisfied than dissatisfied. Overall, these academics were satisfied with the support they were provided while completing the online course, the content of the modules in the course, and their experience of the coaching/mentoring they received. Looking at their satisfaction rating for the item about understanding the requirements for teaching for asynchronous delivery, it appears that the online course provided a solid introduction to, and information about, what was required in teaching these units.

Analysis of Interview Transcripts

Three broad themes emerged from the analysis of the interview transcripts, each of which is discussed next.

Online Compared to Face-to-face Teaching

A number of the interviewees commented that teaching online was very different to teaching face-to-face, giving a range of examples as to why they believed this. For CS, teaching online was “way harder” as it required extensive preparation and a “level of precision and depth” different to face-to-face teaching, and this was a challenge from a workload perspective. LL described themselves as being “unprepared for the differences that were required for teaching asynchronously online”, stating that their discipline is “very face-to-face”. This academic referred to face-to-face teaching as being “the normal way” in their school.

Another difference noted between teaching face-to-face and online which presented a challenge was the content being taught. DM said that he was sure “there’s others, other academics that are trying to teach content that might not be as suited to the online format, and I’m sure they have issues and complaints.” MA expressed similar concerns noting that they have colleagues who don’t want to teach in the AOL delivery mode again and others who said they would not teach in this mode even though they have not experienced it. PK also noted that their colleagues don’t want to teach in the AOL mode due to concerns about workload and negative student evaluations.

One interviewee (MA) said that teaching in the AOL delivery mode, “just doesn’t feel like teaching.” Reasons for this included not really feeling connected to the students, feeling that all they were doing is responding to questions and discussion posts, but not necessarily getting to know them, their strengths and weaknesses, the gaps in their knowledge. RS made the point that “You know, we didn’t become teachers to sit in front of a computer screen. We became teachers to teach people.” LL did not feel comfortable teaching some of the units in AOL mode, not because of the mode per se, but because it was not their field and so they didn’t “feel comfortable doing it”.

Several of the interviewees indicated that teaching in the AOL mode had helped them develop professionally. PK said that teaching the unit again and making changes helped develop their confidence, while RS took up the teaching of their unit because they are "always looking to improve my practice and, you know, learn something myself." SP observed that when teaching their AOL unit, they were "responding to the environment I'm working in, whereas when I'm in a classroom, I'm driving the environment." This academic noted that online teaching is "really confronting to a lot of academics" because of the need to change from a teacher-centred to a student-centred approach.

Communication, Engagement and the Discussion Boards

Although a number of interviewees acknowledged the importance of student autonomy in an AOL course, there were concerns expressed about the level of engagement with students and the ways of communicating and interacting with them in the unit. For example, D described themselves as "somewhat removed from that engagement with the student" but said that "we're still the teacher" and need "some level of engagement". SP indicted that "there is a little bit of an over-reliance on discussion boards" and that students, contrary to the reason for having them, "don't use the discussion boards for the learning, they use it to say I've done it." Similarly, MA felt that "there's quite a disconnect … between the lecturer or the online facilitator and the student" because they were just "answering questions and discussion forums or confirming what someone has said". SP noted that in the AOL environment, the teacher is "receiving what's coming from the students. So therefore I am actually waiting for what's happening." When students don't ask questions, this creates problems according to SP. When students did engage, RS found the experience enjoyable, saying it was "really informative" meeting with students online.

The AOL units at our university are marketed to students as having no timetabled, online sessions, but including the option to have a non-compulsory, online, consultation session weekly. These synchronous sessions were mentioned as an opportunity and means to engage with students. In addition to these weekly drop-in sessions available to all students, some academics offered additional synchronous sessions for students. For example, LL indicated that they had done some "small Zoom meetings" with students on an individual basis. This academic, along with MA, indicated that a regular, weekly meeting would have been beneficial, but noted that workload made this difficult. SP also indicated that they conducted at least two online sessions during the unit, which were designed to answer students' questions about the content and assessment tasks. For MA, a live forum was seen as a "really good way to build rapport with students, but also to connect with them".

In contrast, RS tried to encourage students to come along to "the drop in session" but "didn't really have the buy in from students." This academic noted that after having taught a unit in AOL mode several times, they could see that students were “missing certain things” and tried to work out how to “put that in there”. For them, not teaching face-to-face made it more challenging to show students who they were, as “that perception of you as the teacher obviously comes across quite differently in the online mode.” They indicated that they were recording extra videos and announcements to help address this and as a supplementary communication channel for connecting with students.

Having experience as an online student is valuable for understanding how to teach in an AOL delivery mode according to LL, who said “I think it is a very important remit … for lecturers and online facilitators to know how to engage with … and how to keep people motivated”. Prompt and regular online presence was also mentioned as important for keeping students engaged. CS noted that students appreciate prompt responses to their queries, even though they are studying asynchronously and often late at night or on the weekend. This creates workload issues which are “not necessarily explicitly understood”.

Support and Professional Development are Important

A number of the interviewees specifically mentioned the professional development and support they received. RS noted that the “things that have been the best about my experience have, I would definitely say, was in the first iteration having a mentor.” This was because the mentor’s role in supporting first time teachers “was absolutely key to helping me understand what I was doing and just having a sounding board”. LL also appreciated the professional development they received but said that having this just during the first time an academic taught in AOL mode was not “enough to sustain the academics”. For RS the coaching that was provided helped them to connect with another colleague who was teaching a similar unit in the AOL mode. This became a close professional relationship, where they supported one another and shared their experiences.

Discussion

Survey results and interview transcripts provided by academics teaching units delivered in an AOL environment were analysed to address three research questions. Generally, the results showed that these academics experienced many of the challenges that have previously been identified in the literature relating to online learning generally. They also specifically noted challenges associated with the AOL mode of delivery, as discussed next.

The first research question related to how these academics perceived AOL based on their experience, with a focus on how their teaching practice was impacted given the reliance on announcements and discussion boards for interacting with students. There were mixed views on this from both data sources. But regardless of whether the academic felt the experience was positive or negative, there was agreement that workload could be a significant issue and that teaching in an AOL environment was very different to face-to-face teaching. The reliance on discussion boards and announcements for communication and the lack of immediacy that face-to-face teaching provides were significant challenges that academics said they needed to adapt to when teaching in the AOL mode, which also contributed to their concerns over communicating with and engaging students. Various strategies to address this were described, which notably involved synchronous, online sessions which could be said to model the face-to-face teaching mode. This is problematic as a key feature of this new online venture is that these courses are promoted to students as allowing them to study anytime and anywhere. In addition to describing challenges, some of these academics also indicated that they enjoyed the experience and that there were benefits to having this experience. Interestingly, teaching in the AOL mode did not seem to have a good reputation amongst their colleagues. However, a number of academics did mention the positive benefits of teaching these units, including building their confidence and professional practice.

Our second research question related to what factors facilitated or inhibited their ability to teach effectively in an AOL environment. The inhibitory factors were most numerous and included aspects of the delivery model, adapting to teaching in the AOL environment, and challenges with communicating with and engaging students. Issues with the delivery model included not having editing rights over the content in the unit, which was frustrating when mistakes were identified or when the academic wanted to change or add something. Not having regular, timetabled online sessions that students were expected to attend was also seen as a significant drawback of the model. These academics also felt that the model had the potential for a high workload if not managed. The reliance on discussion boards and announcements for communication and the lack of immediacy that face-to-face teaching provides were significant challenges that academics said they needed to adapt to when teaching in the AOL mode, which also contributed to their concerns over communicating with and engaging students. Various strategies to address this were described, which notably involved synchronous, online sessions that could be said to model the face-to-face teaching mode, which as mentioned previously is not consistent with the model for this new online venture that promises no scheduled classes.

The final research question was about how supported these academics felt in their role and what improvements or enhancements could be made. Responses in the survey and interviews indicated that these academics felt quite supported while teaching their unit in AOL mode the first time, both through their mentor and networking with others. However, at least one academic indicated that this support was needed beyond the initial teaching experience. The online course that was designed to prepare academics for teaching an AOL course was rated positively and appeared to help them understand the requirements of teaching in this delivery mode, even if they were not particularly satisfied with the experience. So, while the online course helped with preparation for teaching, it did not prepare academics for the practicalities of teaching in an AOL environment, even though the course modelled all aspects of the delivery model, including the use of announcements and discussion boards for communicating with and engaging students.

Overall, the results indicate that while the professional development provided to these academics was helpful in preparing for the requirements of the delivery model, the actual experience of teaching in an AOL environment presented some challenges. Specifically, the need to have real-time interaction with students as a means for communicating with them, and engaging and motivating them, was a common theme amongst these academics. As a result, there were mixed views about the experience. While these academics noted many of the challenges already identified in the literature, their concerns about communication and engagement seem to be amplified in the AOL environment due to the lack of opportunities to interact with students synchronously.

Limitations and Future Directions

These results are consistent with findings reported in the literature relating to online learning generally. There is also close agreement between the survey and interview results. However, the sample size for both is quite small and so these results should be regarded as preliminary. Further data collection, using both surveys and interviews, is needed to confirm these findings. Data reported in this paper was collected as part of a larger study and these results suggest that a more detailed investigation of the teaching aspect of these new AOL units at our university is warranted. In particular, understanding how to extend the professional development these academics receive initially to support them whenever they are teaching in the AOL mode, is an important area for future research. Another important area to investigate is how to reconcile the academics' need to synchronously meet with students when these courses are promoted to students as not requiring students to be online at specific times. Many of these academics struggled with the discussion boards, but every unit had at least one weekly discussion topic as part of the design. How to address this requires further investigation – discussion boards play an important role in AOL, so understanding how to use them effectively, both as part of the unit design and its delivery, is critical. Whether there is a need to make the design template less dependent upon them and, if so, how this could be achieved, is an important question yet to be addressed.

Questions about whether the AOL model used in this new online venture at our university needs to be changed, and whether academics need to modify their practice, or both, require further investigation. In any case, the importance of, and challenges with, communication and engagement with students in online learning environments is persistent, especially for AOL environments.

References

Cahyani, N. M. W. S., Suwastini, N. K. A., Dantes, G. R., Jayantini, I. G. A. S. R. & Susanthi, I. G. A. A. D. (2021). Blended online learning: Combining the strengths of synchronous and asynchronous online learning in EFL context. *Jurnal Pendidikan Teknologi dan Kejuruan, 18*(2), 174-184.

Chase, A.-M., Ross, B., & Robbie, D. (2017). Improving digital assessment practice: A case study of a cross institutional initiative. *Journal of University Teaching & Learning Practice, 14*(2). https://doi.org/10.53761/1.14.2.5

Cohen, J., & Donaldson, K. (2021). The impact of learning design interventions on student satisfaction in an online higher education environment. *The International Journal of Technologies in Learning, 28*(1), 45–58. https://doi.org/10.18848/2327-0144/CGP/V28I01/45-58

Croxford, S., Thomas, C., Horvath, D., Buultjens, M., Stirling, E., Larsen, A., Stasis, T. R., Radcliffe, J., Buultjens, P., McDonald, S., & Forsyth, A. (2019). Successful team-based development of an online course with an external partner: An analysis of the perspectives of academics. *Journal of University Teaching & Learning Practice*, *16*(1), 1-20. https://doi.org/10.53761/1.16.1.5

Dawadi, S. (2020). Thematic analysis approach: A step by step guide for ELT research practitioners. *Journal of NELTA*, *25*(1-2). https://doi.org/10.3126/nelta.v25i1-2.49731

Earnshaw, Y. & Al-Sharif, M. A. B. (2024). Supporting tenure-seeking faculty in online programs: A critical discourse analysis. *Online Learning Journal*, *28*(4), 258-278. https://doi.org/10.24059/olj.v28i4.4136

Faulconer, E. K., Bolch, C. & Wood, B. (2023). Cognitive load in asynchronous discussions of an online undergraduate STEM course. *Journal of Research in Innovative Teaching & Learning, 16*(2), 268-280. https://doi.org/10.1108/JRIT-02-2022-0010

Fehrman, S. & Watson, S. L. (2021) A systematic review of asynchronous online discussions in online higher education. *American Journal of Distance Education, 35*(3), 200-213. https://doi.org/10.1080/08923647.2020.1858705

Gassell, C., Lowenthal, P. R., Uribe-Flórez, L. J., & Ching, Y.-H. (2021). Interaction in asynchronous discussion boards: A campus-wide analysis to better understand regular and substantive interaction. *Education and*

Information Technologies, 27, 3421–3445. https://doi.org/10.1007/s10639-021-10745-3

Hussin, A. (2018). Education 4.0 Made Simple: Ideas for teaching. *International Journal of Education and Literacy Studies*, *6*(3), 92-98. http://dx.doi.org/10.7575/aiac.ijels.v.6n.3p.92

Ivancheva, M., & Courtois, A. (2024). EdTech-mediated outsourcing and casualisation of academic labour: Toward a research agenda. *Work Organisation, Labour & Globalisation, 18*(1), 65–82. https://www.jstor.org/stable/48774386

Jensen, J., Smith, C. M., Bowers, R., Kaloi, M., Ogden, T. H., Parry, K. A., Payne, J. S., Fife, P., & Holt, E. (2022). Asynchronous online instruction leads to learning gaps when compared to a flipped classroom. *Journal of Science Education and Technology, 31*(6), 718-729. https://doi.org/10.1007/s10956-022-09988-7

Nguyen, C. & Gilmore, D. M. (2024). Approaches by OPM providers and higher education institutions to drive success, innovation, and remain relevant. In D.M. Gilmore & C. Nguyen (Eds.), Partnering with online program managers for distance education: Approaches to policy, quality, and lLeadership (pp. 2-28). Routledge.

O'Connor, K. (2022). Unbundling the university curriculum: MOOCs, online program management and the knowledge question. Springer. DOI: https://doi.org/10.1007/978-981-19-4656-1

Perrotta, K. & Bohan, C. (2020). A reflective study of online faculty teaching experiences in higher education. *Journal of Effective Teaching in Higher Education, (3)*1, 50-66. https://jethe.org/index.php/jethe/article/view/9/30

Pomerantz, J., & Brooks, D. C. (2017). ECAR study of faculty and information technology 2017, Research report. ECAR. https://library.educause.edu/-/media/files/library/2017/10/facultyitstudy2017.pdf

Reichgelt, J. & Smith, T. C. (2024). The effect of course delivery mode on student performance and student satisfaction: A case study. *Trends in Higher Education*, *3*(4). 872-884. https://doi.org/10.3390/higheredu3040050

Singh, V., & Thurman, A. (2019). How many ways can we define online learning? A systematic literature review of definitions of online learning (1988-2018). *American Journal of Distance Education*, *33*(4), 289–306. https://doi.org/10.1080/08923647.2019.1663082

Stone, C. (2017). *Opportunity through online learning: Improving student access, participation and success in higher education.* The National Centre for Student Equity in Higher Education (NCSEHE). Curtin University: Perth. https://www.acses.edu.au/research-policies/opportunity-through-online-learning-improving-student-access-participation-and-success-in-higher-education-2/

Watson, S., Sullivan, D. P., & Watson, K. (2023). Teaching presence in asynchronous online classes: It's not just a façade. *Online Learning, 27*(2), 288-303. https://olj.onlinelearningconsortium.org/index.php/olj/article/view/3231

Author Details

Christine Armatas
Centre for Education and Innovation
Australian Catholic University
christine.armatas@acu.edu.au

Vikki Pollard
Centre for Education and Innovation
Australian Catholic University
vikki.pollard@acu.edu.au

PROMOTING CRITICAL THINKING THROUGH THE NEWSPIRACY PROJECT: INSIGHTS FROM GREECE

Konstantinos Karampelas
University of the Aegean

Anastasia Pyrini and Georgios Sarrigeorgiou,
PARAGON-eduTech

Konstantinos Tsolakidis
University of the Aegean

and Despina Sarrigeorgiou
PARAGON-eduTech

GREECE

Abstract

In the context of increasing misinformation, this study explores how digital learning platforms can foster critical thinking and media literacy among educators. The research aims to evaluate the effectiveness, usability, and pedagogical value of the NEWSPIRACY platform and its Truth-Track tool. A mixed-methods approach was employed, combining Likert-scale survey data with thematic analysis of open-ended responses from 207 participants in Greece. Findings indicate that interactive, multimodal environments enhance users' analytical awareness, decision-making, and digital autonomy. The study contributes to understanding how educational design supports the development of informed and resilient digital citizens.

Introduction

The digital age has transformed the ways in which information is consumed and disseminated, granting users rapid access to vast digital content. While this democratization of knowledge is beneficial, it has also enabled the proliferation of fake news—deliberately misleading content disguised as legitimate journalism (Lobnikar et al., 2025; Zhang et al., 2022). Addressing the issue of fake news requires not only technical solutions but also educational strategies that promote media literacy and critical thinking. Citizens today act not only as consumers of information but also as distributors. The ability to critically evaluate information is essential for fostering informed and democratic participation. Critical thinking, paired with media literacy, allows individuals to examine claims, assess sources, and resist manipulation. Education plays a crucial role in promoting digital discernment and civic responsibility (Polizzi, 2025).

Critical Thinking and Media Literacy: Tools to Combat Fake News

Fake news tends to spread more rapidly than accurate news, driven by emotional appeal and novelty (Zhang et al., 2022). Vosoughi et al. (2018) found that false information on certain sites was significantly more likely to be shared. This phenomenon is linked to confirmation bias and motivated reasoning, which lead individuals to accept information that aligns with pre-existing beliefs (Bayrak et al., 2025; Pennycook & Rand, 2021). The "illusory truth effect" plays a significant role in the spread of misinformation; repeated exposure to falsehoods increases their perceived credibility (Fazio et al., 2015). Additionally, algorithmic content filtering worsens this problem by creating echo chambers that limit exposure to diverse viewpoints (Polizzi, 2025). Moreover, sharing fake news often serves social functions, such as signaling group loyalty, regardless of the content's accuracy. Therefore, effective interventions must address cognitive biases along with the emotional and social motivations behind dissemination of misinformation. Education is a powerful tool in combating fake news.

Educational Interventions for Digital and Critical Literacy

Zhang et al. (2022) demonstrated that multimedia-based media literacy programs significantly improve individuals' ability to identify fake news. By integrating digital literacy, learners can decode and challenge manipulative content (Lobnikar et al., 2025). Mihailidis (2018) argues that news literacy is a core civic skill that empowers learners to navigate the media landscape critically. Hobbs (2010) advocates for the integration of digital and media literacy across various subject areas, employing inquiry and problem-solving to develop critical media consumers. This educational model increases learner agency, fosters debate, and cultivates informed skepticism.

Digital literacy encompasses both functional skills and critical awareness. Polizzi (2025) introduces the idea of strategic disengagement, emphasizing the intentional avoidance of harmful digital content as a vital digital competency. Buckingham (2007) stresses the importance of understanding media production and its institutional context. This meta-cognitive layer of awareness is essential for resisting manipulation and participating ethically in digital environments. Sarmiento et al. (2025) found that learners with high levels of digital competence were significantly more accurate in distinguishing between reliable and misleading content. Digital competence includes recognizing bias, understanding algorithms, managing privacy, and engaging respectfully in online communication.

Critical Thinking and Reflective Engagement

Critical thinking is essential for resisting misinformation. It involves not only logical reasoning but also emotional intelligence and open-mindedness (Ennis, 1993; Facione, 1990). When applied to media content, critical thinking encourages learners to assess the intent, context, and reliability of information. Reflection-based strategies can enhance critical awareness; however, they must be carefully designed to avoid reinforcing biases, especially when identity is strongly tied to belief (Bayrak et al., 2025). Education systems should encourage diverse perspectives and foster intellectual humility, allowing learners to navigate complexity without resorting to oversimplified or polarized viewpoints.

Effective media literacy requires the integration of curricula and the use of multimodal pedagogical approaches. Zhang et al. (2022) highlight the benefits of visual and interactive content in enhancing memory retention and engagement. Educators must receive training in media pedagogy, and institutions should ensure that learners have equitable access to digital resources (Polizzi, 2025).

Scholars such as Buckingham (2007), Hobbs (2010), and Mihailidis (2018) advocate for inquiry-based, participatory learning strategies that empower learners to challenge dominant narratives, question sources, and engage in democratic discourse. At the policy level, it is crucial to address the digital divide. Without access to digital tools and education, marginalized communities remain vulnerable to misinformation. National and international collaboration can support inclusive and scalable media literacy initiatives, as well as teacher training.

Fake news poses complex educational challenges. A robust response must include critical thinking, media literacy, and digital competence, all fostered through inquiry-based and reflective education. Evidence indicates that multimedia and interactive learning environments effectively promote responsible digital engagement. Equipping learners with these skills enhances personal discernment and strengthens broader democratic resilience. As digital environments continue to evolve, education systems must remain committed to nurturing informed and critical citizens (Polizzi, 2025; Sarmiento et al., 2025; Zhang et al., 2022).

Methodology

This study aimed to evaluate the effectiveness, usability, and pedagogical impact of the NEWSPIRACY digital learning platform and the Truth-Track evaluation tool. As part of the Erasmus+-funded NEWSPIRACY project (2022–2025), coordinated by the University of Maribor in collaboration with partners such as the University of the Aegean, the University of Rijeka, and Valahia University of Targoviste, this initiative promotes digital and media literacy among both pre- and in-service educators, particularly in addressing misinformation through education.

A descriptive mixed-methods approach (Cohen et al., 2017) was adopted to examine participants' perceptions after engaging with the platform. The study focused on three research questions formulated as follows:

(1) How effective is the NEWSPIRACY learning platform in promoting users' understanding of fake news mechanisms and strategies to identify and counter them?
(2) To what extent do interactive digital tools, such as the Truth-Track evaluation environment, enhance users' critical thinking and decision-making when exposed to questionable news content?

(3) What are the perceived strengths, limitations, and usability challenges of the platform and the Truth-Track tool as reported by student teachers and educators?

The study adheres to ethical standards, ensuring voluntary participation, anonymity, and General Data Protection Regulation (GDPR) compliance. The sample comprised 207 student teachers and educators in Greece who used the platform and completed a comprehensive post-intervention survey. Quantitative data were collected using Likert-scale items to assess users' understanding of misinformation, confidence in evaluating content, and satisfaction with platform features. Descriptive statistics were employed to analyze response patterns and trends (Cohen et al., 2017). For example, items such as "I understand the mechanisms behind fake news" and "The Truth-Track tool helped me think critically" were used to capture relevant insights. Qualitative data were derived from open-ended survey questions that asked participants to reflect on their experiences and suggest improvements. Thematic analysis identified key themes, including accessibility, media richness, interactivity, and challenges related to navigation and content overload. These insights enhanced our understanding of user engagement and transformation. This aligns with frameworks for developing reflective judgment and critical dispositions (Bayrak et al., 2025; Mihailidis, 2018).

Findings

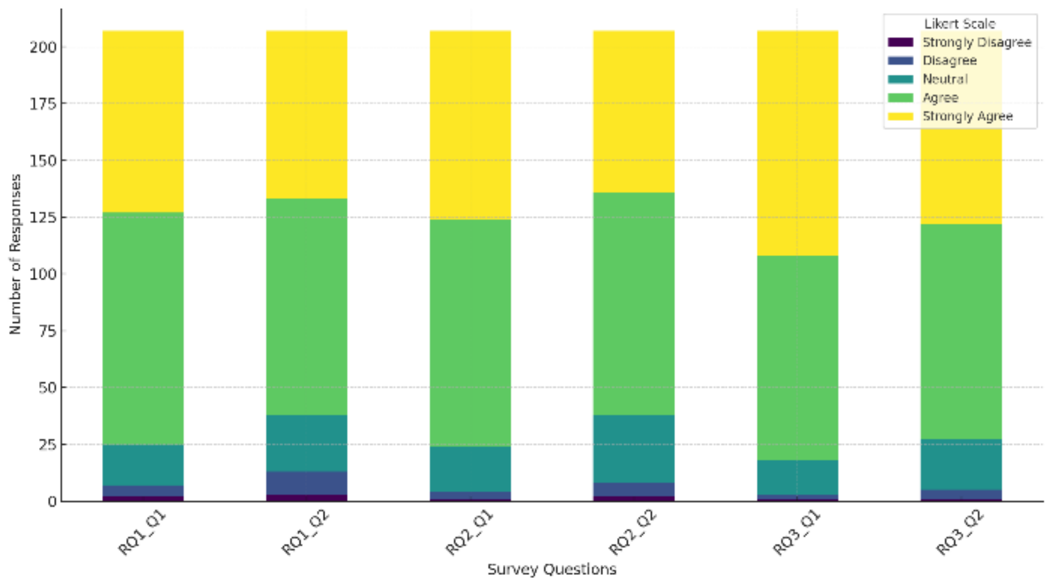
As shown in Figures 1 and 2, the NEWSPIRACY platform significantly enhanced users' competency in understanding fake news, critical thinking, and familiarity with its use, as evidenced by the relatively high mean scores.

Research Question 1: How Effective is the NEWSPIRACY Learning Platform in Promoting Users' Understanding of Fake News Mechanisms and Strategies to Identify and Counter Them?

Participants responded positively to statements measuring their understanding of fake news mechanisms, with mean scores of 4.2 and 4.0 on a scale of 5 (where 5 = strongly agree) for items related to comprehension and confidence. This suggests that the platform effectively enhanced users' foundational awareness of misinformation and its dynamics.

Figure 1

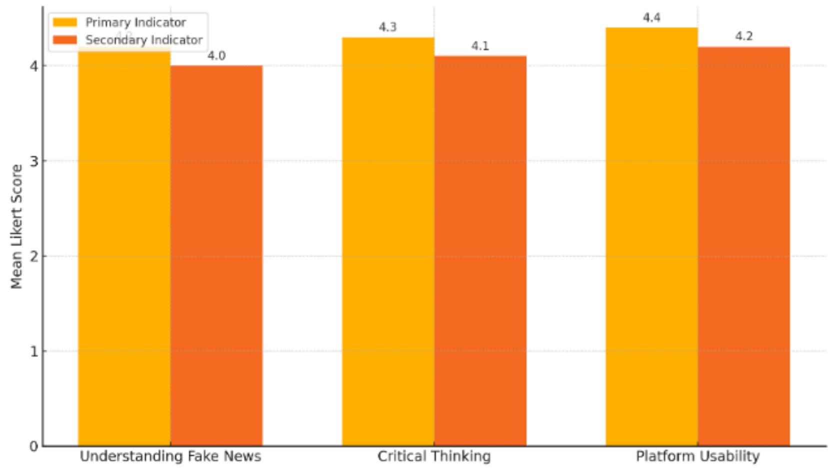
Distribution of Responses across Research Questions



Qualitative responses revealed that users experienced shifts in their interpretation and responses to online content. Several participants highlighted emotional manipulation and algorithmic bias as concepts they had not previously considered. One participant noted their unawareness of how headlines appeal to emotions, while another shared that the training encouraged them to examine who shared content and the motivations behind it. Reflections such as "I realized that I've been manipulated more often than I thought" highlight behavioral and cognitive changes.

Figure 2

Participants' Perceptions by Thematic Area



These reflections are directly linked to theoretical models of media literacy that emphasize reflexive understanding and contextual analysis. Rather than merely echoing content, participants internalized media analysis skills by applying them to real-world behaviors. This connection between theory and application reinforces the value of experiential learning models advocated by Buckingham and others.

Participant remarks on algorithmic awareness and message framing highlight an evolving understanding of the underlying digital structures, supporting the model of critical digital literacy (Polizzi, 2025). This development also reflects conceptions of civic agency within media literacy education (Mihailidis, 2018). Learners have progressed from merely understanding what fake news is to critically examining how and why misinformation operates within media environments.

Emotional and cognitive engagement reported by participants indicates a metacognitive shift, aligning with the disposition-oriented definitions of critical thinking proposed by Ennis (1993) and Facione (1990). Quotes referencing "surprise," "realization," and "reflection" provide evidence that users began to question their prior assumptions, signaling the development of cognitive flexibility. Participants' positive responses to case-based learning underscore the importance of grounding media literacy activities in real-world contexts (Zhang et al., 2022). Several individuals provided concrete examples of behavioral changes following the training, reinforcing the notion that authentic, multimodal tasks facilitate lasting transformation. They developed cognitive and emotional habits that foster skepticism, reflection, and judgment. The data support the claim that structured, example-rich platforms can effectively translate media theory into critical practice. In summary, this research demonstrates that users did not merely learn about fake news. They cultivated cognitive and emotional habits that promote skepticism, reflection, and judgment, which are essential components of digital citizenship (Buckingham, 2007; Mihailidis, 2018; Zhang et al., 2022).

Research Question 2: To What Extent do Interactive Digital Tools, Such As the Truth-Track Evaluation Environment, Enhance Users' Critical Thinking and Decision-making When Exposed to Questionable News Content?

Participants reported that the Truth-Track tool significantly improved their decision-making and critical thinking skills, achieving average Likert scores of 4.3 and 4.1 on key items. These results were further supported by open-ended feedback, which highlighted increased reflection, a tendency to hesitate before sharing content, and enhanced confidence. One participant noted that they paused to consider the motivations behind an article, while another recognized their susceptibility to confirmation bias. Several participants emphasized how the tool's questions prompted them to engage in internal questioning. These reflections indicate that the platform did not merely present information but actively encouraged users to interrogate their own reasoning processes. This outcome aligns well with cognitive theories of reflection and correction. When participants engaged in critical self-examination, they demonstrated the type of reflective learning that Bayrak et al. (2025) associate with accuracy in assessing fake news. The structure of the tool facilitated real-time critical evaluation, consistent with the framework for media literacy interventions (Zhang et al., 2022).

Participant behavior also demonstrated strategic disengagement from harmful content, a core concept in digital literacy (Polizzi, 2025). Reports of users double-checking sources or slowing down their decision-making processes indicate that they engaged in deliberate disengagement based on their growing digital awareness. This behavior represents more than mere learning; it reflects the application of critical judgment. Their learning was not passive; it involved active monitoring and adjustment of their behavior. Formative feedback loops allow users to examine and refine their thinking, thereby reinforcing the metacognitive development outlined in educational theory (Cohen et al., 2017).

What emerges from the data is not merely a change in opinion but a transformation in the relationship between learners and information. Participants began to view themselves as evaluators rather than passive recipients. The platform served not only to inform but also to empower users to become active agents in the digital sphere. This shift supports conceptions of media literacy as a form of civic empowerment (Hobbs, 2010; Mihailidis, 2018). The patterns observed in user feedback align directly with theoretical models that prioritize engagement, reflection, and cognitive autonomy as core learning outcomes. Essentially, this research question demonstrates that critical thinking was not merely taught but also actively exercised. These patterns support learning models that emphasize engagement, reflection, and autonomy within digital environments (Bayrak et al., 2025; Cohen et al., 2017; Hobbs, 2010; Mihailidis, 2018).

Research Question 3: What are the Perceived Strengths, Limitations, and Usability Challenges of the Platform and the Truth-Track Tool as Reported by Student Teachers and Educators?

Participants responded positively to the platform's usability, scoring 4.4 for navigability and 4.2 for the usefulness of multimedia. Comments highlighted the logical layout, visual clarity, and the value of multimedia content. Participants appreciated the ability to work at their own pace and revisit material as needed. One user noted how easy it was to return to earlier sections, while others remarked on the accessibility of various media formats and how video and visual elements clarified complex ideas. These responses align with learning design principles that emphasize cognitive load reduction and universal design.

Some participants noted challenges, such as scrolling fatigue and delays on slower connections. Others expressed interest in incorporating more interactive or social components. These remarks reflect limitations identified in existing research. Concerns about limited interactivity and feedback echo warnings regarding the risk of disengagement in digital learning environments (Lobnikar et al., 2025). User concerns about the absence of peer engagement highlight the view that literacy must incorporate social and emotional dimensions (Mihailidis, 2018).

Participant responses also indicate a demand for multimodal and flexible platforms (Buckingham, 2007). The appreciation for pace control and ease of access aligns with his model of differentiated learning environments. Similarly, the emphasis on accessibility and clarity is echoed in the satisfaction users expressed regarding the platform's structure and media support (Hobbs, 2010). Participants had the freedom to choose how and when to engage with information rather than simply receiving it passively. This autonomy is central to critical literacy. The platform created a space for active, self-directed learning, and the emphasis on autonomy aids in interpreting participant reflections on pacing and navigation (Polizzi, 2025).

Users directly linked comprehension to the availability of infographics and multimedia, reinforcing the pedagogical strength of integrating such content. These features not only facilitated understanding but also aided learners in retaining and applying information. Participant responses clearly supported the notion that retention is enhanced through effective visual design (Zhang et al., 2022). The feedback from participants mirrors the theoretical consensus: effective media education tools must be clear, flexible, and socially responsive. The voices of the participants indicate that the platform largely succeeded in creating an accessible and meaningful digital learning environment. In summary, the data show that the platform's usability features were not neutral; they shaped how participants engaged with content and navigated their learning journey. This aligns closely with theories on multimodal literacy, learner autonomy, and inclusive design (Buckingham, 2007; Hobbs, 2010; Polizzi, 2025; Zhang et al., 2022).

Conclusions

The analysis of all three research questions reveals several important insights into the design and educational value of digital learning platforms aimed at enhancing critical media literacy. Participants consistently reported high levels of engagement, increased confidence, and perceived improvements in their ability to evaluate misinformation, particularly through the use of interactive elements and structured multimedia resources. These findings underscore the necessity of pedagogically sound, learner-centered digital environments in developing informed and resilient digital citizens (Hobbs, 2010; Mihailidis, 2018).

The value of interactivity and feedback-rich design is justified. Multimedia and branching scenarios are effective tools for stimulating cognitive engagement and improving knowledge retention (Zhang et al., 2022). When participants were exposed to simulations and real-world scenarios, they not only became aware of misinformation techniques but also honed their evaluative and reflective judgment (Bayrak et al., 2025; Facione, 1990). These outcomes support that experiential and inquiry-based learning promote higher-order thinking skills (Buckingham, 2007).

A second significant conclusion pertains to the role of usability and accessibility. High mean ratings for navigational clarity and instructional design confirm findings from Cohen et al. (2017), who note that intuitive design is crucial for maintaining learners' focus and motivation. When platforms align effectively with user expectations and provide autonomy in learning, they foster greater user satisfaction and sustained engagement (Polizzi, 2025; Sarmiento et al., 2025). Moreover, the emphasis on multimodal delivery proves that effective digital literacy education must accommodate diverse learning styles (Hobbs, 2010).

Thirdly, the findings reinforce the importance of fostering metacognitive awareness and digital self-regulation. Participants frequently reported greater reflection on their media consumption and a more cautious approach to sharing and evaluating online content. This development echoes Mihailidis' (2018) conception of civic media literacy, which posits that learners must cultivate both critical consciousness and ethical judgment to navigate the emotionally and ideologically complex digital landscape. Such reflective capabilities are essential for individual digital behavior and for promoting democratic resilience at the societal level (Polizzi, 2025).

However, some limitations must be acknowledged. The study relied on self-reported, post-intervention data. Additionally, participants represented a single national context, which may limit the generalizability of the findings. Future research could expand this scope with longitudinal or comparative designs.

Overall, the evidence supports that digital platforms designed with interactive, reflective, and user-responsive features can significantly strengthen learners' ability to identify and resist misinformation. These platforms support individual cognitive development and contribute to the broader educational aim of fostering digital citizenship (Hobbs, 2010; Sarmiento et al., 2025; Zhang et al., 2022). As media landscapes continue to evolve, educators must also adapt their pedagogical strategies. By grounding digital learning in robust theoretical frameworks, such as critical pedagogy, user-centered design, and experiential learning, educators and policymakers can create inclusive and adaptable tools that empower learners to thrive in complex digital environments (Bayrak et al., 2025; Buckingham, 2007).

References

Bayrak, F., Boyacioglu, I., & Yilmaz, O. (2025). Does reflection increase accuracy rather than bias in the assessments of political fake news? *Current Psychology, 44,* 5778–5795. https://doi.org/10.1007/s12144-025-07578-5

Buckingham, D. (2007). Digital media literacies: Rethinking media education in the age of the internet. *Research in Comparative and International Education*, *2*(1), 43–55. https://doi.org/10.2304/rcie.2007.2.1.43

Cohen, L., Manion, L., & Morrison, K. (2017). *Research methods in education* (8th ed.). Routledge. https://doi.org/10.4324/9781315456539

Ennis, R. H. (1993). Critical thinking assessment. *Theory into Practice*, *32*(3), 179–186. https://doi.org/10.1080/00405849309543594

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction: The Delphi report.* (ED315423). ERIC. https://eric.ed.gov/?id=ED315423

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993–1002. https://doi.org/10.1037/xge0000098

Hobbs, R. (2010). *Digital and media literacy: A plan of action* [White paper]. The Aspen Institute.

Lobnikar, B., Kotnik, N., & Krsnik Horvat, B. (2025). Fake news in the digital age: The role of science and media literacy in identifying truth. In P. Šprajc, D. Maletič, N. Petrović, I. Podbregar, A. Škraba, D. Tomić, & A. Žnidaršič

Mohorič,. (Eds.), *Human being, artificial intelligence and organization: Conference proceedings* (pp. 503–516). University of Maribor, University Press. https://doi.org/10.18690/um.fov.2.2025

Mihailidis, P. (2018). Civic media literacies: Re-imagining engagement for civic intentionality. *Learning, Media and Technology*, *43*(2), 152–164. https://doi.org/10.1080/17439884.2018.1428623

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Polizzi, G. (2025). Digital literacy and strategic (dis)engagement: Examining how functional and critical digital literacy shapes participation. *Information, Communication & Society, 1-20,* https://doi.org/10.1080/1369118X.2025.2452282

Sarmiento, A. K. B., Codjo, J., Fisher, S., & Alhayajneh, A. (2025). Impacts of digital media literacy skills on the accuracy of truth discernment. *Journal of Politics and Law*, *18*(1), 47–65. https://doi.org/10.5539/jpl.v18n1p47

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Zhang, L., Iyendo, T. O., Apuke, O. D., & Gever, C. V. (2022). Experimenting the effect of using visual multimedia intervention to inculcate social media literacy skills to tackle fake news. *Journal of Information Science*, *51*(1), 135–145. https://doi.org/10.1177/01655515221131797

Author Details

Konstantinos Karampelas
Pedagogic Department of Elementary Education
University of the Aegean
GREECE
kkarampelas@aegean.gr

Anastasia Pyrini
PARAGON-eduTech, Greece,
GREECE
pyrini@aegean.gr

Georgios Sarrigeorgiou
PARAGON-eduTech, Greece,
GREECE
gsarrigeorgiou@gmail.com

Konstantinos Tsolakidis
Pedagogic Department of Elementary Education
University of the Aegean
GREECE
tsolak@aegean.gr

Despina Sarrigeorgiou
PARAGON-eduTech, Greece,
GREECE
despinasarrigeorgiou@gmail.com

ATTITUDES OF HEAD TEACHERS TOWARD DIGITAL MATERIALS IN PRIMARY SCHOOL SCIENCE LABORATORIES IN GREECE

Konstantinos Karampelas, Nikolaos Raptis, and Maria Kouroutsidou
University of the Aegean, Rhodes
GREECE

Abstract

This study explores the attitudes of 164 elementary school head teachers toward science laboratories, focusing on the role and management of digital and conventional teaching materials. Drawing on quantitative and qualitative data from a national survey, the findings highlight a strong appreciation for both resource types while revealing significant infrastructural and pedagogical challenges. The analysis underscores the need for integrated strategies that combine effective management practices with targeted professional development. This study contributes to the growing discourse on how digital tools and inquiry-based learning can reshape primary-level science education.

Introduction

Science laboratories have become increasingly important in elementary education, offering hands-on experiences that enhance students' understanding and enthusiasm for science. This review examines the literature on the role of science labs, the importance of their effective management, and their integration of various tools. These tools, which support modern teaching practices, may be conventional (e.g., textbooks, extracurricular books, images, photographs, projectors, and scientific instruments) as well as digital (e.g., computers, digital screens, simulations, interactive whiteboards, and microscopes. As the global shift toward digital education intensifies, the strategic use of laboratories in primary schools is seen not only as a means of knowledge delivery but also as a tool for cultivating inquiry, creativity, and 21st-century skills (Polk & Santos, 2025).

Theoretical Insights in Science Laboratory Management

Science labs, which were once primarily emphasized in secondary education, are increasingly becoming recognized as an integral part of elementary schools. The importance of hands-on learning in fostering critical thinking, collaboration, and scientific literacy is widely acknowledged.

In early education, science labs provide an effective way to engage students with complex scientific concepts that may be difficult to convey through abstract teaching alone. These hands-on activities not only deepen cognitive understanding but also enhance student motivation and lead to positive learning outcomes (Kuncorowati et al., 2021). Science labs offer students the opportunity to engage in real-world scientific inquiry. They encourage exploration, observation, and experimentation, allowing students to form connections between theoretical knowledge and practical application. By providing a physical space for experimentation, students develop confidence in their ability to conduct experiments, solve problems, and generate new questions. These experiences also promote a sense of ownership over the learning process, as students actively construct knowledge rather than passively receive information. Such activities play an essential role in fostering problem-solving, critical thinking, and collaboration among students. These skills are increasingly valued in science education as they prepare students for future challenges in scientific fields (Haberbosch et al., 2025; Hofstein & Lunetta, 2004; Polk & Santos, 2025). In elementary education, these laboratory experiences lay the foundation for sustained interest in scientific inquiry and learning, making science more accessible and engaging for young learners.

The Role of the Science Laboratory in Elementary Education

Science laboratories contribute to both conceptual and procedural knowledge, which is especially important when theory is combined with practical experience. Inquiry-based learning models have proven effective in laboratory settings, allowing students to explore phenomena and develop scientific explanations. In this approach, the teacher's role shifts from that of a content provider to that of a facilitator who guides students through observation, reflection, and problem-solving (Bennett & Hogarth, 2009).

Furthermore, inquiry-based learning encourages students to engage in critical thinking and develop higher-order cognitive skills. It allows them to ask questions, conduct experiments, analyze data, and draw conclusions. This active participation in the scientific process strengthens their understanding of how science works and builds their confidence in using scientific methods. The hands-on nature of science laboratories fosters curiosity and encourages students to investigate and explore new concepts; hence, it creates a dynamic learning environment.

Beyond the cognitive benefits, science laboratories significantly impact students' attitudes toward science. Research shows that well-structured lab environments positively influence students' motivation, attitudes, and perceptions of science. For instance, early exposure to hands-on activities helps nurture long-term interest in STEM disciplines, making science engaging and rewarding for young learners.

However, these benefits are contingent upon proper preparation, well-structured classroom environments, and teacher competence in facilitating effective inquiry-based learning (Evana et al., 2021).

The Importance of Science Laboratory Management

Effective management of science laboratories is crucial for realizing their full educational potential. This includes proper planning, organizing resources, and ensuring that lab activities align with curriculum goals. Strategic planning is especially critical in elementary settings where resources are often constrained (Abas & Marasigan, 2020; Evana et al., 2021). Proper management ensures that laboratories are organized to maximize accessibility, safety, and functionality, allowing students to effectively engage in learning activities.

The physical organization of the lab is equally important. Ensuring that materials are properly labeled, stored, and easily accessible helps minimize confusion and fosters independent work. Younger learners particularly benefit from environments that promote order, routine, and self-regulation.

Effective management also involves creating a safe environment where students can confidently engage with equipment and materials. Furthermore, clear procedures and routines are necessary for ensuring the smooth operation of lab sessions, preventing disruptions, and ensuring adherence to safety protocols.

Effective implementation also requires that teachers provide clear guidance and align experiments with specific learning objectives. Teacher support, in addition to the availability of lab assistants, significantly enhances the quality of student engagement (Abas & Marasigan, 2020; Hofstein & Lunetta, 2004).

A regular evaluation of the effectiveness of a laboratory is vital for ensuring continuous improvement. By assessing student engagement and learning outcomes, teachers can refine lab activities to better meet students' needs. Ongoing professional development for teachers is also crucial for keeping them informed on best practices in laboratory management and new teaching strategies. The integration of teacher feedback, alongside student performance evaluations, ensures that the laboratory environment evolves to maintain a high standard of learning (Abas & Marasigan, 2020; Haberbosch et al., 2025; Polk & Santos, 2025).

The Role of Digital Materials in Modern Science Laboratories

The integration of digital materials into science laboratories has become increasingly important in modern science education. Tools such as simulations, virtual labs, and interactive whiteboards provide valuable opportunities for students to visualize scientific phenomena that may otherwise be difficult to observe. These materials extend the learning potential of traditional labs, and have been shown to improve student engagement and deepen understanding of scientific concepts (Kuncorowati et al., 2021).

Virtual laboratories offer a solution to infrastructural limitations by providing interactive environments in which students can conduct experiments without physical resources. This is particularly valuable in schools with limited access to traditional lab materials. Additionally, digital tools support differentiated instruction by adapting content to meet the needs of students at different learning paces. In this way, digital materials promote equity in science education, especially in under-resourced settings. However, the successful integration of digital tools depends on factors such as infrastructure, teacher training, and pedagogical readiness. Besides, the ability of teachers to utilize these tools is critical for ensuring that they effectively enhance the learning experience (Bennett & Hogarth, 2009; Evana et al., 2021).

Digital tools should complement rather than replace hands-on experimentation. Physical manipulation of materials in the laboratory is essential for developing practical skills such as spatial reasoning and fine motor skills, which cannot be fully replicated through digital simulations. Thus, an optimal science lab integrates both digital and conventional resources to create a rich, multifaceted learning environment (Evana et al., 2021).

Summary

This literature review has explored the role of science laboratories, their effective management, and the integration of digital tools into elementary education. Evidence suggests that well-managed laboratories promote conceptual and procedural knowledge, support inquiry-based learning, and enhance student motivation (Hofstein & Lunetta, 2004). Effective management practices such as planning, organizing, and evaluation of lab activities are crucial for maintaining these positive outcomes. The integration of digital tools into science instruction adds flexibility, interactivity, and equity, making science more engaging and accessible for students (Evana et al., 2021).

Given that science education evolves in digitally enriched environments, it is essential for elementary science labs to adapt accordingly. Continued investment in infrastructure, teacher training and resource alignment is necessary to ensure that science labs remain effective, dynamic spaces that foster inquiry, creativity, and a lifelong interest in science (Abas & Marasigan, 2020; Kuncorowati et al., 2021).

Methodology

Building on the theoretical framework outlined in the literature review, this study employed a descriptive survey approach to investigate the views of elementary school head teachers regarding the management of science laboratories, with a particular focus on digital materials. The methodology was designed to capture both general trends and deeper insights into how school leaders perceive lab resources, barriers, and benefits—practices commonly adopted in science education research to bridge policy and practice (Hofstein & Lunetta, 2004). The study addresses the following research questions:

- What are the attitudes of head teachers toward the use of digital materials in science laboratories?
- What barriers constrain the effective implementation of these materials?
- What benefits do head teachers associate with the presence of science laboratories in schools?

The sample consisted of 164 head teachers from a range of educational regions across urban, semi-urban and rural areas, from all 13 prefectures of Greece. Stratified purposive sampling was used to ensure representation from schools with and without science laboratories. This approach allows for robust conclusions by reflecting contextual diversity in educational settings, and involving administrative leaders in this topic is crucial (Haberbosch et al., 2025; Polk & Santos, 2025).

Data was collected through a structured questionnaire incorporating both closed- and open-ended items. Quantitative responses addressed lab infrastructure and material availability as well as challenges, while qualitative responses provided insight into perceived benefits. Quantitative data were analyzed using descriptive statistics and chi-square tests to assess differences across school types and resource categories. These techniques are appropriate for identifying statistically significant relationships in categorical datasets. All chi-square tests were conducted with an overall $\alpha = 0.05$, but to control for Type I error across our five primary comparisons we applied a Bonferroni correction, yielding an adjusted significance threshold of $\alpha = 0.01$ for each test. (Cohen et al., 2017).

Qualitative data were coded thematically to identify patterns related to innovation, engagement, visualization, and interdisciplinary teaching. This thematic approach allowed for a contextual interpretation of the quantitative trends, enriching the overall analysis (Hofstein & Lunetta, 2004). The mixed-methods strategy enabled the triangulation of data sources, which strengthens reliability and contributes to a more holistic interpretation (Kuncorowati et al., 2021).

In short, combining the mixed-methods design with statistical and thematic analysis enabled a comprehensive understanding of how head teachers conceptualize the role and management of science labs as well as the potential of digital materials within these learning environments. This approach aligns with contemporary research priorities in science education that emphasize context-sensitive inquiry, school leadership, and the evolving role of digital technologies (Kuncorowati et al., 2021). Ethical standards were applied as participation was voluntary, responses were anonymous, and informed consent was obtained (Cohen et al., 2017).

Findings

The findings from this study reveal that head teachers agree on the need for both digital and conventional materials in science laboratories. However, challenges exist regarding availability and management (Fraser et al., 2010).

Research Question 1: What are the Attitudes of Elementary School Head Teachers toward the Integration of Digital Materials in Science Laboratories?

The findings from the first research question reveal nuanced perceptions among head teachers. The data presented in Table 1 illustrate that out of the 164 responding schools, 76 possess either a dedicated science laboratory or a specially arranged room, whereas 88 schools lack such facilities.

Table 1

Availability of Science Labs among the Schools of Respondents

Availability of science labs in the schools of respondents	Number
It doesn't exist and it's not planned to happen.	59
There is no laboratory, but a classroom has been set up for this purpose.	43
It exists and is being used.	30
It doesn't exist, but it is planned to happen.	29
It exists but is not utilized.	2
Instruments that we have in abundance for experiments in the classroom are used.	1
TOTAL	164

Table 2 shows no significant difference in availability between conventional and digital resources, as shown by a chi-square analysis ($\chi^2 = 0.00$, $p = 1.0$). The absence of universally accessible essential laboratory materials highlights persistent infrastructure limitations. This observation aligns with literature emphasizing resource scarcity as a critical barrier impacting the quality of science education (Haberbosch et al., 2025; Hofstein & Lunetta, 2004; Polk & Santos, 2025).

As shown in Table 3, a chi-square test comparing head teachers' valuations of conventional versus digital materials produced $\chi^2(2) = 7.84$, $p = 0.049$. Although $p < 0.05$ under a conventional threshold, it does not meet our Bonferroni-adjusted criterion of $\alpha = 0.01$ and is therefore reported here as not statistically significant. This conservative approach helps guard against false positives given multiple parallel tests. Head teachers implicitly associate digital tools with practices, such as interactive demonstrations and simulations, that conventional materials cannot fully provide. These findings reinforce the increasing advocacy for integrating conventional and digital resources into science education as complementary tools. However, the successful integration of both digital and conventional resource types depends significantly on adequate teacher training, administrative support, and institutional readiness, enabling educators to maximize pedagogical effectiveness within diverse classroom settings (Evana et al., 2021; Fraser et al., 2010; Hofstein & Lunetta, 2004; Kuncorowati et al., 2021).

Table 2

Availability in the Schools with Science Labs or Appropriately Designed Rooms

Material Type	Available
School books	29
Organs & instruments	29
Extracurricular books	28
Images and photographs	27
Computers with internet access	26
Projectors	25
Interactive whiteboards	25
Projection screens	23
Computers without internet access	15
Microscopes	11

Table 3

Attitudes toward the Value of Both Material Categories of Head Teachers in Schools without Science Labs or Appropriately Designed Classrooms

Material	Absolutely Valuable	Not Necessarily Valuable	Never Valuable
School books	52	25	8
Extracurricular books	56	26	5
Organs & instruments	80	4	4
Images and photographs	50	26	8
Projectors	59	18	9
Computers with internet access	71	10	4
Computers without internet access	45	21	20
Interactive whiteboards	72	11	5
Projection screens	61	17	6
Microscopes	50	20	17

Research Question 2: What Challenges and Barriers do Head Teachers Perceive Regarding the Use of Digital Tools in Science Labs?

Regarding the second research question, Table 4 highlights the primary barriers reported by head teachers: financial limitations, insufficient teacher training, and inadequate technical support. Statistical analysis using chi-square ($\chi^2 = 4.39$, $p = 0.82$) showed no significant difference in barrier frequencies between schools with and without laboratory facilities, indicating that these challenges are systemic across schools.

Financial constraints hinder procurement and maintenance, while gaps in teacher training and technical support impede the effective use of laboratory resources. These findings support previous research emphasizing that providing resources alone is insufficient without adequate training, leadership, and support (Evana et al., 2021; Kuncorowati et al., 2021).

Table 4

Frequency of Barriers Mentioned by Head Teachers

Barrier	Frequency	Schools with Lab	Schools without Lab
Lack of financial resources	129	58	71
Lack of required training	59	29	30
Lack of time for management	56	25	31
Lack of educational staff	37	19	18
Lack of teaching staff	34	17	17
Absence of relevant legislation	31	15	16
Lack of trained staff	29	16	13
There are no challenges at all.	9	7	2
Lack of trained personnel	6	3	3
Lack of teaching staff.	3	2	1
Other	12	0	0

The open-ended responses further highlighted specific operational constraints. For instance, one head teacher noted, "We struggle to integrate innovative tools due to a rigid curriculum timetable," while another cited, "Digital tools remain unused due to a lack of IT technicians." These comments underline how curriculum constraints and a shortage of technical personnel limit innovation and operational readiness. These findings emphasize the need for holistic strategies to support laboratory-based science education, including professional development, financial investment, technical support, and flexible administrative policies. School leadership must play an active role in securing resources and fostering environments that enable effective use of lab facilities (Abas & Marasigan, 2020; Bennett & Hogarth, 2009; Fraser et al., 2010; Hofstein, 2004; Kuncorowati et al., 2021).

Research Question 3: Head Teachers' Evaluation of the Usefulness and Impact of Digital Materials on Science Teaching and Learning

Regarding the third research question, Table 5 categorizes the responses of head teachers, with the majority associating science labs with enhancing science teaching and providing hands-on experimentation opportunities. Several respondents also noted broader educational benefits such as interdisciplinary learning and increased student engagement. These findings align with the literature on the pedagogical value of inquiry-based, hands-on instruction, which fosters deeper learning (Fraser et al., 2010; Haberbosch et al., 2025; Hofstein & Lunetta, 2004; Polk & Santos, 2025).

Whereas digital materials were not explicitly mentioned by most respondents, categories such as "innovation" and "visual support" suggested a readiness for integrating digital tools into science laboratories. Digital tools can enhance science laboratories by providing access to complex scientific phenomena, thus supporting better understanding (Kuncorowati et al., 2021).

Table 5

Benefits of the Science Lab

Responses	Frequency
Better teaching of science	155
Mainly by conducting experimental activities	148
Opportunity for innovative actions in natural sciences	58
Better teaching of other subjects	47
Opportunity for innovative actions in other subjects	31
Opportunity for innovative actions in natural sciences	6
Other	12

Responses in the "Other" category, such as "students work better in teams when engaged in inquiry" and "it allows interaction with external experts," reflect a broader view of science labs as spaces that foster collaboration, engagement, and real-world connections. These views underscore the role of science labs in enhancing pedagogy and social interaction. Overall, head teachers consider science laboratories not only as places for practical experiments but also as spaces for digitally supported innovation and learning, supporting the need for a flexible design model that integrates both digital and conventional resources (Abas & Marasigan, 2020; Hofstein, 2004).

Conclusions

This study explores the attitudes of elementary school head teachers toward science laboratories, focusing on digital materials and their educational benefits, challenges, and perceived value. The study findings support the theory of an integrative science lab design that combines conventional and digital tools to create dynamic, student-centered learning environments. This highlights a shift toward inquiry-based teaching (Polk & Santos, 2025).

Three key patterns emerged across the research questions. First, head teachers consider conventional and digital materials equally important. This is evident in both their preference data and open-ended responses, which connect digital tools to innovation, visualization, and student engagement. Despite infrastructure challenges, digital resources are seen as essential (Haberbosch et al., 2025; Polk & Santos, 2025). Second, systemic barriers such as insufficient funding, lack of professional development, and inadequate technical support affect both schools with and without labs. The chi-square analysis in this study revealed no significant difference in the types of constraints reported by these schools, emphasizing that challenges are systemic. These results align with previous research, which highlights the need for strategic planning and investment in human resources to implement sustainable science laboratories (Fraser et al., 2010). Third, head teachers consider science labs not only as spaces for experiments but also as pedagogical and community resources. The labs foster teamwork, public engagement, and interdisciplinary learning. This reinforces the literature that positions science labs as transformative spaces that are capable of driving holistic educational development (Abas & Marasigan, 2020; Hofstein, 2004; Kuncorowati et al., 2021).

The study suggests a theoretical model that advocates for the hybridity of digital and conventional resources in science laboratories. This model calls for both types of materials to support instruction, inquiry-based learning, and long-term scientific literacy (Haberbosch et al., 2025; Polk & Santos, 2025). Future research could explore regional differences and incorporate perspectives of science teachers and students (Cohen et al., 2017).

References

Abas, M. C., & Marasigan, A. C. (2020). Readiness of science laboratory facilities of the public junior high in Lanao del Sur, Philippines. *IOER International Multidisciplinary Research Journal*, *2*(2), 12–20.

Bennett, J., & Hogarth, S. (2009). Would you want to talk to a scientist at a party? High school students' attitudes to school science and to science. *International Journal of Science Education*, *31*(14), 1975–1998. https://doi.org/10.1080/09500690802425581

Cohen, L., Manion, L., & Morrison, K. (2017). *Research methods in education* (8th Ed). New York: Routledge. https://doi.org/10.4324/9781315456539

Evana, Y., Fitria, H., & Fitriani, Y. (2021). The management of science laboratory at senior high school in the digital era. *JPGI (Jurnal Penelitian Guru Indonesia)*, *6*(3), 660-667. https://doi.org/10.29210/021105jpgi0005

Fraser, B. J., Aldridge, J. M., & Adolphe, F. S. G. (2010). A cross-national study of secondary science classroom environments in Australia and Indonesia. *Research in Science Education*, *40*(4), 551–571. https://doi.org/10.1007/s11165-009-9133-1

Haberbosch, M., Deiters, M., & Schaal, S. (2025). Combining virtual and hands-on lab work in a blended learning approach on molecular biology methods and lab safety for lower secondary education students. *Education Sciences*, *15*(2), 123. https://doi.org/10.3390/educsci15020123

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, *88*(1), 28–54. https://doi.org/10.1002/sce.10106

Kuncorowati, W., Habibi, M. W., Mosrifa, C. T., & Ilafi, M. M. (2021). The effectiveness of utilization of the Science Laboratory in Integrated Science Learning at MTs Unggulan Al Qodiri 1 Jember. *INSECTA: Integrative Science Education and Teaching Activity Journal*, *2*(2), 165–174. https://doi.org/10.21154/insecta.v2i2.3287

Polk, M. E., & Santos, D. L. (2025). Science practices in action: Group engagement with different degrees of inquiry in general chemistry laboratory. *Journal of Chemical Education*, *102*(4), 1380–1388. https://doi.org/10.1021/acs.jchemed.4c01314

Author Details

Konstantinos Karampelas
Pedagogic Department of Elementary Education
University of the Aegean
GREECE
kkarampelas@aegean.gr

Nikolaos Raptis
Department of Pre-school Education and Education Design
University of the Aegean,
GREECE
nraptis@aegean.gr

Maria Kouroutsidou
Department of Pre-school Education and Education Design
University of the Aegean,
GREECE
kouroutsidou@aegean.gr

EVALUATING GPT-4's PROFICIENCY ON NORWEGIAN EXAMS AND TESTS—AND EXPLORING THE BROADER IMPLICATIONS FOR EDUCATIONAL PRACTICE

Rune Krumsvik and Lise Jones
University of Bergen
NORWAY

Abstract

Meta-analyses show Intelligent Tutoring Systems excel one-to-one, yet none support Norwegian. This position paper treats GPT-4—though not designed as an ITS—as a candidate tutor, testing it on Norwegian exams in medicine, nursing, psychology, dentistry, military theory, driving, university entrance, citizenship and maths teaching, plus IQ, social and multimodal medical-image tasks. GPT-4 averaged 94.3% accuracy, handled descriptive-procedural questions, adapted to Sami, and surpassed conventional ITSs in linguistic and cultural flexibility. Findings reveal broad multilinguistic, cognitive and multimodal strengths with significant implications for formative and summative assessment across education.

Keywords: AI; GPT-4; Multilingualism; Exams, Tests, Norwegian; Performance Introduction

Introduction

With the launch of language models such as XLNet, BERT, ChatGPT, GPT-4, Gemini Advanced, Claude, we are facing a technological paradigm shift that may also influence how we perceive Intelligent Tutoring Systems (ITS) and related areas in the future. Since the 1970s, research has explored ITS and the potential of AI to provide personalized tutoring, inspired by Bloom's (1984) well-known 2-sigma finding. Numerous meta-analyses comparing traditional teaching methods with ITS have found that, under certain conditions, ITS can effectively provide one-on-one tutoring. However, existing ITSs do not support the Norwegian language. While large language models like GPT-4, Gemini Advanced, and Claude are not specifically designed as ITS, they share significant similarities. The knowledge base shows that these models have the potential to address certain educational challenges in both the education and healthcare sectors, opening up a broader discussion about their role in the future of education. GPT-4, an advanced language model developed by OpenAI, has proven capable in various English-speaking academic fields, exams, and tests (Ray, 2023; Agarwal et al., 2023; Brin et al., 2023; Brodeur et al., 2023; Deng et al., 2025; Goh et al., 2024; Hirunyasiri et al., 2023;

Jin et al., 2024; Karthikesalingam and Natarajan, 2024; Kim et al. 2020; Liu et al, 2024; McDuff et al., 2025; Nori et al., 2023; Rajpurkar et al., 2020; Phung et al., 2023). However, the current state of knowledge lacks studies on how it performs in Norwegian in Norwegian-language exam and test contexts. This position paper, based on a case study, aims to evaluate how GPT-4 manages multilingual challenges with Norwegian as an exam/test language, focusing on descriptive-procedural questions and various exam and test contexts both within and outside academia. These contexts include exams in medicine, nursing, psychology, dentistry, a military theory test, the Norwegian driving test, the Swedish university entrance exam (SweSAT), the Norwegian citizenship test, and a national teacher exam in mathematics. A primary objective is to assess the reliability and generalizability of this type of AI in academic settings and in the Norwegian language. More specifically it focuses on whether GPT-4 is capable of answering various exams and tests that are primarily given in Norwegian in Norwegian educational and societal contexts, how reliable it is, and what implications this might have for both multilingualism summative and formative assessment elements within and outside academia.

This abovementioned literature review and our former studies (Krumsvik, 2024, 2025a, 2025b, 2025c) generated a number of explorative questions and reflections around this topic: How effectively can GPT-4 handle multilingual challenges, particularly in Norwegian, across both academic exam tasks and general IQ and social tests? What is its precision rate, and how reliable is its performance in these contexts? Does GPT-4's ability to manage descriptive and procedural questions align with international findings, and how well does it adapt to Norwegian exam and test contexts? Can it demonstrate linguistic and cultural adaptability that conventional ITSs lack? Furthermore, what are the limitations of GPT-4 in these contexts, and how does a case study contribute to our understanding of its multilingual and cognitive capabilities in summative assessments? Can GPT-4 serve as an effective tool in formative assessment contexts, and how well does the case study design perform in research environments characterized by rapid development? These preliminary questions and reflections can be summarized in one main research question we will examine in this position paper:

How capable is GPT-4 of answering selected Norwegian exams and tests, and what potential implications might this have for formative and summative assessment in educational sciences and healthcare education?

Methodology

This position paper is based on a case study which is exploratory and intrinsic (Stake, 1995, 2006). I conducted a cumulative data collection and analysis process (Creswell & Guetterman, 2021), based on performance on nine exams and tests inside and outside academia in Norway where I applied chain-of-thought prompting with the exact same wording as in the exam and test text in Norwegian (Figure 1).

Figure 1

The Research Process of the Intrinsic Case Study

Chain-of-thought prompting is an approach in which a user explicitly asks a language model to reveal its step-by-step reasoning process before giving the final answer, improving transparency and often boosting solution quality for complex tasks. Furthermore, in the supplementary data collection (blue arrows in second and third position in the figure), I integrated the research questions into the dialogue (A) of the results from 1, and interacted with GPT-4 around the preliminary findings. Finally, further digital fieldwork (B) was conducted to check for possible biases and misinterpretations, ongoing fine-tuning of GPT-4, as well as in light of the aforementioned current state of knowledge on this topic.

The main test period was carried out from March 25, 2023, to August 5, 2023, and the exams and tests were from different areas both inside and outside academia to check GPT-4s ability to handle different contexts. Four of the exams were fullscale exams, while the five other exams and tests were based on random selection (two sub-tasks in one test had to be omitted due to task drawings that GPT-4 could not perceive and "see"). All the exams and tests were in the Norwegian language (except the Swedish SweSAT) and consisted mainly of text questions. Scoring of all the exams was based on the grading guidelines (sensorveiledning) derived from different sources. Interaction with GPT-4 was conducted based on the questions in nine exams and tests, which were posed to GPT-4 using chain-of-thought prompting, and responses were recorded (each response was considered final).

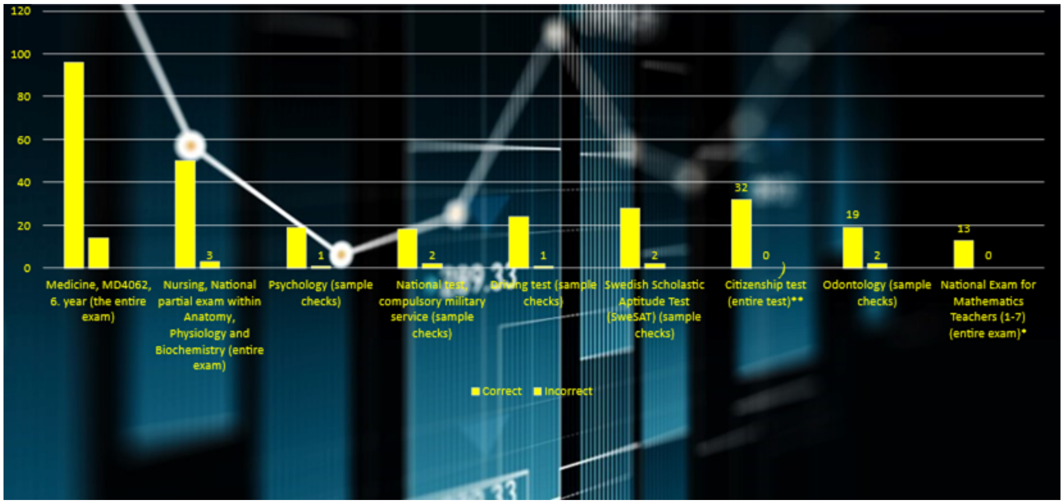
Data analysis, step 1, was based on GPT-4's performance on the nine exams and tests selected randomly from previous exam sets. The supplemental data was collected from August 2023 to April 2024 and consisted of comprehensive interactions with GPT-4 and digital fieldwork (described above).

Results

The tests were conducted from March 20 to August 10, 2023. Figure 2 illustrates the number of questions in each of the nine exams and tests.

Figure 2

Number of Questions in Each of the Nine Exams and Tests



Note. Sample checks: When only sample checks of the exam/test were performed. Entire exam: When a test of the entire exam/test was conducted.

* Two sub-tasks had to be omitted due to a task drawing that GPT-4 cannot see (thus, 13 out of 15 sub-tasks were completed).

** This test currently consists of 36 questions, but the version publicly available and tested consisted of 32 questions.

Table 1 shows that the average precision rate of 94.26% indicates that GPT-4 performs very well across various fields within and outside academia, spanning a relatively broad range of exams, tasks, tests, and domains. It demonstrates multilingual and cognitive skills a high level and GPT-4 generally has capabilities comparable to the human level in such exam and test contexts. While all nine exams/tests have an element of descriptive knowledge (knowing that), the medical exam includes a number of exam tasks that lean towards procedural knowledge (knowing how) (Anderson, 2005) as they are formulated as patient cases (and not factual knowledge per se). Additionally, about 10 percent of the 110 exam tasks contain image illustrations related to the tasks (X-rays, images of skin rashes, organ images, etc.), which are helpful for students in addition to the task text itself (which often small case descriptions about patients). This multimodality could not be "seen" or interpreted by GPT-4 in spring 2023, and thus, for these tasks, it could only respond based on text descriptions. Nevertheless, we see that GPT-4 achieves 87.3% correct answers (96 out of 110) on this exam, and when looking at the detailed and reasoned responses it provides, this shows a good academic level. Below I present a dialogue (A) with GPT-4 regarding these results.

Table 1

GPT-4's Performance on Exams and Tests inside and outside Academia in Norway

Field	Correct (%)	Incorrect (%)
Medicine (entire exam)	87.3	12.7
Nursing (entire exam)	96.2	3.8
Psychology (sample checks)	95	5
Military Conscription (IQ-test) (sample checks)	90	10
Driving Test (Car) (sample checks)	96	4
Swedish Scholastic Aptitude Test (SweSAT) (sample checks)	93.3	6.7
Citizenship Test (entire test)	100	0
Dentistry (sample checks)	90.5	9.5
Teacher education (Mathematics) (entire exam)	100	0
Average Precision Rate	**94.26**	

Summary of Phases A and B

Overall, GPT-4 shows advanced capabilities in understanding and generating accurate responses across diverse and complex tasks, particularly in Norwegian-language exams and medical image analysis.

The results indicate that in the medical exam, descriptive and procedural knowledge are in a dialectical relationship as GPT-4 cannot answer the exam questions without possessing both types of surface and deep knowledge. Exams in nursing, psychology, dentistry, the Swedish Scholastic Aptitude Test (SweSAT), and teacher education also exhibit this combination to some extent. Other tests, like the military conscription test (IQtest), also include this combination but are more oriented towards general knowledge. It can be added that a smaller version than GPT-4, GPT-3, managed 73 out of 80 tasks in the SweSAT in 2022 (Svensson, 2022), and many were surprised at how well it handled abstract metaphors. The driving test and citizenship test primarily assess descriptive knowledge. From this, we can see that the GPT-4's descriptive and procedural abilities can also be related to Anna Sfard's (1998) two metaphors for learning (acquisition metaphor and participation metaphor) but in a more situated context within school or academia, not limited to a specific exam or test situation inside and outside academia.

Overall, GPT-4s scores across the nine different exams/tests demonstrate its ability to handle multilinguist and relatively complex Norwegian-language questions, at times at a high academic level. Additionally, phases A and B show that it also handles multimodal image analysis very well. This suggests a need for a broader epistemological discussion about new forms of AI-generated communities of practice (CoP and whether GPT-4 can be considered a highly capable dialogue partner and tutor for Norwegian-speaking students preparing for this form of summative assessment (school exams). These findings are consistent with our tentative knowledge summaries and case studies (Krumsvik, 2024, 2025a, 2025b, 2025c), which find a similar trend across various English-language exams/tests internationally (Ray, 2023).

Discussion

The results from testing GPT-4 on various Norwegian-language exams and tests on multilingual and cognitive capabilities in such contexts, aligns with our previous findings (Krumsvik, 2024, 2025a, 2025b, 2025c). GPT-4s cognitive capabilities also align with the pre-print from Bubeck et al. (2023), Ray (2023) and Deng et al. (2025). Particularly notable in this study is GPT-4's multilinguistic abilities and performance on Norwegian-language exams, despite the model primarily being trained on English-language data. This indicates a good ability to generalize knowledge across languages. Exams such as medicine, nursing, and psychology include both descriptive and procedural knowledge, requiring deeper understanding and processing. GPT-4's ability to handle such complex tasks suggests that the model can go beyond mere memorization of facts and engage in more sophisticated cognitive processing. Such findings are supported within ITS by VanLehn (2011), who emphasizes the importance of deep learning in effective tutoring systems. VanLehn (2011) points out that human tutoring has an effect size of $d = 0.79$, while ITSs show a similar effect size of $d = 0.76$ in his study, and in Tlili et al. (2025) metanalysis this is $g=1.07$.

The results of our study suggest that GPT-4, as part of an ITS, can offer a comparable level of support as human tutors. This is especially relevant in light of previous research showing that traditional classroom instruction often does not reach the same level of effectiveness as one-on-one tutoring (Bloom, 1984). At the same time, it is important to note that, according to Ma et al. (2014), there are some distinctive features of ITS (as mentioned earlier) that GPT-4 does not inherently possess. These are especially oriented towards calculating inferences from student responses, constructing multidimensional models of the student's learning status, and placing the student's current learning status in a multidimensional domain model. This can be partially achieved by establishing a domain-specific "chatbot within the chatbot" by integrating a training basis on top of GPT-4 and simultaneously embedding a script in this chatbot, tuning it more specifically towards having an ITS-related functionality (along with established ITSs). Integrating GPT-4 in ITS-related areas can potentially expand tutoring opportunities in the educational sector. With GPT-4's ability to generate educational content, analyze student input, and offer real-time feedback, GPT-4 can significantly enhance tutoring opportunities and AI-CoP both for students who have Norwegian as their native language, but also for foreign students who may interact with GPT-4 in English about Norwegian-language exam questions, tests, etc. This can be a good sparring partner for such language thresholds and be a valuable supplement for foreign students in addition to other measures and conventional tutors in higher education.

Conclusion and Implications

The research question in position paper focused on how capable GPT-4 is of answering exams and tests in Norwegian and what implications could this have for education. Despite some limitations, the position paper confirms that GPT-4 has significant multilinguistic and cognitive capabilities, making it a valuable tool both inside and outside academia as a sparring partner in various exam and test contexts. With an average precision rate of 94.26%, the model demonstrates the ability to answer both more factual questions and more complex and varied questions in a manner comparable to human performance in such contexts. Given that GPT-4 also masters Norwegian well, it is particularly relevant in the nine exam and test contexts studied, which are primarily in Norwegian, targeting the Norwegian educational and societal context (the exception being the SweSAT, where GPT-4 also shows strong proficiency in Swedish). GPT-4's good performance in Norwegian as early as 2023 is probably because, e.g. medical knowledge is globally standardized and largely overlaps with the English-language material the model was trained on. It handles written Norwegian and technical terminology well, and many exam tasks require pattern recognition rather than deep reasoning—an area where GPT-4 excels. The exams were highly standardized and not dependent on Norwegian-specific legal or cultural context, allowing the model to apply its global knowledge base effectively in Norwegian. This suggests that large language models can play a complementary role in various tutoring contexts in higher education, in developing AI-CoPs, and in the further development of ITS.

In summary, such language models are intellectual artifacts mastering contextual language games (Wittgenstein, 1997), representing a significant leap from earlier language models, ITS, and chatbots. But it requires mastery of the granularity in prompts based on "chain of thought" prompting, which often shows variability among students, citizens, and learners. This illustrates that such digital competence will become increasingly important both inside and outside academia in the coming years to fully exploit the potential of such language models for both formative and summative assessment contexts in the future. With such reservations, one implications of the findings might be that non-Norwegian-speaking students, university staff, and citizens in general who wish to learn Norwegian now have a highly competent "Norwegian teacher" by their side with GPT-4. However, AI is at the same time an ethical minefield and the study also underscores the need for vigilance and careful implementation to mitigate biases and ethical issues associated with AI use in education.

Methodological Considerations – Strengths and Weaknesses

As previously mentioned, the case study underpinning this position paper is based on pre-testing of GPT-4 in the absence of actual student participation. As such, it carries several limitations, yet also demonstrates certain strengths. Given that the AI field is a "moving target" developing very rapidly, case studies with triangulation, cumulativeness, and the possibility for retesting over a year can be an effective design in such research settings. The ability to track progress over an entire year provides valuable insights into the model's ability to adapt and improve over time. Although this case study has several strengths, there are also important limitations that need to be considered related to the methodological choices made. For instance, a full-scale testing of all nine exams and test areas was not conducted. Instead, a series of sample checks were carried out in this case study in five of the test areas, and such sample checks have several limitations that should be noted. The tasks were not translated to English and kept in Norwegian. Additionally, although the study followed the development over one year, this may still be a relatively short period to fully understand the long-term effects and improvements in AI models. Longer follow-up periods could provide more comprehensive insights. The results from the case study cannot be generalized and derive their strength from the depth perspective. The selected sample also has its biases. These limitations highlight the need for caution in interpreting the results and the importance of further research to validate and extend the findings.

Declaration of AI Use

This article explores the use of GPT-4, and artificial intelligence (AI) is therefore the object of study in this case-based research. As such, GPT-4 has been used in various ways throughout the research process, including pre-testing, documentation, and analytical reflection. However, all parts of the article—including the structure, argumentation, interpretation of findings, and final wording—have been designed, authored, and critically reviewed by the authors themselves.

References

Anderson, J. R. (2005). *Cognitive Psychology and Its Implications* (6th ed.). Worth Publishers.

Agarwal, N., Moehring, A., Rajpurkar, P., & Salz, T. (2023). *Combining human expertise with artificial intelligence: Experimental evidence from radiology (Working Paper No. 31422).* National Bureau of Economic Research. https://www.nber.org/papers/w31422

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13*(6), 4-16. https://doi.org/10.3102/0013189X013006004

Brin, D., Sorin, V., Vaid, A., Soroush, A., Glicksberg, B. S., Charney, A. W., Nadkarni, G., & Klang, E. (2023). Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. *Scientific Reports*, *13*, 16492. https://doi.org/10.1038/s41598-023-43436-9

Brodeur, P. G., Buckley, T. A., Kanjee, Z., Goh, E., Ling, E. B., Jain, P., Cabral, S., Abdulnour, R.-E., Haimovich, A. D., Freed, J. A., Olson, A., Morgan, D. J., Hom, J., Gallo, R., McCoy, L. G., Horvitz, E., Mombini, H., Lucas, C., Fotoohi, M.,…Rodman, A. (2024). Superhuman performance of a large language model on the reasoning tasks of a physician. *ArXiv,* 2412.10849. *https://arxiv.org/abs/2412.10849*

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. *ArXiv,* 2303.12712. https://doi.org/10.48550/arXiv.2303.12712

Creswell, J. W., & Guetterman, T. C. (2021). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (6th ed.). Pearson.

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, *227*, 105224. https://doi.org/10.1016/j.compedu.2024.105224

Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., Cool, J. A., Kanjee, Z., Parsons,A. S., Ahuja, N., Horvitz, E., Yang, D., Milstein, A., Olson, A. P. J., Rodman, A., & Chen, J. H. (2024). Large language model influence on diagnostic reasoning: A randomized clinical trial. *Journal of the American Medical Association (JAMA) Network Open, 7*(10*),* Article e2440969. doi:10.1001/jamanetworkopen.2024.40969

Hirunyasiri, D., Thomas, D. R., Lin, J., Koedinger, K. R., & Aleven, V. (2023, July 7). Comparative analysis of GPT-4 and human graders in evaluating praise given to students in synthetic dialogues [Conference paper]. In *Proceedings of the AIED 2023 Workshop: Towards the Future of AI-Augmented Human Tutoring in Math Learning* (CEUR Workshop Proceedings, Vol. 3491, pp. 1–12). CEUR-WS.org. https://ceur-ws.org/Vol-3491/paper4.pdf

Jin, H. K., Lee, H. . & Kim, E. (2024). Performance of ChatGPT-3.5 and GPT-4 in national licensing examinations for medicine, pharmacy, dentistry, and nursing: A systematic review and meta-analysis. *BMC Med Educ 24*, 1013. https://doi.org/10.1186/s12909-024-05944-8

Karthikesalingam, A. & Natarajan, V. (2024, January 12). *AMIE: A research AI system for diagnostic medical reasoning and conversations*. Google Research. https://blog.research.google/2024/01/amie-research-ai-system-for-diagnostic_12.html

Kim, H.-E., Kim, H. H., Han, B.-K., Kim, K. H., Han, K., Nam, H., Lee, E. H., & Kim, E.-K. (2020). Changes in cancer detection and false-positive recall in mammography using artificial intelligence: A retrospective, multireader study. *The Lancet Digital Health*, 2(3), e138–e148. https://doi.org/10.1016/S2589-7500(20)30003-0

Krumsvik, R. J. (2024). Artificial intelligence in nurse education – a new sparring partner? GPT-4 capabilities of formative and summative assessment in National Examination in Anatomy, Physiology, and Biochemistry. *Nordic Journal of Digital Literacy*, *19*(3), 172–186. https://doi.org/10.18261/njdl.19.3.5

Krumsvik, R. J. (2025a). GPT-4's capabilities in handling essay-based exams in Norwegian: An intrinsic case study from the early phase of intervention. *Frontiers in Education, 10,* 1444544. https://doi.org/10.3389/feduc.2025.1444544

Krumsvik, R. J. (2025b). Chatbots and academic writing for doctoral students. *Education and Information Technologies, 30*, 9427–9461. https://doi.org/10.1007/s10639-024-13177-x

Krumsvik, R. J. (2025c). GPT-4's capabilities for formative and summative assessments in Norwegian medicine exams—an intrinsic case study in the early phase of intervention. *Frontiers in Medicine, 12*, 1441747. https://doi.org/10.3389/fmed.2025.1441747

Liu M., Okuhara, T., Chang, X., Shirabe, R., Nishiie, Y., Okada, H., & Kiuchi, T. (2024). Performance of ChatGPT across different versions in medical licensing examinations worldwide: Systematic review and meta-analysis. *J Journal of Medical Internet Research, 26,* e60807. doi: 10.2196/60807.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901-918. edu-a0037123.pdf

McDuff, D., Schaekermann, M., Tu, T., Palepu, A., Wang, A., Garrison, J., Singhal, K., Sharma, Y., Azizi, S., Kulkarni, K., Hou, L., Cheng, Y., Liu, Y., Mahdavi, S. S., Prakash, S., Pathak, A., Semturs, C., Patel, S., Webster, D. R.,… Natarajan, V. (2025). Towards accurate differential diagnosis with large language models. *Nature, 642*, 451–457. https://doi.org/10.1038/s41586-025-08869-4

Svensson, E. (2022, June 27). Does GPT-3 know Swedish? *Medium*. https://medium.com/@emil_svensson/does-gpt-3-know-swedish-3e22bfd7b58f

Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of GPT-4 on medical challenge problems. *ArXiv*, 2303.13375. https://doi.org/10.48550/arXiv.2303.13375

Phung, T., Pădurean, V.-A., Cambronero, J., Gulwani, S., Kohn, T., Majumdar, R., Singla, A., & Soares, G. (2023). Generative AI for programming education: Benchmarking ChatGPT, GPT-4, and human tutors. In *Proceedings of the 2023 ACM Conference on International Computing Education Research - Volume* 2 (pp. 41-42). https://doi.org/10.1145/3568812.3603476

Rajpurkar, P., O'Connell, C., Schechter, A., Asnani, N., Lee, J., Kiani, A., Ball, R. L., Mendelson, M., Maartens, G., van Hoving, D. J., Griesel, R., Ng, A. Y., Boyles, T. H., & Lungren, M. P. (2020). CheXaid: deep learning assistance for physician diagnosis of tuberculosis using chest x-rays in patients with HIV. *NPJ Digital Medicine, 3*, 115. https://doi.org/10.1038/s41746-020-00322-2

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, *27*(2) 4-13. https://doi.org/10.3102/0013189X027002004

Stake, R. E. (1995). *The art of case study research*. Sage.

Stake, R. E. (2006). *Multi case study analy*sis. Guilford Press.

Tlili, A., Saqer, K., Salha, S., & Huang, R. (2025). Investigating the effect of artificial intelligence in education (AIEd) on learning achievement: A meta-analysis and research synthesis. *Information Development*. https://doi.org/10.1177/02666669241304407

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197-221. https://doi.org/10.1080/00461520.2011.611369

Wittgenstein, L. (1997). *Filosofiske undersøkelser* (M. B. Tin, Trans.). Pax.

Author Details

Rune Krumsvik
University of Bergen
NORWAY
Rune.Johan.Krumsvik@uib.no

Lise Jones
University of Bergen
NORWAY

EVALUATING CHATGPT'S EFFECTIVENESS IN FORMATIVE ASSESSMENT PRACTICES FOR CS EDUCATION: A SWOT ANALYSIS

Jacqui Chetty
University of Birmingham
United Kingdom

Abstract

Recently, large language models (LLMs) such as OpenAI's ChatGPT and AI-powered IDEs like GitHub Copilot have become increasingly integrated into teaching and learning. This study evaluates the effectiveness of ChatGPT for formative assessment through a Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis. The SWOT analysis serves as an evaluative tool to critically assess the effectiveness of ChatGPT in assessment practices. The findings suggest that ChatGPT is effective as an automated tutor; supports students in engaging with programming concepts; provides a real-time environment for students to rapidly assess and reflect on their code; and enhances their learning experience. However, caution is necessary, as ChatGPT may produce inaccuracies, lack deep conceptual awareness, and pose risks to academic integrity.

Introduction

In late 2022 OpenAI launched ChatGPT, seen as a breakthrough Large Language Model (LLM) that could generate text and maintain human-like conversations (Rahman & Watanobe, 2023). ChatGPT and similar LLMs have the potential to create opportunities, present challenges, pose threats, raise ethical concerns and disrupt education fields. For example, for university students, ChatGPT can assist in research and writing tasks and develop critical thinking and problem solving (Kasneci et al., 2023). It can facilitate group and remote learning and empower learners with disabilities by combining speak-to-text or text-to-speech solutions. For educators, ChatGPT can assist with personalising student learning, lesson planning, research, and writing as well as assessment and evaluation. However, the negative impacts of easy cheating and plagiarism, ChatGPT solving problems instead of students acquiring a skill set, and providing incorrect knowledge (Malinka et al., 2023), can outweigh the advantages.

Although a very new research field, many researchers are investigating the use of LLMs in education. The primary focus is on improving the learning process and aiding students as well as re-designing repetitive processes for educators. For example, Qureshi (2023) explored the use of ChatGPT as a tool for learning and assessment in undergraduate computer science, highlighting the opportunities and challenges. This would be no different for higher education as ChatGPT is seen as a potential disruptor to teaching-and-learning. Banerjee et al. (2025) conducted an impact analysis by evaluating the capability of ChatGPT for instructional purposes in the field of computer science and engineering. The article explores the opportunities and limitations of ChatGPT as well as performing a student survey to highlight ChatGPT anomalies and concerns. Tlili et al. (2023) discuss whether ChatGPT is friend or foe, including the extent to which ChatGPT has additional qualities, such as personality, emotion, its usefulness, ethical considerations, cheating, and truthfulness. The article highlights that to incorporate ChatGPT into instructional design, the way knowledge is assessed must change.

Much of the research focus thus far is aimed at strengths, weaknesses and threats. Accordingly, conducting a SWOT analysis (Strengths, Weaknesses, Opportunities, and Threats) is a logical course of action to evaluate ChatGPT in large-scale assessment practices. The SWOT would highlight ChatGPT's potential benefits, limitations, and future implications for educational settings. The SWOT analysis is framed as an evaluative framework to critically examine ChatGPT's effectiveness in assessment practices.

The aim of the paper is to address the following research questions:

> **RQ1:** What are the strengths and weaknesses of ChatGPT as an educational tool for formative assessment in CS education?
> **RQ2:** What external opportunities and threats influence the effectiveness of ChatGPT in formative assessment for CS education?

To answer these research questions a SWOT Analysis was conducted.

Background Motivation

Existing Assessment Practices

Covid-19 changed teaching-and-learning and assessment practices dramatically world-wide. Overnight, alternatives to traditional ways of teaching were adopted and e-learning was the best option available to ensure learning continued and students and educators were kept safe (Maatuk et al., 2021). The UK adopted a similar approach where teaching was online, and assessments were "take home" assignments. Students completed these, uploading the completed assessment to a learning management system (LMS). Post-Covid, traditional assessment practices (exam-styled) have been re-introduced, in combination with online assessments that are completed remotely, uploaded to an LMS.

Current Challenges

During the Covid era university enrolment increased dramatically. Post Covid, these numbers have remained high, placing pressure on educators and students alike. Large cohorts enrolled in a course means that educators are under pressure to provide quality learning, assessment practices, feedback, and support, to large numbers of students.

With the rise of publicly accessible large language models (LLMs) such as ChatGPT and less one-to-one access to educators, students are increasingly relying on these tools when learning programming. This can potentially be detrimental when learning a skill like programming as coding requires practice. The use of ChatGPT may interfere with the learning process. However, ChatGPT offers an appealing shortcut, particularly for beginners.

Current challenges faced by students and educators are:

- Students struggle to assess their performance (Tam, 2021).
- Post Covid, students struggle to transition from "take home" assessments to exam-style ones (Aboagye et al., 2020).
- The continued reliance on ChatGPT may contribute to students not developing the essential problem-solving and coding skills required (Hermans, 2021).
- The reliance on ChatGPT does not guarantee enhanced performance (Becker et al., 2023; Xue et al., 2024).
- Time and attention given to smaller groups of students is reduced.
- Assessments completed as "take home" or online, over an extended period, can lead to students approaching programming tasks with the support of other students and the use of ChatGPT.
- Educators are unable to assess students' abilities when assessments are completed online.
- ChatGPT raise concerns about academic integrity and fairness.

Research shows that ChatGPT may be a good tool to use as part of a layered approach to learning (Becker et al., 2023). For example, students can verify their code using ChatGPT and constructive feedback can further their learning so ChatGPT can then be used towards formative assessment.

Methodology

This study employs a SWOT analysis framework to evaluate ChatGPT's effectiveness for formative assessment using a quantitative research approach.

Description of the Module

For the 2024/25 academic year the cohort (n=318) registered for a continuous assessment (CA) module, aimed at master's students learning programming for the first time. Students are either completing a year in computer science (YiCS) or registered for a degree conversion. Sixty percent of students have little to no programming experience. Within one term (12 weeks) students learn and engage with Python, used as a vehicle, to teach the fundamental programming concepts of programming, as well as more advanced concepts such as object-oriented programming and inheritance.

Formative Assessment Opportunities

Weekly formative assessment takes place in a lab. Students are tasked with completing worksheets, based on the content taught to them for that week. They are then asked to complete a survey, shown in Table 1, regarding the learning and the worksheet completed. Surveys are only conducted for the first half of the term, due to survey fatigue.

Table 1

Weekly Surveys Regarding Labs

Question	***Multichoice***	***Week 2 (n=88)***	***Week 3 (n=41)***	***Week 4 (n=29)***
I was able to comfortably master the concepts taught to me in the lecture this week.	*Absolutely*	*22 (25%)*	*5 (12%)*	*5 (17%)*
	Somewhat	*50 (57%)*	*30 (73%)*	*21 (72%)*
	Not really	*16 (18%)*	*6 (15%)*	*3 (10%)*
I managed to complete most of the exercises on the worksheet this week.	*Absolutely*	*59 (67%)*	*14 (34%)*	*18 (62%)*
	Somewhat	*28 (32%)*	*27 (66%)*	*11 (38%)*
	Not really	*1 (1%)*	*0 (0%)*	*0 (0%)*

Additionally, Table 2 shows the topics that students found difficult.

Table 2

Difficulty Learning Fundamental Programming Concepts

Week 2		Week 3		Week 4	
Concept	*Number reporting difficulty*	*Concept*	*Number reporting difficulty*	*Concept*	*Number reporting difficulty*
Calling functions	13	If…else	2	I still struggle with loops	6
Writing functions	17	While loops	12	Lists	2
Function signatures	22	For loops	11	Dictionaries	8
*args	44	Lists	48	Understanding larger applications	18
*kwargs	59				
Total	**155**		**73**		**34**

The challenge within this educational setting is the teaching assistant (TA) to student ratio, which averages 1:35. To alleviate the pressure on TAs, the use of ChatGPT was integrated into the lab for weeks 3 and 4 (see Table 3). The aim was for students to solve a problem and develop a solution in Python. They were instructed to engage with ChatGPT (as opposed to the TA's) to assist them with any difficulties they encountered when coding. As seen in Table 3, for week 3, students struggled to implement their own solution, however by week 4, there struggle was less notable. Table 3 also shows that students engaged with ChatGPT to explore incorrect coding; when they had difficulty creating a solution; and when they were grappling to formulate a solution. Further investigation is required to determine if the explanations provided by ChatGPT bridged the gap.

In week 9 students were tasked with completing a programming problem. They were also instructed to ask ChatGPT for a solution to the same problem and compare their solution to that of ChatGPT. Finally, they completed a quiz where the following question was put to them:

> "I have completed the solution as instructed in exercise 3 for this week's lab worksheet. I have studied the feedback and alternatives given to me by ChatGPT. I feel that I should be awarded the following grade (0 to 10) for my solution".

Table 3

Weekly Surveys Regarding ChatGPT Experience

Question	***Multichoice***	***Week 3***	***Week 4***
The solution that I implemented was like that of ChatGPT	*Yes*	*16*	*23*
	No	*19*	*13*
	I struggled to implement a solution	*25*	*5*
After reading ChatGPT's explanation of the code can you describe the code to someone	*Yes*	*49*	*39*
	No	*11*	*2*
After ChatGPT explained the code, I better understood how to solve the problem	*Absolutely*	*33*	*26*
	Somewhat	*25*	*14*
	Not really	*2*	*7*
Using ChatGPT helps me learn how to code	*Absolutely*	*38*	*27*
	Somewhat	*21*	*14*
	Not really	*1*	*0*
After learning how to create this function with the help of ChatGPT, could you now code something similar on your own?	*Yes*	*49*	*24*
	No	*11*	*16*
Do you trust that ChatGPT is providing you with correct knowledge?	*Yes*	*38*	*30*
	No	*21*	*11*

Table 4 shows the outcome of the number of students that allocated a grade to themselves based on their solution and comparing it to ChatGPT's solution. Although the results show that the average grade was 8.9, it does seem that students graded themselves towards the higher end. However, when reviewing the reasons for the grades, students graded themselves very fairly and provided valid reasons for the grade. In many cases they often grading themselves downwards.

Table 4

Student Self-evaluation of Their Programming Solution (0 – 10)

0	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
2	*0*	*0*	*4*	*1*	*2*	*12*	*15*	*64*	*66*	*53*

Summative Assessment Opportunities

Although not part of this study, it is important to note that summative assessments were also conducted. These consisted of two proctored (invigilated) two-hour assessments. For the proctored assessments, students are allowed access to learning materials. Additionally, students are allowed access to selected educational sites; however, access to ChatGPT is not allowed. During proctored tests, invigilators and plagiarism detection systems ensure compliance. Prior to the proctored assessments, students are encouraged to complete mock tests. They are expected to complete these mocks independently; however, support is provided. For all proctored tests, to ensure integrity, multiple monitoring systems are employed:

- Extended time and small rooms are provided to students with reasonable adjustment plans (RAPs).
- Invigilators supervise the timed tests.
- Manual grading enables instructors to identify ChatGPT-generated solutions (often abstract or generic and misaligned with coding techniques emphasised in the curriculum).
- Automated plagiarism detection tools compare student submissions.

The grading is managed by six educators, each having access to a shared spreadsheet that finely details how scores are allocated. Additionally, a column for feedback is also included. The grade and the feedback are provided to students. Table 5 shows the overall performance of the cohort comparing 2023/24 (inflated grades due to the Covid era of "take home" assessments) and 2024/25 ("take home" assessments were replaced with proctored assessments to ensure academic fairness).

Table 5

Performance for the Module over 2 Years

Academic year	***Students (n)***	***Pass rate (%)***	***Failures (n, %)***	***t-test: Pass Rate (p-value)***
2024/25	*318*	*67%*	*74 (23%)*	*p = 0.015 * (2023/24 – 2024/25)*
2023/24	*187*	*75%*	*5 (3%)*	*p = 0.042 * (2022/23 – 2023/24)*

Formative Assessment: SWOT Analysis

A SWOT is a structured planning tool used in research (and other sectors) to evaluate internal and external factors regarding a topic. As a research methodology SWOT can assist in assessing the effectiveness regarding an area of interest.

To answer the two research questions, the data from the formative assessments was analysed using a SWOT analysis.

Results and Discussion

Table 6 presents the results of the SWOT analysis derived from the quantitative data in Tables 2, 3 and 4.

Table 6

SWOT Analysis to Answer RQ1 & RQ2

Strengths	Weaknesses	Opportunities	Threats
1. Effective learning support 2. Encourages independent problem-solving 3. Alleviates teaching assistant (TA) workload 4. Self-evaluation and reflection	1. Potential for inaccuracy and misconceptions 2. Struggles with implementation 3. Limited personalisation and adaptive feedback	1. Enhancing automated tutoring capabilities 2. Developing AI literacy among students 3. Bridging the TA gap in large cohorts 4. Gamification and interactive learning	1. Academic integrity concerns 2. Varying accuracy and bias in AI responses 3. Resistance to AI adoption

To provide a more comprehensive explanation of the SWOT analysis mapped in Table 6, further discussion is required to reflect on the pedagogical, technical and ethical considerations when contemplating ChatGPT as an educational tool to support formative assessments.

Strengths

1. Effective learning support
   - Most students found that ChatGPT helped them to code (Week 3, Week 4).
   - Students reported improved understanding after engaging with ChatGPT (Week 3, Week 4).
2. Encourages independent problem-solving
   - Many students felt confident in coding similar problems after using ChatGPT (Week 3, Week 4).
   - ChatGPT explanations improved students' ability to describe code (Week 3, Week 4).
3. Alleviates TA workload
   - Given the high TA-to-student ratio (1:35), ChatGPT serves as an additional learning resource, reducing reliance on human assistance (Week 3).
4. Self-evaluation and reflection
   - Students compared their solutions with ChatGPT (Week 9), developing self-assessment skills (Table 4).

Weaknesses

1. Potential for inaccuracy and misconceptions
   - Some students did not trust ChatGPT's responses (week3, Week 4).
   - ChatGPT lacks deep conceptual awareness as students still struggled with function parameters and loops (seen in Table 2).
2. Struggles with implementation
   - Some students struggled to implement solutions even after using ChatGPT (Week 3, Week4).
3. Limited personalisation and adaptive feedback
   - Unlike human instructors, ChatGPT does not adapt explanations to individual student needs in real-time (Week 4).

Opportunities

1. Enhancing automated tutoring capabilities
   - Integrating ChatGPT with personalised hints and adaptive scaffolding could improve effectiveness.
2. Developing AI literacy among students
   - Using ChatGPT helps students critical analyse AI-generated content, fostering AI literacy and debugging skills.
3. Bridging the TA gap in large cohorts

- Expanding ChatGPT's role in routine formative assessments can further support students without additional staffing costs.

4. Gamification and interactive learning
   - Incorporating ChatGPT into gamified coding challenges could make learning more engaging.

Threats

1. Academic integrity concerns
   - Students may become over-reliant on AI-generated solutions impacting original problem-solving skills.
2. Varying accuracy and bias in AI responses
   - ChatGPT-generated code may contain errors or inefficiencies potentially reinforcing misconceptions (Week 3, Week 4).
3. Resistance to AI adoption
   - Some educators and students may distrust or resist using AI in assessment and learning (Week 3, Week 4).

Summary of SWOT Analysis

In relation to the research questions, the SWOT analysis underscores that ChatGPT offers notable educational benefits for supporting formative assessment in CS education. However, it also draws attention to potential challenges and risks associated with its use. ChatGPT can function as a valuable personal tutor, fostering independent learning. At the same time, over-reliance on such tools may impede students' conceptual understanding and problem-solving development. While ChatGPT has the potential to enhance problem solving abilities, it may inadvertently hinder learning if students adopt inefficient or incorrect coding solutions without critically evaluating the code. Thus, ChatGPT presents a double-edged sword: its effectiveness is highly dependent on how it is integrated into pedagogical practice.

Conclusion

The SWOT analysis highlights ChatGPT's potential to serve as an automated tutor and a valuable educational tool in supporting students with formative assessments. However, its integration must be approached with caution, given notable limitations such as concerns around academic integrity, ethical implications, accuracy, over-reliance, and constraints in delivering personalised feedback.

Future work will focus on the development of robust AI-assisted learning frameworks; the enhancement of critical AI literacy amongst students and educators; and proactive measures to address ethical considerations related to academic integrity.

References

Aboagye, E., Yawson, J. A., & Appiah, K. N. (2020). COVID-19 and E-Learning: The challenges of students in tertiary institutions. *Social Education Research, 2*(1), 1-8. https://doi.org/10.37256/ser.212021422

Banerjee, P., Srivastava, A. K., Adjeroh, D. A., Reddy, R, & Karimian, N. (2025). Understanding ChatGPT: Impact analysis and path forward for teaching computer science and engineering. *IEEE Access, 13*, 11049-11069. https://ieeexplore.ieee.org/abstract/document/10833612

Becker, B. A., Craig, M., Denny, P., Keuning, H., Kiesler, N., Leinonen, J., Luxton-Reilly, A., Malmi, L., Prather, J., & Quille, K. (2023). Generative AI in Introductory Programming. *Computer Science Curricula*, 1-25.

Hermans, F. (2021). *The Programmer's Brain: What every programmer needs to know about cognition*. New York USA: Manning Publications Co.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günneman, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T.,. . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Maatuk, A. M., Elberkawi, E. K., Aljawarneh, S., Rashaideh, H., & Alharbi, H. (2021). The COVID-19 pandemic and E-learning: challenges and opportunities from the perspective of students and instructor. *Journal of Computing in Higher Education, 34*, 21-38. https://doi.org/10.1007/s12528-021-09274-2

Malinka, K., Peresíni, M., Firc, A., Hujnák, O., & Janus, F. (2023). On the educational iImpact of ChatGPT: Is artificial intelligence ready to obtain a university degree? *ITiCSE 2023: Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1 (pp.* 47 - 53). https://doi.org/10.1145/3587102.3588827

Qureshi, B. (2023). (2023). Exploring the use of ChatGPT as a tool for learning and assessment in undergraduate computer science curriculum: Opportunities and challenges. *arXiv:2304.11214*. doi:http://dx.doi.org/10.48550/arXiv.2304.11214

Rahman, Md. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. Applied Science, 13(9), 5783. https://doi.org/10.3390/app13095783

Tam, A. C. F. (2021). Students' perceptions of and learning practices in online timed take-home examinations during Covid-19. Assessment & Evaluation in Higher Education, 47(3), 477-492. https://doi.org/10.1080/02602938.2021.1928599

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, 15. https://doi.org/10.1186/s40561-023-00237-x

Xue, Y., Chen, H., Bai, G. R., Tairas, R., & Huang, Y. (2024). Does ChatGPT help with introductory programming? An experiment of students using ChatGPT in CS1. *ICSE-SEET '24: Proceedings of the 46th International Conference on Software Engineering: Software Engineering Education and Training* (pp. 331 - 341). https://doi.org/10.1145/3639474.3640076

Author Details

Jacqui Chetty
School of Computer Science, College of Engineering and Physical Sciences
University of Birmingham
UNITED KINGDOM
j.chetty@bham.ac.uk

d Jacqui Chetty

TOWARDS EFFECTIVE AUTOMATED GRADING IN CS1: A COMPARISON BETWEEN GenAI AND THE IN-HOUSE GRADING TOOL

Pieter Joubert, Wendy Yanez-Pazmino, and Jacqui Chetty
University of Birmingham
United Kingdom

Abstract

The adoption of Generative Artificial Intelligence (GenAI) in automating assessment has become increasingly popular in introductory computer science (CS1) modules, especially for large student cohorts. This study evaluates the effectiveness of a GenAI-based grading tool built using OpenAI API, in comparison to a traditional automated grading system. We conducted the study on formative assessments submitted by first-year students in the Object-Oriented Programming (OOP) module, which was graded by both systems. The findings reveal that the grades generated by the GenAI were as accurate as those produced by the automated grading system. These results suggest that the integration of GenAI into the grading process for formative assessments can optimise the marking and grading for educators and potentially improve student learning.

Introduction

Artificial intelligence (AI) is a distinctive research field within computer science (CS) (Crompton & Burke, 2023) and its recent advancements, particularly in Generative Artificial Intelligence (GenAI) are reshaping higher education (HE). In HE, the inclusion of AI and GenAI is becoming increasingly popular. This trend has heated intense debate surrounding the advantages, disadvantages, threats and opportunities that AI creates. Without robust policies, ethical frameworks, and collaborative guidance, the adoption of AI in HE may lead to unintended consequences (Bond et al., 2024). Despite a steady rise in research on GenAI in education over the last 5 years, key gaps remain, especially around assessment and grading, where academic integrity concerns are increasing (Tobler, 2024).

Studies show that GenAI is increasingly used for tasks such as grading essays, evaluating free-text responses, and analysing cognitive engagement (Crompton & Burke, 2023). However, traditional coursework assessment and grading are under threat, encouraging educators to rethink how assessments are designed and evaluated (Chan, 2023; Raman & Kumar, 2022). Additionally, these should include the use of GenAI as Rudolph et al. (2023) advise against the idea of policing students or focusing on academic misconduct for using GenAI. To remedy the current need for evaluating grading methods and providing grading tools that are reliable within CS HE, in this study, a comparison was drawn between the designed automated grading system and a GenAI grading tool.

This study addresses the current need for reliable, scalable tools in computer science education by comparing traditional automated grading systems with a GenAI grading tool. In particular, the study investigates the accuracy and efficacy of the two marking methods when assessing programming assessments. The main contributions of this work are: (a) a GenAI grading tool using different prompts that can be used by lecturers in CS1 modules to automatically grade formative assessments and (b) an empirical comparison of GenAI and automated grading systems in terms of accuracy and performance.

Related Work

Automated grading systems and tools for programming assessments have evolved significantly over the past decade. Traditional approaches often rely on static code analysis or unit testing to evaluate the correctness and functionality of code (Ala-Mutka et al., 2004; Rahman & Nordin, 2007). However, these methods do not fully capture the reliability, efficiency, and complexity of the code, which are critical in educational settings (Van Verth, 1985). Other approaches, like static analysis, are becoming popular among automated testing techniques (Antonucci et al., 2015), as they mainly consider the source code and its abstraction representation, resulting in a fairer evaluation. Similarly, recent work has been done on the use of big data to collect data from programming assessments to identify behavioural patterns and learning flaws, but there is still a large amount of wasted data and tools that cannot fully capture code changes (Antonucci et al., 2015).

Recently, there has been a proliferation of GenAI tools, specifically, Generative Pre-Trained Transformer (GPT-)3 being labelled by *The Economist* as "eerily human-like" and ChatGPT seen as "scary good, crazy-fun" (Kantrowitz, 2022) for automated programming assignment grading. Both GPT-3 and ChatGPT are owned by OpenAI, an organisation that has transformed from a non-profit to a for-profit corporation. Research indicates that ChatGPT can be used for automatic code checking, enabling teaching and grading large groups of students without burdening teacher times (Bang et al., 2023; Jukiewicz, 2024; Mekterovic & Brkic, 2017; Rahman & Nordin, 2007; Rahman & Watanobe, 2023; Raman & Kumar, 2022; Ullah et al., 2018). However, the effectiveness of using GenAI to grade formative assessments for large cohorts, compared to traditional unit test–based automated tools, remains underexplored.

Methods

Description of the Module

This study was conducted within an introductory Object-Oriented Programming (OOP) module offered in the autumn term to both B.Sc. and M.Sc. Computer Science students. The module provides foundational programming skills, covering topics such as control structures, data structures, classes and objects, inheritance, and file handling. Although the module includes both formative and summative assessments, this study focuses on a single formative task, which required students to implement and evaluate their code using lecturer-designed unit tests. Students were expected to verify test outcomes prior to submission. The module is supported by weekly two-hour lectures, two-hour lab sessions, lecturer office hours, and additional one-to-one drop-in sessions.

Motivation for the Project

Over the past four years, assessments in the OOP programming module have been graded using standardised, test-case-based automated grading systems, primarily through unit testing. Lecturers design the unit tests, which are validated by teaching assistants (TAs) to ensure accuracy, reliability, and alignment with the assessment brief. Although the development of assessments and implementation of unit tests is time-intensive, automated grading significantly reduces marking time and promotes fairness by applying consistent evaluation criteria. Students are provided with unit tests to validate their code prior to submission, while an extended set of tests is used for final grading. The system assigns marks mainly based on test outcomes, requiring submissions to be valid, compilable Java code. However, this binary approach, which can be rigid, often results in full marks or zero, limiting edge assessment cases and occasionally necessitating additional moderation.

Data Collection and Analysis

As the purpose of the study was to determine whether GenAI grades assessments as accurately as the automated grading system given the large cohort of students, quantitative research was the most suitable methodology to adopt.

Assessment Structure

A formative programming assessment was designed to evaluate the creation of a single Java class that makes use of methods, constructors, getters, and setters. Students were provided with instructions and a Maven template project as a resource to create and develop the assessment. This project included a selection of unit tests that students could make use of to test their work before submission. It also included empty Java classes that students needed to update to complete the assessment. The students were given a week to complete the formative assessment and were required to submit the entire project folder to allow the automated grading system to grade it.

Automated Grading

For this module, an automated marking system based on "Junit" tests was used to evaluate students' submitted code, as a project, and to provide a grade. The automated marking system required a working project that could be compiled to run the designed unit tests. The automated marking system also provided limited feedback in the form of a breakdown of the test results in a human-readable format, as shown in Figure 1.

Figure 1

Example Feedback from the Automated Grading System

```
E0Feedback for: xxxxxxxxxx
Marks:(TestName Mark)
    add 1
    multiply 1
    subtraction 1
    division 1
    circle area 0
    sphereVolume    1
Total [ %]:  83.0
Problems found before running tests:
```

GenAI grading

To perform the GenAI grading, a Python script was used to access the OpenAI API, using the "gpt-3.5-turbo-0125" model. Firstly, a prompt without specific assessment information, shown in Figure 2, was run on both assessments. The instructions given to the GenAI grading tool, seen in Figure 2, produced a result in a JSON format to allow for easy analysis.

Figure 2

Prompt 1 Used to Mark Formative Assessments 1 and 2

> *You are a precise markings assistant designed to mark first year java programs and designed to output JSON. In the JSON name each method with its exact method signature and mark each method out of 1. Do not give half marks.*

Secondly, the specific assessment instructions for each formative assessment (shown in Figure 3 for formative assessment 1 and Figure 4 for formative assessment 2) were added to the prompt seen in Figure 2 above. All these prompts were run in conjunction with a user prompt to instruct the GenAI grading tool.

Figure 3

Prompt 2 Used to Mark Formative Assessment 1

You are a precise markings assistant designed to mark first year java programs and designed to mark first year java programs and designed to output JSON. In the JSON name each method with its exact method signature and mark each method out of 1. Do not give half marks. *Base the marks on the following question:* *Complete the class named BankAccount, which has 4 attributes: clientName, clientID, accountBalance and a boolean to check whether the account is closed or not.* *The class includes a constructor with 3 parameters 4 getters (1 for each attribute) and the following methods:* *deposit() method: the balance increases with the depositAmount;* *withdraw() method: the balance decreases with the withdrawalAmount;* *transferTo( ) method: the amount is transferred from the current account to another account (make use of the methods to update both balances accordingly);* *closeAccount() method: upon closing an account, the balance should be set to zero* *The class and method signatures are provided in the template file BankAccount.java* *All methods are tested in BankAccount.java* *Please follow the submission instructions on Canvas.*

Figure 4

Prompt 2 Used to Mark Formative Assessment 2

You are a precise markings assistant designed to mark first year java programs and designed to output JSON. In the JSON name each method with its exact method signature and mark each method out of 1. Do not give half marks. Base the marks on the following question: Complete the class named BillingManager, which includes a default constructor with VAT = 20%, a second constructor that takes in the VAT parameter so that the default value can be modified and three overloaded computeBill() methods for a book store: *• When computeBill() receives a single parameter, it represents the price of one book ordered. Add the VAT and return the total due.* *• When computeBill() receives two parameters, they represent the price of the book, and the quantity ordered. Multiply the two values, add the VAT and return the total due.* *• When computeBill() receives three parameters, they represent the price of the book, the quantity ordered and a voucher value. Multiply the price and quantity, reduce the result by the voucher value, and then add the VAT and return the total due.*

For each student, the prompts in Figure 2, Figure 3 and Figure 4 were run for both formative assessments. This was repeated five times to improve consistency and minimise the impact of possible GenAI hallucinations. An example of the result generated by the GenAI grading tool is shown in Figure 5. One at the end of the statement indicates a mark; otherwise, if left blank, no mark was awarded.

Figure 5

GenAI Grading Tool Result

```
{
    " BankAccount (String, int,double)":1,
    " getName() " :1,
    " getID() " :1,
    " getBalance()":1,
    " getClosed() ":1,
    "deposit(double)":1,
    "withdraw(double)":,
    "closeAccount()":,
"transferTo(BankAccount,double)":1

}
```

The total grades for each student's submission were used in the comparison with the marks generated from the automated grading system.

Results and Discussion

To evaluate the statistical differences between the automated grading system and the GenAI grading tool using prompt 1 and prompt 2, a paired sample t-test and a correlation coefficient were used for this study. It is worth noting that prompt 1 does not include the question context, while prompt 2 does include the question context. The total number of student code submissions was 781.

Results

Table 1 presents the individual score for each run, as well as the mean scores for both GenAI prompt 1 and prompt 2 for both formative assessments. The prompts were run separately five times to evaluate the consistency of each prompt. The average score for formative assessment (FA) 1 - prompt 1 across the five runs was 99.39, while FA 1 - prompt 2 achieved an average score of 98.40 across all five runs. The average score for FA 2 - prompt 1 across the five runs was 97.28, while FA 2 - prompt 2 achieved an average score of 92.72 across all five runs.

Table 1

A Comparison of GenAI Prompt 1 and Prompt 2 over Five Runs for Both Formative Assessments

	Formative Assessment 1 (FA 1)		Formative Assessment 2 (FA 2)	
Runs	**Prompt 1**	**Prompt 2**	**Prompt 1**	**Prompt 2**
1	99.34	98.43	97.28	93.28
2	99.52	98.66	97.31	91.73
3	99.45	97.72	97.19	92.43
4	99.48	98.82	97.27	93.01
5	99.15	98.37	97.36	93.20
Mean	99.39	98.40	97.28	92.72

Table 2 presents a comparative analysis of the automated grading system against GenAI prompt 1 and prompt 2. The results show the mean of the automated marking and the mean of all the runs using prompt 1 and prompt 2; the difference in means between the automated marking and the means of running prompt 1 and prompt 2; a significance value generated from a T-test comparing both prompts to the automated marking; and a correlation coefficient comparing both prompts to the automated marking.

Table 2

A Comparison of Automated Grading System with GenAI Grading Tool

	Formative Assessment 1			Formative Assessment 2		
	Automated Grading System	GenAI Grading Tool		Automated Grading System	GenAI Grading Tool	
Analysis		Prompt 1	Prompt 2		Prompt 1	Prompt 2
mean	93.98	99.39	98.40	95.72	97.28	94.79
mean difference	-	5.41	4.42		1.56	3.00
significance	-	8.06E-12	5.94E-09		0.013	1.3415E-05
correl. coefficient	-	0.246	0.326		0.432	0.436

Table 3 shows a comparison similar to that of Table 2 between the automated grading system and GenAI grading tool; however, in this case all grades of zero provided by the automated marker, and their corresponding results from GenAI grading tool, were removed. The reason for this is due to the automated grading system only marking correctly compiled projects. Students whose code was technically correct but submitted in the incorrect format or with minor syntax errors, would receive zero from the automated grading system, while GenAI grading tool would assess the code.

Table 3

A Comparison of Automated Marking with GenAI, Zeros Removed

Analysis	**Formative Assessment 1**			**Formative Assessment 2**		
	Automated Grading System	**GenAI Grading Tool**		**Automated Grading System**	**GenAI Grading Tool**	
		Prompt 1	**Prompt 2**		**Prompt 1**	**Prompt 2**
mean	98.92	98.92	99.60	97.98	99.39	98.61
mean difference	-	-0.678	0.005	-	0.699	3.44
significance	-	0.006	0.984	-	0.377	8.91E-25
correl. coefficient	-	0.361	0.406		0.920	0.616

The distribution of grades from the automated grading system is shown in Table 4.

Table 4

Marks Distribution from the Automated Grading System

Grade	**Occurrence**
0	39
20	3
40	1
60	5
80	15
100	716

The findings indicate a mixed set of results, showing that, given a set of circumstances, GenAI grading tool performs as well as the automated marking system on efficacy. The weak to moderate, positive correlation for prompt 1 ($r=0.361$) and prompt 2 ($r=0.406$) for formative assessment 1, seen in Table 3, indicates that there is a relationship between the automated grading system and the GenAI grading tool. Similarly, for formative assessment 2 (seen in Table 3), prompt

1 (r=0.919) and prompt 2 (r=0.615) show high levels of correlation. These results are discussed next.

Discussion

Several interesting points are worthy of discussion based on the results in the previous section. Firstly, Table 1 shows the consistency of the GenAI grading tool for both prompts. In other words, GenAI hallucinations were low. This could be due to a strong focus on providing GenAI with: (a) clear and specific prompts; (b) enough information; and (c) avoiding ambiguous prompts that could lead to misinterpretation.

Secondly, Table 2, based on the low significance and correlation coefficient between the results of prompt 1 and prompt 2 in comparison to the marks from the automated grading system, shows that the automated grading system and GenAI grading tool results were notably different.

Thirdly, to further identify whether this difference was due to errors in the way that the GenAI grading tool marks the assessments, the individual grades were inspected. It became clear that the automated grading tool marks had a higher bimodality, with most results being either 100% or 0% (see Table 4).

Lastly, to investigate the impact of this bimodality, all grades of zero were removed from the data and the same comparisons were performed. The results of this in Table 3 indicate that the GenAI grading tool performs equally well as the automated marking system (based on means, significance and correlation coefficient). This is due to the fact, as mentioned above, that the automated marking system graded only correctly compiled code.

The difference in results between prompt 1 and prompt 2 for formative assessment 1 (in Table 3) is of interest. The GenAI grading tool with prompt 2 (i.e., with the question context included) more closely matches the automated grading system on all metrics (mean difference, significance, and correlation coefficient). From this, it seems that the inclusion of the question context in the prompt results in grading that is as accurate as that done by the automated grading system. In contrast for formative assessment 2, there is a much larger discrepancy between prompt 2 and the automated grading system. It seems clear that the inclusion of more context changes the result of the GenAI grading tool-based marking process.

Conclusion

The results suggest that GenAI could be a potential alternative to the automated grading system, especially in CS1 modules with large student cohorts. The use of

GenAI could save lecturer marking time as well as assist in the development of unit tests and provide potential solutions. It could also enhance the student learning experience by using GenAI as a supportive tool. The findings further underscore the importance of conducting multiple runs when using GenAI grading tools to ensure consistency and reliability of assessment outcomes, and to mitigate the impact of potential hallucinations. Furthermore, it is interesting to note that providing GenAI with the assessment question, as well as eliminating the zeros, may infer that GenAI grading tools marked as accurately as the automated grading system.

While these insights showcase the advantages GenAI may offer, it is crucial to reflect on broader pedagogical implications. Beyond reducing marking time, educators should also consider how GenAI can enhance fairness and transparency in assessment processes, ensuring that assessment practices are not only efficient but also equitable and explainable, thus providing meaningful feedback to students. Moreover, GenAI has the potential to democratise access to formative assessment while upholding academic standards.

Nonetheless, this study highlights several limitations. The interpretability and transparency of GenAI grading decisions remain an area of concern, especially when complex or subjective criteria form part of the grading process. Further research needs to be conducted to understand how GenAI interprets and evaluates student work when providing more contextual information, including assessment instructions and marking criteria, to name a few. Ethical considerations around fairness, bias, and data privacy should be addressed to ensure responsible adoption. Additionally, the tendency of GenAI to generate inconsistent outputs across multiple runs necessitates further research into strategies that minimise discrepancies and enhance GenAI grading tools robustness.

References

Ala-Mutka, K., Uimonen, T., & Järvinen, H. M. (2004). Supporting students in C++ programming courses with automatic program style assessment. *Journal of Information Technology Education: Research, 3*, 245-262. https://jite.org/documents/Vol3/v3p245-262-135.pdf

Antonucci, P., Estler, C., Nikolić, D., Piccioni, M., & Meyer, B. (2015). An incremental hint system for automated programming assignments. *ITiCSE '15: Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education (*pp. 320 - 325). https://doi.org/10.1145/2729094.2742607

Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, W., Do, Q. V., Xu, Y., & Fung, P. (2023). A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. *Computation and Language*. https://doi.org/10.48550/arXiv.2302.04023

Bond, M., Khosravi, H., De Laat, M., Bergdahl, M., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21*, 4. https://doi.org/10.1186/s41239-023-00436-z

Chan, H. C. B. (2023). Grading generative AI-based assignments using a 3R framework. In *2023 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, Auckland, (pp. 1-5). **DOI:** 10.1109/TALE56641.2023.10398408

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education, 20,* 22. https://doi.org/10.1186/s41239-023-00392-8

Jukiewicz, M. (2024). The future of grading programming assignments in education: The role of ChatGPT in automating the assessment and feedback process. *Thinking Skills and Creativity, 52,* 101522. https://doi.org/10.1016/j.tsc.2024.101522

Kantrowitz, A. (2022, December 2). *Finally, an A.I. chatbot that reliably passes "the Nazi test"*. Slate. https://slate.com/technology/2022/12/chatgpt-openai-artificial-intelligence-chatbot-whoa.html

Mekterovic, I., & Brkic, L. (2017). Setting up automated programming assessment system for higher education database course. *International Journal of Education and Learning Systems, 2, 287-294.*

Rahman, K. A., & Nordin, M. J. (2007). A review on the static analysis approach in the automated programming assessment systems. In *Proceedings of the 2007 National Conference on Programming (pp.* 1–15).

Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Science, 13*(9), 5783. https://doi.org/10.3390/app13095783

Raman, A., & Kumar, V. (2022). Programming pedagogy and assessment in the era of AI/ML: A position paper. In V. Choppella & A. Karkere (Eds.), *COMPUTE '22: Proceedings of the 15th Annual ACM India Compute Conference* (pp. 29 - 34). https://doi.org/10.1145/3561833.3561843

Rudolph, J., Tan, S., & Tan S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Tobler, S. (2024). Smart grading: A generative AI-based tool for knowledge-grounded answer evaluation in educational assessments. *MethodsX 12, 102531*. https://doi.org/10.1016/j.mex.2023.102531

Ullah, Z., Lajis, A., Jamjoom, M., Altalhi, A., Al-Ghamdi, A., & Saleem, F. (2018). The effect of automatic assessment on novice programming: Strengths and limitations of existing systems. *Computer Applications in Engineering Education, 26*(6), 2328-2341. https://doi.org/10.1002/cae.21974

Van Verth, P. B. (1985). *A system for automatically grading program quality (Metrics, software metrics, program complexity)*. [Doctoral dissertation, Computer Science Department, State University of New York at Buffalo].

Author Details

Pieter Joubert
School of Computer Science, College of Engineering and Physical Sciences
University of Birmingham
p.joubert@bham.ac.uk

Wendy Yanez-Pazmino
School of Computer Science, College of Engineering and Physical Sciences
University of Birmingham, UK
w.yanez@bham.ac.uk

Jacqui Chetty
School of Computer Science, College of Engineering and Physical Sciences
University of Birmingham, UK
j.chetty@bham.ac.uk

STUDENT EXPERIENCES WITH CHATGPT IN HIGHER EDUCATION: INSIGHTS FROM A TWO-YEAR GLOBAL STUDY

Aleksander Aristovnik and Dejan Ravšelj
Faculty of Public Administration, University of Ljubljana
SLOVENIA

Abstract

This study examines global student experiences with ChatGPT in higher education during the 2023–2024 and 2024–2025 academic years, based on survey data from 23,218 and 22,963 students, respectively. Focusing on usage patterns, satisfaction, and attitudes, the findings reveal a marked increase in the adoption and integration of generative artificial intelligence into academic routines. ChatGPT was most frequently used for study support, summarization, and research. As usage intensified, students reported greater satisfaction, perceiving the tool as more useful, accessible, and effective than traditional resources. These trends highlight ChatGPT's growing role in enhancing learning in higher education.

Introduction

The introduction of the conversational chatbot ChatGPT in November 2022 marked a major step forward in the use of artificial intelligence within higher education. Developed by OpenAI in San Francisco, California, ChatGPT quickly attracted widespread interest among students due to its natural language processing abilities, which allow for seamless and intuitive communication with users (Alessandri-Bonetti et al., 2024; Mohmad, 2023). Although ChatGPT was originally designed to simulate human conversation, its functionality extends well beyond this purpose by supporting a wide variety of academic and practical tasks (Boubker, 2024; Das, 2024).

Widely regarded as one of the most advanced and rapidly adopted consumer applications of artificial intelligence, ChatGPT has received significant attention from the global education community, inspiring both strong support and serious concerns within higher education institutions (Tlili et al., 2023; Twinomurinzi & Gumbo, 2023). Supporters of ChatGPT emphasize its benefits for higher education, including the ability to provide real-time feedback, personalized learning support, cross-platform accessibility, and more effective use of open educational resources. These features are seen as having the potential to positively influence learning outcomes and skills development. At the same time, critics point to a number of risks such as data privacy issues, algorithmic bias, reduced student motivation, academic dishonesty, and the spread of false or misleading information, often referred to as artificial intelligence hallucinations. These concerns have led to calls for stronger ethical safeguards, clearer oversight, assessment practices that prioritize originality, and more rigorous content verification to ensure the integrity of academic work (Michalon & Camacho-Zuñiga, 2023; Williams, 2024; Ravšelj, Keržič, et al., 2025).

Existing studies have identified a range of ways in which ChatGPT can be used by students. These include support for academic writing, study assistance, language learning, idea generation, research help, and personal organization (Boubker, 2024). Such uses highlight the flexibility and usefulness of ChatGPT as a tool that can support students in enhancing their educational experience. This opens up many opportunities to integrate artificial intelligence into routine academic tasks. The capabilities of ChatGPT continue to grow, especially following the release of the Generative Pre-trained Transformer (GPT-4o) in May 2024. This version introduced significant improvements in multimodal functionality and also made some features accessible to users without paid subscriptions. These advancements further establish ChatGPT as a transformative technology for educational settings (Dong et al., 2024).

According to the existing literature, students also express strong satisfaction and favorable perceptions of ChatGPT, particularly valuing its immediate, in-depth answers and support in understanding complex subjects (Ajlouni et al., 2023; Park, 2023). The tool is frequently praised for its ability to simplify difficult concepts, offer explanations in natural language, and provide guidance tailored to individual learning needs. Greater familiarity with ChatGPT and more positive attitudes toward its use are associated with increased engagement, especially among students in the final stages of their academic programs, who often rely on such tools for research and exam preparation (Pallivathukal et al., 2024). This satisfaction is largely driven by the perceived efficiency, availability, and personalized learning experience that ChatGPT offers, which helps students study more effectively and independently. Additionally, ChatGPT supports academic research by providing quicker access to scholarly materials, summarizing large volumes of information, and generating initial ideas or frameworks, thereby enhancing students' overall academic productivity and satisfaction. Nonetheless, persistent concerns remain regarding the reliability, accuracy, and potential bias of its responses, highlighting the importance of cross-verifying information with credible academic sources and promoting responsible, critical use among students (Ait Baha et al., 2024; Ravšelj, Keržič, et al., 2025).

While higher education has been extensively studied in the context of emerging digital technologies, there remains a notable gap in understanding the specific impacts introduced by ChatGPT. Most existing research has focused on broader trends in digital transformation, leaving the particular dynamics of artificial intelligence adoption underexplored. Therefore, this study seeks to contribute to the existing body of knowledge by providing empirical insights into how students across the globe are engaging with ChatGPT in its early stages of adoption. It focuses specifically on students' usage patterns, their satisfaction with the tool, and their attitudes toward its role in academic life. The remainder of the study is structured as follows. The next section describes the methodology, including data collection techniques and analytical procedures. This is followed by a section presenting the main results, highlighting key evidence-based findings. Finally, the concluding section summarizes the key findings and reflects on their broader implications.

Methodology

Data collection was conducted in two waves using the Global ChatGPT Student Survey, which was initiated by the Faculty of Public Administration at the University of Ljubljana, Slovenia. The first wave, carried out between October 2023 and February 2024, gathered students' initial perceptions of ChatGPT (Ravšelj, Aristovnik, et al., 2025), while the second wave, conducted between October 2024 and February 2025, explored how these perceptions evolved over time (Aristovnik et al., 2025). The survey instrument was pilot tested with students from Slovenia to improve its clarity and usability (see Aristovnik et al., 2024). It was developed in line with established best practices in survey design to ensure content relevance and practical applicability. To ensure broad international participation, the survey was made available in seven languages: English, Italian, Spanish, Turkish, Japanese, Arabic, and Hebrew. Participants were higher education students aged 18 or older who were legally capable of providing informed and voluntary consent to take part in the anonymous survey (Ravšelj, Keržič, et al., 2025). A convenience sampling strategy was employed, with the survey disseminated through classroom activities and institutional communication channels, following a commonly used approach in educational research to engage students who were easily accessible and willing to participate (Boubker, 2024; Sarstedt et al., 2018).

The online survey instrument consisted of several sections that examined dimensions both explicitly and implicitly connected to the use of ChatGPT. These included socio-demographic characteristics, knowledge and experiences, capabilities, ethical governance and concerns, satisfaction and attitude, study issues and outcomes, skills development, labour market and skills mismatch, emotions, study and personal information, and general reflections. The majority of these dimensions were assessed through closed-ended items using a 5-point Likert scale, with responses ranging from 1 (e.g., strongly disagree) to 5 (e.g., strongly agree)

(Ravšelj, Aristovnik, et al., 2025). The structure of the questionnaire remained consistent across both waves of data collection, with the exception of a slight revision to the item regarding the use of ChatGPT and other generative artificial intelligence tools. In the initial wave, students were asked exclusively about their use of ChatGPT, whereas in the second wave, the question was broadened to encompass a range of tools, including ChatGPT (OpenAI), Microsoft Copilot, Google Gemini (formerly Bard), Perplexity AI, Claude AI (Anthropic), and an open-ended option labelled "Other." While the survey addressed a wide range of topics, the present study specifically investigates students' usage patterns, as well as their satisfaction and attitudes, through a comparative perspective across two successive academic years.

The data were analysed using two primary statistical approaches. The first involved descriptive analysis through the computation of Top 2 Box (T2B) scores, indicating the percentage of respondents who selected the highest two categories ("agree" and "strongly agree") on a 5-point Likert scale. The second method consisted of an independent samples t-test, which was employed to compare mean values between students' initial and evolving perceptions of their usage patterns, as well as their satisfaction and attitudes. This parametric procedure is widely acknowledged as a reliable and standard technique for identifying mean differences between two independent groups (Rasch et al., 2007).

Results

By the end of the first data collection wave (2023–2024), a total of 23,218 students from 109 countries and territories had taken part in the survey. In the second wave (2024–2025), participation included 22,963 students from 120 countries and territories (see Table 1). The socio-demographic profile of the participants remained largely consistent across both waves, enabling meaningful comparisons between the two datasets. In both waves, the majority of respondents were female (58.8% in the first wave and 61.5% in the second), and most were enrolled in undergraduate (first-level) study programs (83.4% and 78.5%, respectively). Most students were studying in the field of social sciences (41.4% in the first wave and 43.0% in the second), followed by applied sciences, with fewer students in natural and life sciences, as well as arts and humanities. The dominant modes of study were traditional learning (47.3% in the first wave and 42.5% in the second) and blended learning (43.2% and 49.7%, respectively), while a smaller share of students participated in fully online learning.

Table 1

Socio-demographic Characteristics of the Survey Participants

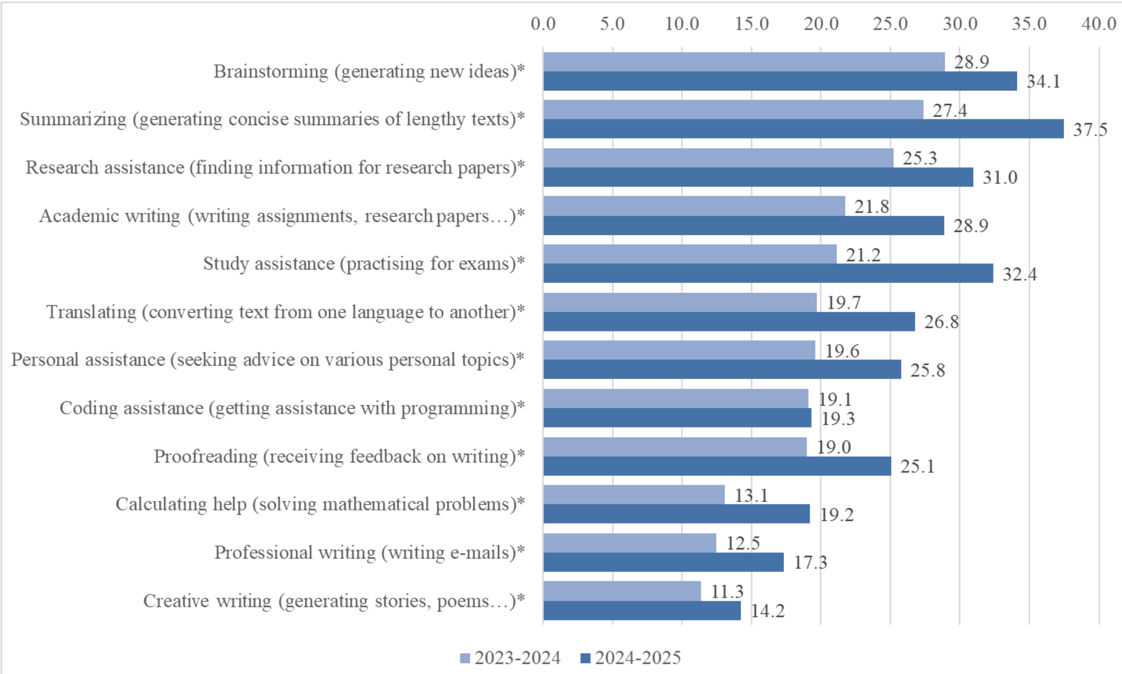
Socio-demographic characteristics	First wave (2023-2024) Number (#)	Share (%)	Second wave (2024-2025) Number (#)	Share (%)
Gender				
Male	9346	41.2	8649	38.5
Female	13365	58.8	13797	61.5
Level of study				
First	18935	83.4	17574	78.5
Second	2867	12.6	4058	18.1
Third	912	4.0	758	3.4
Field of study				
Arts and humanities	2740	12.1	2247	10.1
Social sciences	9356	41.4	9575	43.0
Applied sciences	7809	34.5	7899	35.5
Natural and life sciences	2717	12.0	2527	11.4
Mode of study				
Traditional learning	10754	47.3	9533	42.5
Online learning	2159	9.5	1735	7.8
Blended learning	9833	43.2	11150	49.7

Note: Due to incomplete responses, some socio-demographic variables do not align precisely with the total number of participants in the final sample.
Source: Authors' calculations based on the Global ChatGPT Student Survey.

Since its launch in November 2022, ChatGPT has seen growing adoption among students. During the 2023–2024 academic year, 71.4% of students reported using the tool, a figure that rose to 91.1% in the subsequent year. There was also a rise in usage intensity. While 18.0% of students indicated considerable or extensive use in the first year, this percentage increased to 27.1% in 2024–2025. Nevertheless, the frequency of use differed depending on the type of task (Figure 1). Initially, students used ChatGPT most often for brainstorming, summarizing, and research support, emphasizing its function in aiding fundamental academic activities such as idea development, comprehension, and information retrieval. Tasks like academic writing, study help, translation, personal assistance, coding, and proofreading were used moderately. In contrast, tasks involving mathematical problem-solving, professional writing, and creative writing were the least commonly reported. By 2024–2025, students had adopted ChatGPT more broadly and consistently across their academic work. The most significant increases in use were for study support (an 11.3 percentage-point [p.p.] rise) and summarizing (a 10.1 p.p. rise), indicating that students increasingly recognized ChatGPT's potential for enhancing learning efficiency and managing academic workloads.

Figure 1

Students' Usage Patterns of ChatGPT

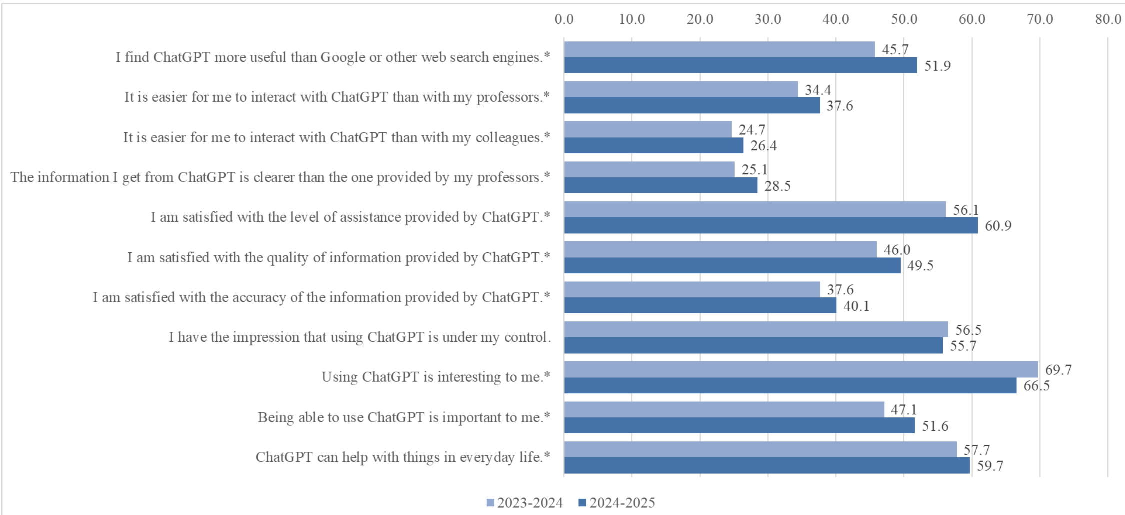


Note: An asterisk (*) indicates a statistically significant result of the t-test ($p \leq 0.05$).
Source: Authors' calculations based on the Global ChatGPT Student Survey.

In the 2023–2024 academic year, students expressed generally positive but moderate satisfaction and attitudes toward ChatGPT (Figure 2). Many found it useful compared to tools like Google or other web search engines, noted that it was relatively easy to interact with, and believed it provided clearer information than their professors. There was also a fair level of satisfaction with the quality, accuracy, and overall assistance offered by ChatGPT. Most students found using the tool interesting and considered the ability to use it important for both academic and everyday purposes. In 2024–2025, these attitudes became more favourable, with more students viewing ChatGPT as more useful than conventional search engines, reporting greater satisfaction with its assistance, and finding its explanations even clearer. The ease of interaction was noted more positively, and a growing number of students emphasized the importance of being able to use ChatGPT, suggesting it had become more integrated into their academic routines. While overall interest remained high, there was a slight drop in how interesting students found the tool (a 3.2 p.p. decrease). Comparing both years, the strongest positive changes were in the perception of ChatGPT's usefulness over traditional search engines (a 6.2 p.p. rise), satisfaction with the assistance it provides (a 4.7 p.p. rise), and the importance placed on being able to use it (a 4.5 p.p. rise), while the only decline was in students' interest in using ChatGPT (a 3.2 p.p. decrease), possibly indicating that as the tool became more normalized in students' daily academic lives, its novelty began to fade.

Figure 2

Students' Satisfaction and Attitudes toward ChatGPT



Note: An asterisk (*) indicates a statistically significant result of the t-test ($p \leq 0.05$).
Source: Authors' calculations based on the Global ChatGPT Student Survey.

The results show a clear connection between the way students use ChatGPT and their overall satisfaction and attitudes toward it. As students began to rely more heavily on the tool for tasks such as study support and summarizing, their perceptions of its usefulness and the quality of assistance it provides also improved. This growing reliance reflects a shift in how students view ChatGPT, as it moved from being seen as a novel tool to becoming a dependable part of their academic routine. The increased satisfaction with its clarity, ease of interaction, and value compared to traditional search engines suggests that students not only used ChatGPT more frequently but also recognized its role in helping them manage academic challenges more effectively. Although interest in the tool declined slightly, this likely reflects its integration into everyday academic life rather than a decrease in its perceived value.

Conclusion

This two-year global study provides one of the most comprehensive insights to date into the evolving relationship between students and ChatGPT in higher education. The findings clearly indicate that the adoption and integration of generative artificial intelligence tools, particularly ChatGPT, have increased significantly between the 2023–2024 and 2024–2025 academic years. Not only did usage rates rise sharply, but students also expanded the scope of tasks for which they used the tool, increasingly relying on it for academic support functions such as summarization, study assistance, and information retrieval (Boubker, 2024; Pallivathukal et al., 2024). These trends reflect a broader shift toward normalization and mainstream acceptance of artificial intelligence–driven support in educational routines (Tlili et al., 2023; Ajlouni et al., 2023).

Furthermore, students' satisfaction and attitudes toward ChatGPT have generally improved over time. The tool is perceived as more useful than traditional search engines, easy to interact with, and capable of providing clear, personalized support (Park, 2023; Ravšelj, Keržič, et al., 2025). The rising importance students place on being able to use ChatGPT highlights a growing expectation that proficiency in such tools is becoming an essential academic and life skill (Michalon & Camacho-Zuñiga, 2023). Interestingly, the slight decline in perceived novelty suggests that ChatGPT is moving from an emerging innovation to an embedded element of everyday student life.

However, several limitations of the study must be acknowledged. First, the use of a convenience sampling strategy may limit the generalizability of the results, as it does not fully represent the diversity of the global student population (Sarstedt et al., 2018). Second, although the survey was translated into multiple languages to encourage broad participation, linguistic nuances may have affected how some questions were interpreted by respondents. Third, the reliance on self-reported data introduces potential biases, such as recall inaccuracy or social desirability effects. Fourth, the study lacks advanced statistical analyses, which may limit the causal interpretability and robustness of the findings. Finally, given the rapid pace at which generative artificial intelligence technologies evolve (Dong et al., 2024), some findings may become outdated quickly, emphasizing the need for ongoing, longitudinal research in this area.

Nonetheless, the study offers robust empirical evidence that generative artificial intelligence is reshaping the student experience in higher education. As artificial intelligence tools become more sophisticated and accessible, understanding their pedagogical implications will be essential for educators, policymakers, and students alike. Future research should explore not only how students use these tools but also how their learning processes, cognitive engagement, and academic achievements are influenced over time.

Acknowledgements

The authors acknowledge the financial support of the Slovenian Research and Innovation Agency (research program No. P5-0093, start-up research program "Generative artificial intelligence in higher education for the future labour market", project No. J5-60095 and project No. Z5-4569). During manuscript

preparation, limited use of ChatGPT (version 4o, OpenAI) was made to support language editing. All intellectual content, analysis, and conclusions are the original work of the authors.

References

Ait Baha, T., El Hajji, M., Es-Saady, Y., & Fadili, H. (2024). The impact of educational chatbot on student learning experience. *Education and Information Technologies*, *29*(8), 10153-10176. https://doi.org/10.1007/s10639-023-12166-w

Ajlouni, A. O., Wahba, F. A.-A., & Almahaireh, A. S. (2023). Students' attitudes towards using ChatGPT as a learning tool: The case of the University of Jordan. *International Journal of Interactive Mobile Technologies, 17*(18), 99–117. https://doi.org/10.3991/ijim.v17i18.41753

Alessandri-Bonetti, M., Liu, H. Y., Giorgino, R., Nguyen, V. T., & Egro, F. M. (2024). The first months of life of ChatGPT and its impact in healthcare: A bibliometric analysis of the current literature. *Annals of Biomedical Engineering*, *52*, 1107-1110. https://doi.org/10.1007/s10439-023-03325-8

Aristovnik, A., Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., Kessler, S. H., Schäfer, M., et al. (2025). *Higher education students' evolving perceptions of ChatGPT: Global survey data from the academic year 2024-2025* (Version 2) [Data set]. Mendeley Data. https://doi.org/10.17632/nv2343nwsb.2

Aristovnik, A., Umek, L., Brezovar, N., Keržič, D., & Ravšelj, D. (2024). The role of ChatGPT in higher education: Some reflections from public administration students. In S. K. S. Cheung, F. L. Wang, N. Paoprasert, P. Charnsethikul, K. C. Li, & K. Phusavat (Eds.), *Technology in education: Innovative practices for the new normal (ICTE 2023)*. *Communications in Computer and Information Science* (Vol. 1974). Springer. https://doi.org/10.1007/978-981-99-8255-4_22

Boubker, O. (2024). From chatting to self-educating: Can AI tools boost student learning outcomes? *Expert Systems with Applications*, *238*(Part A), 121820. https://doi.org/10.1016/j.eswa.2023.121820

Das, S. R., & Madhusudan, J. V. (2024). Perceptions of higher education students towards ChatGPT usage. *International Journal of Technology in Education*, *7*(1), 86-106. https://doi.org/10.46328/ijte.583

Dong, M. M., Stratopoulos, T. C., & Wang, V. X. (2024). A scoping review of ChatGPT research in accounting and finance. *International Journal of Accounting Information Systems*, *55*, 100715. https://doi.org/10.1016/j.accinf.2024.100715

Michalon, B., & Camacho-Zuñiga, C. (2023). ChatGPT, a brand-new tool to strengthen timeless competencies. *Frontiers in Education, 8*, 1251163. https://doi.org/10.3389/feduc.2023.1251163

Mohmad, P. (2023, January 19). *How does ChatGPT become popular so quickly and how is it growing*. *Analytics Insight*. https://www.analyticsinsight.net/how-does-chatgpt-become-popular-so-quickly-and-how-is-it-growing/

Pallivathukal, R. G., Soe, H. H. K., Donald, P. M., Samson, R. S., & Ismail, A. R. H. (2024). ChatGPT for academic purposes: Survey among undergraduate healthcare students in Malaysia. *Cureus, 16*(1), e53032. https://doi.org/10.7759/cureus.53032

Park, J. (2023). Medical students' patterns of using ChatGPT as a feedback tool and perceptions of ChatGPT in a leadership and communication course in Korea: A cross-sectional study. *Journal of Educational Evaluation for Health Professions, 20*, 29. https://doi.org/10.3352/jeehp.2023.20.29

Rasch, D., Teuscher, F., & Guiard, V. (2007). How robust are tests for two independent samples? *Journal of Statistical Planning and Inference, 137*(8), 2706–2720. https://doi.org/10.1016/j.jspi.2006.04.011

Ravšelj, D., Aristovnik, A., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N. Alahad, N., Abdulla, A. A., Akopyan, A., Segura, M. W. A., AlHumaid, J., Allam, M. F., Alló, M., Andoh, R. P. K., Andronic, O., Arthur, Y. D., Aydın, F., Badran, A., Balbontín-Alvarado, R.,... Zlatkin-Troitschanskaia, O. (2025). *Higher education students' early perceptions of ChatGPT: Global survey data* (Version 2) [Data set]. Mendeley Data. https://doi.org/10.17632/ymg9nsn6kn.2

Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., A. Iahad, N., Abdulla, A. A., Akopyan, A., Segura M. W. A., AlHumaid, J., Allam, M. F., Alló, M., Andoh, R. P. K., Andronic, O., Arthur, Y. D., Aydin, F., Badran, A., Balbontín-Alvarado, R., Saad, H. B., ...Aristovnik, A. (2025). Higher education students' perceptions of ChatGPT: A global study of early reactions. *PloS one*, *20*(2), e0315011. https://doi.org/10.1371/journal.pone.0315011

Sarstedt, M., Bengart, P., Shaltoni, A. M., & Lehmann, S. (2018). The use of sampling methods in advertising research: A gap between theory and practice. *International Journal of Advertising*, *37*(4), 650-663. https://doi.org/10.1080/02650487.2017.1348329

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*, 15. https://doi.org/10.1186/s40561-023-00237-x

Twinomurinzi, H., & Gumbo, S. (2023). ChatGPT in scholarly discourse: Sentiments and an inflection point. In A. Gerber & M. Coetzee (Eds.),*:* *Proceedings of the South African Institute of Computer Scientists and Information Technologists (SAICSIT 2023). Communications in Computer and Information Science, Vol. 1878* (pp. 230–245). Springer. https://doi.org/10.1007/978-3-031-39652-6_17

Williams, R. T. (2024). The ethical implications of using generative chatbots in higher education. *Frontiers in Education, 8*, 1331607. https://doi.org/10.3389/feduc.2023.1331607

Author Details

Aleksander Aristovnik
Faculty of Public Administration
University of Ljubljana
SLOVENIA
aleksander.aristovnik@fu.uni-lj.si

Dejan Ravšelj
Faculty of Public Administration
University of Ljubljana
SLOVENIA
dejan.ravselj@fu.uni-lj.si

GENERATIVE ARTIFICIAL INTELLIGENCE IN PRIMARY AND SECONDARY EDUCATION IN PORTUGAL: ACCEPTANCE AND USE BY STUDENTS

Nuno Dorotea and Célia Ribeiras
UIDEF – Unidade de Investigação e Desenvolvimento em Educação e Formação
Institute of Education, University of Lisbon
PORTUGAL

Abstract

Generative Artificial Intelligence (GenAI) has penetrated the world globally, prompting several structural changes and new ways of dealing with knowledge. Major challenges are emerging in education. This research aims to analyse the factors that influence the use and acceptance of GenAI by primary and secondary school students in Portugal through UTAUT2 (Unified Theory of Acceptance and Use of Technology). With 478 participants, the data was collected in 2024 and analysed using the partial least squares method. Results indicate that Habit emerged as the most influential factor on Behavioural Intention, followed by Performance Expectancy, Hedonic Motivation, and Personal Innovation. Habit and Behavioural Intention demonstrated significant impact on Behavioural Intention.

Introduction

Generative Artificial Intelligence (GenAI) is a consequence of technological evolution and of the human desire to surpass their own limits, creating systems that reproduce intelligent behaviour in an artificial way (Oliveira, 2019, p. 2). Like any other technology, its presence in education is inevitable. Since November 2022, with the public release of ChatGPT, this topic has flooded the global educational landscape. On the one hand, Liu et al. (2023, p. 73) points out that GenAI "…can improve the learning process and experience for students"; on the other hand, its constant and ongoing emergence and development demands "more research (…) to determine its effectiveness in different contexts" (Su & Yang, 2023, p. 362) and requires understanding their capabilities and limitations. This is a new, emerging, and overwhelming technology, creative and generative, that, when it appears, marks only the beginning of something much greater, in constant growth, and whose impact on education is not yet fully understood. What is clear is that AI will have significant consequences for education. As Holmes, Bialik, and Fadel (2019) point out, "however, while many assume that artificial intelligence in education means students being taught by robot teachers, the reality is more prosaic yet still has the potential to be transformative. Nonetheless, the application of AI to education raises far-reaching questions" (p. 80). The interaction between Generative AI and education extends beyond the classroom to include teaching about AI and preparing for Human-AI collaboration. The introduction of AI into education raises questions about pedagogy, access, ethics, equity, and sustainability, and calls for a continuous reassessment of the foundational principles of education. The pedagogical advantages of using generative AI in education are numerous. Liu et al. (2023) argue that GenAI technologies, "together with other forms of AI, can enhance the learning process and experience for students due to their ability to access and generate information" (p. 73).

However, as an emerging technology, there is still much to learn, identify, and explore, especially due to the challenges it presents, such as the errors and inaccuracies it may produce, the biases in its results, not only due to the algorithms used but also the data employed for machine learning, as well as the so-called "hallucinations" (Adiguzel et al. 2023; Su & Yang, 2023; Sullivan et al., 2023; Tlili et al., 2023; Liu et al., 2023), and the ethical, moral, and legal issues associated with it. In order to better understand how students in non-higher education perceive, accept, and use Generative AI, it is crucial to investigate whether studies exist on the acceptance and use of this technology in primary and secondary education, not only in Portugal but also abroad. And ultimately, if no alternatives are available, at other levels of education. Based on these premises, the present study aims to identify and analyse the factors that influence the acceptance and use of Generative AI in academic contexts by students in primary and secondary education in Portugal.

Methodology

To analyse the factors influencing the adoption and use of Generative AI, a literature review was conducted to identify the most appropriate theoretical model for measuring levels of technology acceptance.

Theories of technology acceptance and use have emerged over many years of research, resulting in different models with similar purposes. Some of the most used models include the TAM (Technology Acceptance Model), the TPB (Theory of Planned Behaviour), the MM (Motivation Model), the Model of PC Utilization (MPCU), the IDT (Innovation Diffusion Theory), and the SCT (Social Cognition Theory. All these models contributed to the construction of the UTAUT model (Unified Theory of Acceptance and Use of Technology) and, consequently, to the UTAUT2 model. Venkatesh et al. (2003) selected the constructs and respective theories they considered most effective in identifying the factors that most impact technology acceptance and use "both in organizational and non-organizational contexts" (Venkatesh et al., 2012). For the present study, the UTAUT2 model was adopted, which is an extension of the UTAUT model developed by Venkatesh, Thong, and Xu (2012).

We considered the following constructs from UTAUT2: Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), Facilitating Conditions (FC), Hedonic Motivation (HM), Habit (HB), Personal Innovation (PI), Confidence (CO), and Perceived Risk (PR). The aim was to identify the impact of these factors on the constructs Behavioural Intention (BI) and Frequency of Use (FU). Additionally, two moderating variables were considered: Gender and School Level.

Performance Expectancy (PE) considers the extent to which individuals believe they can improve their performance by using a given technology.

Effort Expectancy (EE) refers to the ease of using a technology, that is, the level of effort required to use the technology.

Social Influence (SI) refers to the influence that other people (whether relevant or not to the individual) have on an individual's use of a given technology. This construct is a key determinant of Behavioural Intention, as it is known that an individual's behaviour is influenced by how they believe others will perceive them as a result of using the technology (Venkatesh et al., 2003).

Facilitating Conditions (FC) reflect the degree to which an individual believes there is support for using a technology. Directly linked to the technological and organizational environment surrounding individuals, this support should contribute to solving problems that may arise and consequently remove barriers.

Hedonic Motivation (HM) is synonymous with pleasurable feelings: "the fun or pleasure derived from using a technology, and it has been shown to play an important role in determining technology acceptance and use" (Venkatesh et al., 2012, p. 161).

Habit (HB) aims to determine the extent to which individuals behave automatically when handling a technology. This automatism is directly linked to the learning individuals have acquired or developed through the use of that technology.

Personal Innovation (PI) "refers to an individual's willingness and ability to adopt and use new technology in their daily life" (Strzelecki, 2023, p. 4). According to the same author, this construct is an essential addition to the UTAUT2 model, generally defined as the level of willingness to embrace new technologies while simultaneously demonstrating comfort and confidence in handling them.

Confidence (CO) in technology refers to "the users' belief that the use of technology is reliable and trustworthy" (Al-Azawei & Alowayr, 2020, p. 5). In other words, it concerns confidence in the outputs of Generative AI, and their credibility, which should not be “blind” but rather measured, informed, and cautious.

Perceived Risk (PR) is directly related to security, knowledge about dangers and issues associated with Generative AI, data protection and privacy, ethical concerns, as well as misinformation and biases/prejudices exhibited by these applications. These are linked to overconfidence, which can result in a lack of critical thinking and creativity. These challenges are so widespread that there exists, both within and beyond education, a sense of distrust, threat, and discomfort regarding Generative AI. In other words, this construct determines an individual’s assessment of the potential risks or uncertainties of a given situation, which in this case is the use of Generative AI. According to Yao et al. (2024), “previous research has shown that risk perception plays a crucial role in shaping individuals’ attitudes and intentions towards adopting new technologies” (p. 6).

Behavioural Intention (BI) refers to the “likelihood or subjective intention of an individual to use a particular technology in the future” (Venkatesh et al., 2012, cited in Strzelecki, 2023, p. 4).

Frequency of Use (FU) refers to how often an individual uses a particular technology.

In all constructs, the concept of technology was adapted to Generative AI (GenAI). In total, 49 items were considered, aiming to identify the factors that contribute to the adoption and frequency of use of Generative AI (GenAI) by students in primary and secondary education. Based on these constructs, 13 hypotheses were developed to demonstrate the relationships among them.

Validation of the Questionnaire

The quality analysis of the questionnaire was conducted for all constructs, except for Frequency of Use, as it contained only a single item. Table 1 presents the results of the reliability analysis of the different constructs in the questionnaire, as well as of the instrument as a whole.

According to George and Mallery (2003), the Cronbach’s Alpha value for each construct should be above 0.7 to ensure that the internal consistency of the data is acceptable, i.e., to ensure data reliability. As shown in Table 1, the Cronbach’s alpha values for all constructs are above 0.8, with most being very close to or above 0.9, indicating that the questionnaire demonstrates near-excellent internal consistency. The overall reliability of the scale in this study is 0.952. Similarly, the composite reliability values are very close to or above 0.9.

Table 1
Measurement Model Indicators

Constructs	Number of items	Average	Standard Deviation	Cronbach's Alpha (α)	Composite reliability	Average Variance Extracted (AVE)
Facilitating Conditions (FC)	5	4.91	1.27	.808	.880	.602
Confidence (CO)	4	4.20	1.52	.892	.924	.753
Performance Expectancy (PE)	6	4.95	1.42	.934	.948	.752
Effort Expectancy (EE)	4	5.27	1.36	.903	.932	.774
Habit (HB)	4	3.69	1.66	.897	.928	.765
Behavioural Intention (BI)	3	4.60	1.61	.912	.944	.850
Personal Innovation (PI)	4	4.16	1.54	.879	.918	.738
Social Influence (SI)	8	3.94	1.41	.928	.941	.671
Hedonic Motivation (HM)	4	4.75	1.44	.917	.942	.804
Perceived Risk (PR)	6	4.80	1.28	.871	.898	.598
Global	**48**	**4.51**	**.94**	**.952**	**-**	**-**

The various constructs were also analysed in terms of convergent validity and discriminant validity.

The Average Variance Extracted (AVE) is a metric used to assess the convergent validity of a construct in Structural Equation Modeling (SEM). It helps measure how much of the variance in a set of indicators is explained by the construct they are intended to measure. Thus, it ensures that the constructs in a model are adequately represented by the indicators and helps to ensure the robustness of the analyses based on that model.

For a construct to demonstrate convergent validity, an AVE value greater than 0.50 is a mandatory requirement (Hair et al., 2014; Henseler et al., 2009). This indicates that more than 50% of the variance in the indicators is explained by the construct. Analysing the AVE values for each construct (Table 1), it is observed that all are above 0.598, which suggests that the constructs are capable of explaining at least 60% of the variance they are intended to represent.

Discriminant validity was also assessed using the Fornell and Larcker (1981) criteria, including the Heterotrait-Monotrait ratio (HTMT), and cross-loadings.

According to Fornell and Larcker (1981) and Hair et al. (2014), discriminant validity is assessed using the Fornell-Larcker criterion, which is confirmed when the square roots of the AVE values are greater than the correlations between constructs (i.e., the bold value for each construct must be higher than all the values in the intersection of that construct with the others). In this study, this requirement is met. On the other hand, it is important to analyse the cross-loadings, whose ideal values require that each item loads more highly on the construct to which it is theoretically linked than on any other construct. The Heterotrait-Monotrait Ratio (HTMT) is another metric, considered more precise for assessing discriminant validity in structural equation models (SEM). Discriminant validity refers to the extent to which a construct is truly distinct from other constructs in the model.

A value below 1.0 generally indicates good discriminant validity, suggesting that the constructs are distinct from one another. HTMT values below 0.85 are acceptable and suggest adequate discriminant validity.

Based on the analysis of the model HTMT values (below 0.83), it can be concluded that the model meets the criteria for convergent and discriminant validity, thereby ensuring the consistency of its structure and the reliability of subsequent statistical inferences.

Since all conditions for all measured constructs were met, it can be concluded that these constructs are suitable for estimating their impacts on Behavioural Intention and Frequency of Use of Generative AI applications.

Participants

A total of 478 students from Primary and Secondary Education participated in the study, with an average age of 15 years. Of these, 50.4% were male and 49.6% were female. The data were collected between January and May 2024 and subsequently analysed statistically using the Smart PLS-SEM software.

Results

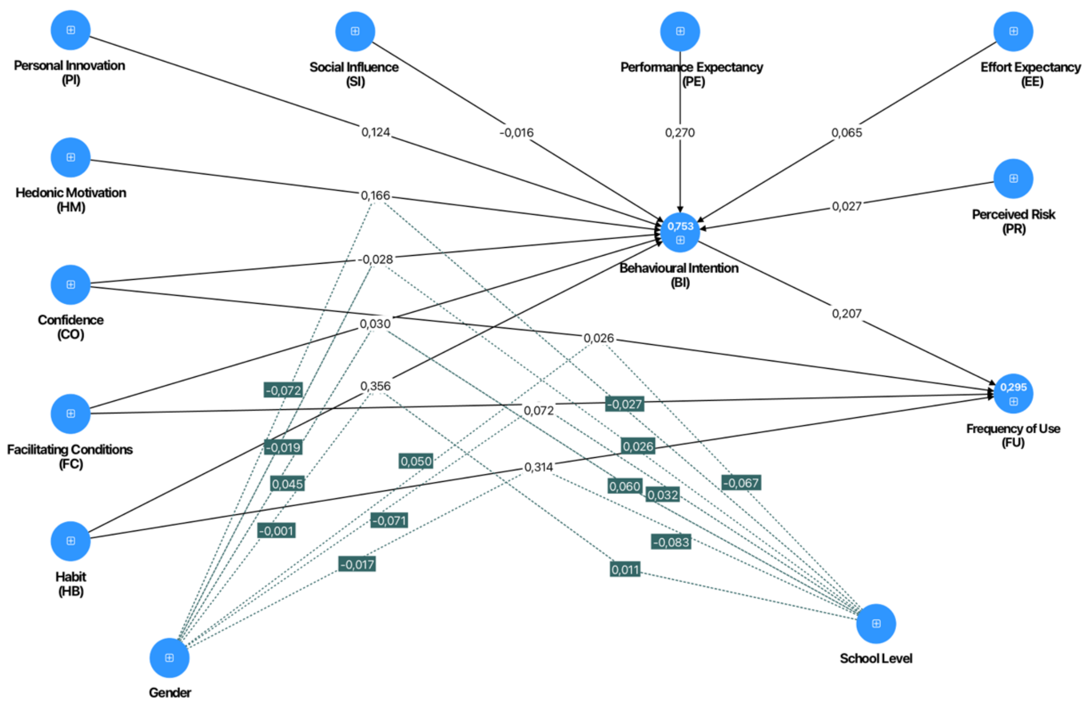
Partial Least Squares Structural Equation Modelling (PLS-SEM) provides diagrammatic representations that visually illustrate hypotheses and relationships between constructs (Hair et al., 2021). In these models, constructs, or latent variables, are depicted as circles or ovals. The relationships between constructs, and between constructs and indicators, are represented by unidirectional arrows, suggesting predictive or, when supported by robust theory, causal relationships. PLS-SEM comprises two main components: the structural model (or inner model), which connects the constructs and displays the relationships among them, and the measurement models (or outer models), which show the relationships between constructs and their indicators.

To estimate the model, we used the PLS-SEM algorithm with the path weighting scheme through the SmartPLS4 software (Version 4.1.0.8), running 5,000 bootstrap samples to determine the statistical significance of the PLS-SEM results, as recommended by Ringle et al., (2022).

Below, we present the results of the relationships between constructs and their influence on Behavioural Intention and Frequency of Use.

Figure 1

Results of the GenAI Acceptance and Utilisation Model: Structural Model of Student Acceptance and Use of AI Defined for the Study



The coefficient of determination R^2 is used to determine the explanatory power of each construct and of the overall model. Ranging between 0 and 1, higher R^2 values indicate greater explanatory power. According to Hair et al. (2021), R^2 values of 0.25 are considered weak, 0.50 moderate, and 0.75 substantial. Figure 1 presents the results of the PLS-SEM analysis, indicating the relationships between the constructs and the R^2 values, which are displayed inside the circles. As we can see, 75.3% of the variance in Behavioural Intention (BI) can be explained by the other constructs, and only 29.5% of Frequency of Use (FU) can be explained by

Behavioural Intention (BI), Confidence (CO), Facilitating Conditions (FC) and Habit (HB).

According to Sarstedt, Ringle, and Hair (2022), f^2 (f-squared) determines the effect size of a construct. Values around 0.35 correspond to large effects, 0.15 to medium effects, and 0.02 to small effects. Values of f^2 below 0.02 suggest no effect. Table 2 shows the effect size of the hypotheses confirmed for this study.

Table 2

Path Coefficients and Significance Test Results

Hypo-thesis	Relationships	*Path Coefficients*	*P values*	f^2	Confirmed
H1	(PE) Performance Expectancy -> (BI) Behavioural Intention	.270	**.000**	.114 ++	**Yes**
H2	(EE) Effort Expectancy -> (BI) Behavioural Intention	.065	.097	.008	No
H3	(SI) Social Influence -> (BI) Behavioural Intention	-.016	.570	.001	No
H4	(FC) Facilitating Conditions -> (BI) Behavioural Intention	.030	.404	.002	No
H5	(FC) Facilitating Conditions -> (FU) Frequency of Use	.072	.096	.005	No
H6	(HM) Hedonic Motivation -> (BI) Behavioural Intention	.166	**.001**	.039 +	**Yes**
H7	(HB) Habit -> (BI) Behavioural Intention	.356	**.000**	.180 ++	**Yes**
H8	(HB) Habit -> (FU) Frequency of use	.314	**.000**	.047 +	**Yes**
H9	(BI) Behavioural Intention -> (FU) Frequency of use	.207	**.003**	.019	**Yes**
H10	(PI) Personal Innovation -> (BI) Behavioural Intention	.124	**.013**	.022 +	**Yes**
H11	(CO) Confidence -> (BI) Behavioural Intention	-.028	.395	.002	No
H12	(CO) Confidence -> (FU) Frequency of Use	.026	.816	.000	No
H13	(PR) Perceived Risk -> (BI) Behavioural Intention	.027	.406	.002	No

Note*: (+) f 2 > .02 = low effect; (++) f 2 > 0.15 medium effect (Sarstedt et al., 2022)*

The values of the relationships between the different constructs are also shown in Table 2, under the Path Coefficients (pc) column. A Path Coefficient closer to +1 indicates a strong positive relationship (as one construct increases, so does the other). A path coefficient closer to -1 indicates a strong negative relationship (as one construct increases, the other decreases). A Path Coefficient of 0 means there is no relationship between constructs.

The analysis of these internal relationships between the model's constructs, which allows identifying its capacity to predict "Behavioural Intention" and "Frequency of Use", suggests that the strongest predictors of "Behavioural Intention" in descending order, are "Habit" (pc = .356; p = .000), "Performance Expectancy" (pc = .270; p = .000), "Hedonic Motivation" (pc = .166; p = .001), and "Personal Innovation" (pc = .124; p = .013), which together explain 75.7% of the variance in "Behavioural Intention".

Regarding "Behavioural Intention," positive effects were also observed for "Effort Expectancy" (pc = .065; p = .074), "Facilitating Conditions" (pc = .030; p = .404), and "Perceived Risk" (pc = .027; p = .406), but the effect size (f^2) of these relationships is not significant (< .02).

Concerning the predictors of "Frequency of Use" the results suggest that the strongest predictor is "Habit" (pc = .314; p = .000), followed by "Behavioural Intention" (pc = .207; p = .003). These constructs account for 29.1% of the variance in "Frequency of Use". Positive effects were also observed for "Facilitating Conditions" (pc = .072; p = .096) and "Confidence" (pc = .026; p = .011), but again, the effect size (f^2) of these relationships is not significant (< .02).

Thus, of the 13 hypotheses defined for the study, only H1, H6, H7, H8, H9, and H10 are confirmed, as they show statistical significance at the 5% level. The remaining hypotheses are not accepted. Based on the statistical analysis results, the study corroborated six of the thirteen initially proposed hypotheses.

These confirmed hypotheses highlight the influence of Performance Expectancy (PE), Hedonic Motivation (HM), Habit (HB), and Personal Innovation (PI) on Behavioural Intention (BI), as well as the influence of Behavioural Intention (BI) and Habit (HB) on Frequency of Use (FU). Notably, Habit showed a medium influence (pc = 0.356, $p < .001$, $f^2 = 0.180$) and Performance Expectancy (PE) also had a medium effect (pc = 0.270, $p < .001$, $f^2 = 0.114$) on Behavioural Intention ($p < 0.05$).

Conversely, hypotheses H2, H3, H4, H5, H11, H12, and H13 did not receive sufficient statistical support to be accepted in the context of this investigation.
The model presented in the study also considers the moderating effects of "Gender" and "School Level". The results show that the moderating variable "School Level" had a significant impact on Frequency of Use (pc = .107; p = .015) and influenced the relationship between Facilitating Conditions and Behavioural Intention (pc = .060; p = .031). On the other hand, the moderating variable "Gender" had no significant impact on the tested relationships between the predictors and the dependent variables.

Conclusions

Digital technologies are now inseparable from analogue ones in all aspects of human life, including education. The use of technologies in education has been a growing and widely discussed topic over the past 30 years. Artificial Intelligence, more specifically its generative aspect, represents a further step in this evolution, especially since November 2022, when ChatGPT was made available to the public. Regarding the influence of the various dimensions on Behavioural Intention and Frequency of Use, Habit directly and significantly impacts the former, making it the primary factor influencing the intention to use Generative AI (GenAI). Other important factors include the expectation that GenAI can contribute to better performance, both in terms of time efficiency and success, the enjoyment derived from using these technologies, and the feeling among students that they are learning something new. Simultaneously, both Habit and Behavioural Intention have a positive direct impact on Frequency of Use. In other words, the stronger the habit and the greater the intention to use GenAI, the higher the frequency of its use.

Among the moderating variables, only the effect of the School Level on Frequency of Use and on the relationship between Facilitating Conditions and Behavioural Intention was confirmed. These were the 6 hypotheses supported out of the 13 initially proposed.

The results obtained can thus provide valuable contributions to the understanding of the adoption and use of GenAI in the context of primary and secondary education, as well as support the development of educational strategies that effectively integrate this technology.

References

Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, *15*(3), ep429. https://doi.org/10.30935/cedtech/13152

Al-Azawei, A., & Alowayr, A. (2020). Predicting the intention to use and hedonic motivation for mobile learning: A comparative study in two Middle Eastern countries. *Technology in Society, 62*, 101325. https://doi.org/10.1016/j.techsoc.2020.101325

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. https://doi.org/10.2307/3151312

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference* (4th ed.). Pearson Higher Education.

Hair, J. F., Sarstedt, M., Hopkins, L. & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *European Business Review, 26*(2), 106–121.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). *Partial least squares structural equation modeling (PLS-SEM) using R. Classroom Companion: Business*. Springer International Publishing. https://doi.org/10.1007/978-3-030-80519-7

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *New Challenges in International Marketing (Advances in International Marketing, Vol. 20*), 277-319. Emerald Group Publishing Limited.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in education: Promises and implications for teaching & learning.* Center for Curriculum Redesign.

Liu, B. L., Morales, D., Roser Chinchilla, J. F., Sabzalieva, E, Valentini, A., Vieira do Nascimento, D., & Yerovi, C. (2023). *Harnessing the era of artificial intelligence in higher education: A primer for higher education stakeholders.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386670

Oliveira, A. (2019). *Inteligência Artificial*. Lisboa: Fundação Francisco Manuel dos Santos.

Ringle, C. M., Wende, S., & Becker, J. M. (2022). *SmartPLS 4.* SmartPLS. https://www.smartpls.com/

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2022). Partial least squares structural equation modeling. In C. Homburg, M. Klarmann, & A. Vomberg (Eds.), *Handbook of market research* (pp. 587–632). Springer. https://doi.org/10.1007/978-3-319-57413-4_15

Strzelecki, A. (2023). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. Interactive Learning Environments, 1–14. https://doi.org/10.1080/10494820.2023.2209881

Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education, 6*(3), 355-366. https://doi.org/10.1177/20965311231168423

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching, 6*(1), 31-40. https://doi.org/10.37074/jalt.2023.6.1.17

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*, 15. https://doi.org/10.1186/s40561-023-00237-x

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://www.jstor.org/stable/30036540

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly, 36*(1), 157–178. https://doi.org/10.2307/41410412

Yao, E., Guo, D., Liu, S., & Zhang, J. (2024). The role of technology belief, perceived risk and initial trust in users' acceptance of urban air mobility: An empirical case in China. *Multimodal Transportation, 3*(4), 100169. https://doi.org/10.1016/j.multra.2024.100169

Author Details

Nuno Dorotea
UIDEF, Unidade de Investigação e Desenvolvimento em Educação e Formação.
Institute of Education, University of Lisbon
nmdorotea@ie.ulisboa.pt

Célia Ribeiras
Institute of Education, University of Lisbon
celiaribeiras@edu.ulisboa.pt

This work was supported by National Funds through FCT-Portuguese Foundation for Science and Technology, I.P., under the scope of UIDEF - Unidade de Investigação e Desenvolvimento em Educação e Formação, UIDB/04107/2020, https://doi.org/10.54499/UIDB/04107/2020

ARTIFICIAL INTELLIGENCE (AI) AND SCHOOL LEADERSHIP — SCHOOL LEADERS' REFLECTIONS ON PROFESSIONAL USE OF AI

Marcia Håkansson Lindqvist
Department of Education, Mid Sweden University
SWEDEN

Fanny Pettersson
Department of Education, Umeå University
SWEDEN

Abstract

Artificial intelligence (AI) continues to influence all aspects of school and education. Although AI is acknowledged for introducing both opportunities and constraints for school leadership, there remains limited research on how school leaders use and understand AI in schools. This paper examines how school leaders perceive and engage with the integration of AI for their learning processes and leadership practices. School leaders' professional use, as well as how they model the use of AI in their educational settings, may in turn support teachers' and students' application of AI for teaching and learning, and it may benefit the school as an organization.

Introduction

Artificial intelligence (AI) has had an ongoing influence on all aspects of school and education, including teaching, learning, and organization (Karakose & Tülübas, 2024). Thus, in recent years, the debate in AI in education (AIED) on whether to use AI (i.e., ChatGPT) in education has shifted towards the need for further research on how AI tools can be critically and effectively adopted and used for teaching, learning, and leadership (Strzelecki, 2023). Research has shown that although AI can serve as a teaching and learning tool to stimulate reflection, provide ideas, assist in assessments, and correct language, it also poses risks such as lack of control, cheating, decreased creativity, and development of academic dishonesty (Neumann et al., 2023).

Early research on AI appears to provide a picture of opportunities and constraints for teachers and students (Neumann et al., 2023; Rudolph et al., 2023). However, research on the impacts of AI on school leadership has been "extremely limited," and according to several studies, only "early evidence" has been assessed on how AI influences leadership in school (Duran & Ermiş, 2024; Fullan et al., 2023; Wang, 2021). Simultaneously, Fullan et al. (2023) concluded that the limited research that exists "suggests that such technology can help educational leaders perform routine, mechanical tasks, thus allowing them to focus on other more productive and creative issues that demand their human skills and their social intelligence" (Fullan et al., 2023, p. 342).

The aim of this paper is to explore and analyse how school leaders perceive and engage with the integration of AI for their learning processes and leadership practices. The following research question was posed: 1) How do school leaders perceive opportunities and constraints related to their professional use of AI in learning and leadership practices?

Background

In recent years, advancements in AI have had a direct impact on all levels of education (Tyson & Sauers, 2021). Fullan et al. (2023) described AI in terms of its "enormous potential to improve learning, teaching, pedagogical innovations, assessment, and educational administration through intelligent tutoring systems, chatbots, robots, learning analytics dashboards, adaptive learning systems and automated assessment" (p. 340). Neumann et al. (2023) discussed this in terms of challenges and opportunities in education. Challenges involve the limited knowledge of how students utilize AI, uncertainties in evaluating AI in school, varying perceptions of acceptable use, increased time demands for assessments, and the unknown potential of AI. Conversely, opportunities lie in enhanced student support, fostering creativity, and potential for driving educational innovation. This means that AI introduces a wide range of ethical, moral, and practical challenges for all actors in school (Strzelecki, 2023).

However, there is limited knowledge regarding school leaders and their use of AI in educational contexts. Fullan et al. (2023) argued that AI will fundamentally reshape both the perception and execution of leadership in educational contexts. Current research shows that AI has the potential to reshape school leadership by automating and streamlining administrative processes, providing advanced data analysis, supporting student learning strategies, and optimizing communication with parents, teachers, students, and the broader educational community (Dogan & Arslan, 2025). Thus, these capabilities can enhance efficiency, facilitate informed decision-making, and allow school leaders to focus on strategic and pedagogical priorities. AI has been associated with both increased and decreased workload, alterations of teaching profession, and the powerful processing abilities. This work involves several perspectives. Policies and easy-to-understand guidelines are needed for the use of language models in learning and teaching, proper use of these tools, and the consequences for cheating (Rudolph et al., 2023).

To meet these opportunities and constraints, professional development is needed. For example, educators may need training to instruct students on academic integrity and educators on how to critically evaluate any resources and adapt the use of AI to their specific context (Rudolph et al., 2023). Other studies have pointed to the importance of expertise, experience, and understanding of students' use and the opportunities and constraints that the use of AI encompasses for them (Cooper, 2023). To support all these aspects, it will be necessary to improve administration and professional development (Hutami, 2024). In summary, according to Van Quaquebeke and Gerpott (2023), "The question is not anymore whether AI will play a role in leadership, the question is whether we will still play a role. And if so, what role that might be. It is high time to start that debate" (p. 272).

Method

This paper focuses on school leaders' reflections on professional use of AI. As part of this explorative pilot study, school leaders were asked to discuss how they perceive and engage with the integration of AI for their learning processes and leadership practices. The data on school leaders' written reflections were gathered from learning reflections ($N = 15$). With inspiration from Moon's (2006) notion of learning journals, the school leaders were asked to elaborate on their professional use of AI through learning journals. The learning reflections were written in the spring of 2024 by school leaders who were in middle of the Swedish National School Leader Programme. The school leaders represented all levels of school, from preschool to upper secondary school. The school leaders had time to reflect on this question for a brief period during a lesson regarding AI and digital technologies for leading, teaching, and learning. The learning reflections were short texts. Using reflexive thematic analysis (Braun & Clarke, 2019), the learning reflections were analysed. This involved reading and rereading in a reflective approach as well as determining which themes emerged in reflection. The school leaders' reflections are identified as "School leader" (SL 1–15).

Findings

The findings are presented in this section. First, the theme *Professional Support* is presented. Thereafter, *Opportunities* and *Constraints* are presented.

Professional Support

In the analysis, the following themes regarding AI as professional support emerged: *Own learning, Source of knowledge,* and *Administrative tasks.*

In the category *Own learning*, school leaders saw opportunities when using AI as a pedagogical tool to support their own learning. For many of the school leaders, this was evident in their ongoing training as school leaders: “I have used it to help me understand some books during my principal training. I think this can be linked to my learning as a head teacher” (SL10). For another school leader, AI was a source of support regarding the course literature: “I find some course literature difficult to interpret as it sometimes does not have a clear connection to subjects I am used to studying; here I think AI can be a help” (SL12).

AI as a *source of knowledge* was also noted in the school leaders’ reflections. As one school leader noted:

> AI can be used as a quick tool in my principal role when I may be looking for answers to questions and help on where to turn for answers. AI can serve as an idea bank and give you new perspectives on areas that you may not have dealt with before. (SL5)

This involved AI as a source of information in everyday activities: “I can use AI, for example, as an information base linked to various issues that I face in my everyday life as a principal” (SL3). However, AI as a tool also offered support in specific areas such as the Education Act: “I see that one possibility is to get quick answers to questions I am wondering about, such as the School Act” (SL4). Another school leader reflected on the use of AI in school law:

> I still wonder if AI could not be a help and support in [school] law. Some legal texts are sometimes difficult to interpret and apply to different situations that arise. Here, perhaps questions to AI can help with the interpretation and thus contribute to more informed decision-making. (SL11)

For one school leader, the use of AI opened opportunities to deepen knowledge in specific areas, as AI “often provides better answers than search engines, and the knowledge can be deepened through follow-up questions. Instead of asking an expert, I can get an overview of the state of knowledge and dig further into what is most relevant” (SL7).

School leaders also saw the use of AI for support in *Administrative tasks*. For many of the school leaders, AI was used to this end. One school leader explained, “I have the possibility to get answers quickly to specific questions” (SL1). This involved a wide variety of school administration tasks, such as school speeches: “I have support in speeches that I give at school starts and graduations” (SL7). This was also expressed by another school leader in further detail:

> [I] already used it to help with a speech; [it's] amazing how quickly I got help with the keywords I put in as important. [I] was then able to use certain parts, phrases, and sentences. It saved me a considerable amount of time. (SL8)

Other school administrative tasks included "Scheduling, distribution of duties, plotter diagram payroll, [and] statistics" (SL1). Another school leader saw the uses for support in scheduling meetings as well as gathering, understanding, and presenting information:

> AI can be used by getting help to understand content in different texts. [It] helped to see how I can structure meetings and content in different presentations . . . [I'm] thinking we could use it in scheduling. [It] would be interesting to gather facts and use AI to make decisions that benefit my organization. (SL8)

One of the school leaders provided another example: "[It is useful] as support in letters and texts, but it is important that you see it as support and not as a finished delivered result" (SL5) in order to "[find] structures to work in the systematic quality work and with the governance and management of the organization" (SL6). Many of the administrative tasks were seen to perhaps support organizational administrative tasks that were resource efficient: "Today, I often see AI as a labour-saving tool" (SL11); however, another school leader saw that AI provided "learning in different types of operational and development issues that the school needs to work on" (SL6). In summary, one of the school leaders elaborated on the opportunities in relation to the current challenges with AI: "Challenges right now are to become familiar with the opportunities AI can give me. I think the opportunities are ultimately resource efficiency, saving of time, and methodical work" (SL1). This support was also important regarding the work with data in school:

> I think AI will help us a lot in producing data and analysing data in our organization. We often have large and complicated data sets where AI can be a valuable tool to see connections and help us evaluate the organization. (SL11)

For several of the school leaders, this also involves being a model for AI support in work: "I would like to set an example and show my colleagues how it can be used in a good way in everyday school life" (SL8). In a similar line of thought, one school leader saw that AI support could imply more time for pedagogical leadership: "If AI can help with, for example, analysis, administrative tasks, etcetera, it gives me more time to focus on pedagogical leadership. [It is important] to set an example to the staff so they can apply it to their contact with students" (SL12). In relation to the many challenges that may be related to the use of AI, a school leader reflected on the need to learn about and use AI: "I have to learn the tool to be able to use it myself when I ask teachers to use it. I see that teachers have to use AI in order not to become user-hostile" (SL4). Another school leader discussed their own organization:

> A good prerequisite in my organization is that all staff, from the janitor to the operations manager, have received basic AI training that we school leaders can build on. I myself intend to utilize the benefits where possible and, in this way, try to be a role model for the staff. (SL13)

Opportunities

The school leaders also saw opportunities in providing support for teachers and students in teaching and learning in the classroom. According to the school leaders, this could involve new approaches with teachers: "I can use AI to develop different pedagogical approaches and structures for tasks that I want to implement with teachers" (SL3). This could also involve pedagogical leadership and support and the guidance of teachers: "I can use AI to get tips on conference arrangements/assignment arrangements for my management of staff. I can guide educators to use AI in their teaching to generate time for staff who are always short on time" (SL4). One school leader saw opportunities for teachers to utilize AI, which in turn could support students from several perspectives:

> I see that AI can compensate and help students with difficulties in terms of quickly getting a text template or quickly retrieving facts in a subject they are going to learn. I see that AI can compensate for students with language disabilities in the production of text and easily help them get quick answers. (SL4)

One school leader saw opportunities in AI for teaching and learning for both students and teachers, as AI can "help many, for example, as a support function and source of knowledge for staff and students in schools" (SL3). This support could also provide support for teachers in planning activities and skills needed for work with students: "I could suggest to the teachers to use AI to get different suggestions for activities or projects to do with the children. They could use it in their planning for different tips, ideas, and thoughts" (SL10).

Other school leaders thought that AI can support teachers as staff. This involves understanding teachers as a group: "[It can] give me a summarized picture of what is happening in the activities in forums where educators give their picture. In this way, [it] can give me a learning experience of how educators think" (SL14). Another school leader also expressed a similar reflection:

> I think that AI could possibly help me formulate deeper questions and ideas about a specific area where I need deeper reflection in order to create greater understanding. One such example could be different types of personnel

> matters where I need a greater understanding of how educators think and reflect. (SL15)

One school leader noted the idea of the use of AI as a pedagogical tool for teachers to support students: "I feel a strong belief in utilizing AI as a complement in everything I do and, above all, a tool for educators to find faster and easier ways to help themselves to help the children" (SL14). Although many of the school leaders mentioned using AI to support teachers and students as groups, one school leader reflected on the use of AI to support individual students: "I have also used AI in challenging situations around individual students and then got good inspiration for measures at the group, individual, and organizational level" (SL7).

Constraints

The next theme involved school leaders' reflection on constraints related to AI. The category *Constraints* involved the following: *Improper use, Time for professional development in the use of AI including prompting,* and *Critical review.*

Improper use was noted in several of the school leaders' reflections. One school leader elaborated on the need for the following: "The school/educators have to make new tasks so that cheating does not occur" (SL4). Another school leader discussed improper use and the consequences of improper use in detail:

> Challenges with AI for me as a principal and responsible for student learning is that students produce finished texts that they then present as their own, which means that teachers judge it as cheating and the student fails the module. This of course affects the students' learning. Then the matter comes to my desk, whether the student should fail the entire course and be given the opportunity to reread or review. Opportunities for students are to find questions—for example, for the upper secondary school work—and tips on investigations that can be done, which can inspire students to stronger motivation in various subjects that allow them to pass—an overall goal for me as a principal in upper secondary school. (SL6)

Professional development was also noted as a constraint. Several of the school leaders reflected on the need for professional development. This often involved the need for more knowledge on the beneficial uses of AI: "I have not yet familiarized myself with how to use AI in the best way. I do not yet have sufficient knowledge on how to use it as a tool" (SL3); and "I need more knowledge on how to use it in a good way" (SL9). New knowledge was necessary to achieve the opportunities of saving time: "Right now I just feel . . . saving time would be great. However, I don't know how to do it because I don't know enough about AI" (SL14). The need for training was also related to prompting in AI: "I see a challenge in that principals and educators need to train themselves to prompt so that AI becomes a good tool

. . . I also don't think that teachers have started to use AI in planning or in connection with challenging students yet" (SL7). However, professional development also requires time: "What is still missing is the time for competence training in this" (SL14).

Critical review was also seen as important in the use of AI. This could involve dialog about AI: "The question of whether we should or should not use it, I think, is not relevant. It is here, and we cannot stop technology. However, I think it's important that we talk about the ethics of using it" (SL2). According to the school leaders, another constraint was that AI as a supplement could be too efficient: "Challenges may perhaps be that it creates an attachment if the help proves too effective—that you would rely on AI in favour of your own judgment. However, I think that awareness of this may avoid this" (SL11). Another risk was disinformation: "The challenges I see are that there can be disinformation shared . . . to be able to deal with AI in the best way" (SL8). One of the school leaders summarized the challenges: "resistance from staff, potential for cheating, rapid developments (difficult to keep up), and technical knowledge" (SL12).

In summary, the most prominent support is when AI is perceived as an inspiration, expert, and source of knowledge that can provide quick and profound answers, which shows that AI is primarily used to support school leaders in administrative tasks. Opportunities are seen in providing support for teachers and students in teaching and learning in the classroom. Constraints involve improper use, as well as time for professional development in the use of AI, including prompting and critical review.

Discussion

The aim of this paper was to examine and analyse how school leaders perceive and engage with the integration of AI for their learning processes and leadership practices. The following research question was posed: 1) How do school leaders perceive opportunities and constraints related to their professional use of AI in learning and leadership practices?

In regard to professional use of AI in learning and leadership practices, the school leaders' perceptions are seen on several levels. Regarding their own learning, they appear to have identified activities and uses of AI for increased efficiency (Hutami, 2024; Strzelecki, 2023). This is apparent in terms of saving time in administrative tasks and accessing information. Several of the school leaders viewed access to information as supportive in reformulating and summarizing texts they are required to use in their profession, such as laws and regulations that may be difficult to understand in practice. This use saves time as it gives school leaders direct access to information and profound answers. Several of the school leaders described the use of AI as beneficial in their studies in the national principal program as well. AI is utilized to summarize and make difficult texts more accessible, which in turn also allows school leaders to save time.

The school leaders in this study confirmed the importance of supporting teachers and students in their learning. For many of the school leaders, this involves acting as a role model, discussing, and sharing efficient use as well as ethical use and critical evaluation (Rudolph et al., 2023). Many of the school leaders discussed ethical use—that is, supporting teachers in the shift from AI as a method for cheating, as well as propagating the new methods in the formulation of assignments and assessments that support students' learning (Neumann et al. 2023; Rudolph et al., 2023). This work will be important for school leaders in supporting teachers' use of AI to support students. Further issues of discussion for school leaders include guidelines for the use of AI (e.g., supporting students in their learning activities with the support of AI), which may offer opportunities for new forms of leadership (Dogan & Arslan, 2025).

For school leaders, as expressed by those interviewed in this study, more knowledge about the efficient and ethical use of AI in learning and leadership will be necessary (Dogan & Arslan, 2025; Fullan et al., 2023). Here, the need for professional development is identified. In addition, school leaders require professional development as school leaders. Finally, school leaders' professional use, as well as how they model the use of AI in their educational settings, may in turn support teachers' and students' application of AI for teaching and learning, and it may benefit the school as an organization.

References

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health, 11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology, 32*, 444–452. https://doi.org/10.1007/s10956-023-10039-y

Dogan, M., & Arslan, H. (2025). The role of artificial intelligence in school leadership. *Revista de Pedagogie Digitala*, *4*(1), 23–30. doi.org/10.61071/RPD.2531

Duran, A., & Ermiş, U. F. (2024). A qualitative focus on school leaders' perceptions of roles, responsibilities, and expectations in enhancing equitable access to educational technology in the era of Gen-AI. *Ahi Evran Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *10*(1), 208-227. https://doi.org/10.31592/aeusbed.1440249

Fullan, M., Azorín, C., Harris, A., & Jones, M. (2023). Artificial intelligence and school leadership: Challenges, opportunities and implications. *School Leadership & Management*, *44*(4), 339–346. https://doi.org/10.1080/13632434.2023.2246856

Hutami, S. (2024). Utilizing technology and artificial intelligence in educational administration to enhance school performance at junior high school. *PPSDP International Journal of Education*, *3*(2), 197–212. https://doi.org/10.59175/pijed.v3i2.302

Karakose, T., & Tülübaş, T. (2024). School leadership and management in the age of artificial intelligence (AI): Recent developments and future prospects. *Educational Process: International Journal*, *13*(1), 7-14.

Moon, J. A. (2006). *Learning journals: A handbook of reflective practice and professional development.* Routledge. https://doi.org/10.4324/9780203969212

Neumann, M., Rauschenberger, M., & Schön, E.-M. (2023). "We need to talk about ChatGPT": The future of AI and higher education. *2023 IEEE/ACM 5th International Workshop on Software Engineering Education for the Next Generation (SEENG)*, (pp. 29-32). IEEE. doi:10.1109/SEENG59157.2023.00010

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Strzelecki, A. (2023). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments*, *32*(9), 5142-5155. https://doi.org/10.1080/10494820.2023.2209881

Tyson, M. M., & Sauers, N. J. (2021). School leaders' adoption and implementation of artificial intelligence. *Journal of Educational Administration*, *59*(3), 271–285. https://doi.org/10.1108/JEA-10-2020-0221

Van Quaquebeke, N., & Gerpott, F. H. (2023). The now, new, and next of digital leadership: How artificial intelligence (AI) will take over and change leadership as we know it. *Journal of Leadership & Organizational Studies, 30*(3), 265–275. https://doi.org/10.1177/15480518231181731

Wang, Y. (2021). Artificial intelligence in educational leadership: A symbiotic role of human–artificial intelligence decision-making. *Journal of*

Educational Administration, *59*(3), 256–270. https://doi.org/10.1108/JEA-10-2020-0216

Author Details

Marcia Håkansson Lindqvist
Department of Education
Mid Sweden University
SWEDEN
marcia.hakanssonlindqvist@miun.se

Fanny Pettersson
Department of Education
Umeå University
SWEDEN
fanny.pettersson@umu.se

DEVELOPMENT OF A THEORETICAL FRAMEWORK FOR SELF-EVALUATION OF ADAPTIVE DIGITAL LEARNING PLATFORMS BASED ON ARTICIAL INTELLIGENCE: A SYSTEMATIC REVIEW

Ana Pedro, Nuno Dorotea, and Bárbara Azevedo
UIDEF – Unidade de Investigação e Desenvolvimento em Educação e Formação
Institute of Education, University of Lisbon
PORTUGAL

Abstract

This study aims to develop a Theoretical Framework for the self-evaluation of adaptive digital learning platforms through Artificial Intelligence, focused on promoting meaningful interaction and personalizing content based on user profiles. A Systematic Literature Review was conducted with the aim of exploring processes and indicators of effectiveness, as well as the challenges and opportunities associated with digital adaptive learning platforms, with a special focus on the role of automatic, real-time self-assessment as an essential tool for the continuous improvement of the performance, personalization and quality of these platforms. The results point to gaps in the literature related to the integration of AI in the self-assessment of adaptive digital learning platforms and suggest the need for new approaches to improve self-assessment systems.

Keywords: Adaptive digital learning platforms, Artificial Intelligence, PRISMA protocol, Systematic Literature Review

Introduction

The evolution of digital innovation has led to significant transformations across various sectors of society, including education. In the current context, adaptive digital learning platforms play an increasingly important role in the dissemination of knowledge and the facilitation of personalized learning processes. Technological advances, particularly in the field of Artificial Intelligence, have driven substantial changes in these platforms, impacting key areas of education such as pedagogical objectives, content, teaching methods, and assessment processes (Kalota, 2024).

This Systematic Literature Review aims to explore the effectiveness, challenges, and opportunities associated with adaptive digital learning platforms, with a particular focus on the role of automatic and real-time self-assessment as a key tool for the continuous improvement of performance, personalization, and the overall quality of these platforms. The focus is not only on the effectiveness of these platforms but also on the challenges and opportunities related to their implementation in educational settings. Additionally, given that self-assessment remains an underexplored area, the review also seeks to investigate current and potential approaches to automatic self-assessment within these platforms.

Methodology

To analyze the impact of adaptive digital learning platforms in the educational context, a systematic literature review was conducted based on studies published in open-access databases. The objective of this review was to gather and synthesize existing evidence, ensuring a rigorous and objective analysis. This rigorous process allows for the development of credible and reliable research (Ramos et al., 2014), which, by following a structured analysis process, ensures consistency of results and the validity of conclusions. Likewise, the development of systematic research enables the construction of comprehensive and unbiased syntheses of publications within a given scientific domain, reporting data and results rather than theories or concepts. Aromataris and Pearson (2014) state

> the 'systematic review,' also known as the 'research synthesis,' aims to provide a comprehensive, unbiased synthesis of many relevant studies in a single document. While it has many of the characteristics of a literature review, adhering to the general principle of summarizing the knowledge from a body of literature, a systematic review differs in that it attempts to uncover 'all' of the evidence relevant to a question and to focus on research that reports data rather than concepts or theory. (p. 54)

This systematic literature review follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) model, as described by Galvão et al. (2015), which aims to assist authors in improving the reporting of systematic reviews and meta-analyses.

Considering the central objective of this research – to develop a Theoretical Reference Framework for the self-assessment of adaptive digital learning platforms, with an emphasis on the integration of Artificial Intelligence (AI) – the systematic review was structured to include not only the effectiveness of AI-powered adaptive learning platforms but also to explore approaches to their self-assessment, an area still underexplored in the literature. The following steps were outlined for the development of the analysis.

The objectives of the systematic literature review were defined as follows: (i) to identify studies on the effectiveness of adaptive digital learning platforms, with a focus on AI integration; (ii) to identify the main challenges and opportunities in the implementation of these platforms in educational settings; and (iii) to explore the concepts and approaches to automatic self-assessment of adaptive digital platforms – a topic still rarely addressed in the literature, yet essential for the construction of the intended Theoretical Reference Framework.

A set of inclusion criteria was also established: (i) studies on adaptive digital learning platforms that use AI, focusing on personalized learning and self-assessment; (ii) studies defining and characterizing personalized and adaptive digital learning platforms; and (iii) studies on strategies for overcoming challenges in implementing adaptive digital learning platforms in educational environments.

Exclusion criteria included: (i) articles that do not address adaptive digital learning platforms, such as those focused on other methodologies or teaching models; (ii) articles whose full content is not available online; (iii) duplicate articles; (iv) articles that do not directly address the integration of Artificial Intelligence in adaptive digital learning platforms or are not relevant to the research objectives; and (v) articles published in languages other than English or Portuguese.

After defining the inclusion and exclusion criteria, a search key was created based on the research theme and the defined criteria. Keywords were combined using Boolean logic operators (AND, OR, NOT) to establish links between research terms.

The following search strings were used, with terms in both Portuguese and English (see Table 1):

Table 1

Search Strings in Portugese and English for Articles on Adaptive Learning Platforms

Portugese	English
Plataforma Digital <OR> Plataforma de Aprendizagem <AND> Inteligência Artificial <OR> Ensino Personalizado ALL FIELDS <AND> Autoavaliação <OR> Avaliação IN TÍTULO <OR> Desafios ALL FIELDS <NOT> Ensino Tradicional ALL TEXT	Adaptive Learning Platform <OR> Learning Platform <AND> Artificial Intelligence <OR> Personalized Learning ALL FIELDS <AND> Evaluation <OR> Assessment IN TITLE <OR> Challenges ALL FIELDS <NOT> Traditional Teaching ALL TEXT

These search strings were applicable and replicable in the selected databases using the advanced search mode. The databases chosen for the search were EBSCO, MDPI Open Access Journals, SCOPUS, and Web of Science.

To select the studies to be included in this systematic review, the PRISMA model, was used as a methodological guide. This model is a methodological tool used in systematic literature reviews, consisting of sequential steps including: (i) identification of the initial research question; (ii) search and selection of relevant studies; (iii) quality assessment of the studies; (iv) data extraction and synthesis; and (v) results.

Results

Overview of Relevant Studies

The research identified 18 relevant articles on the use of Artificial Intelligence in adaptive digital learning platforms, published between 2020 and 2024. These articles were grouped into three main categories: Adaptive Digital Learning Platforms, AI Technologies and Personalization, and Implementation Challenges and Opportunities.

Adaptive Digital Learning Platforms: This category includes studies that analyze how adaptive digital learning platforms can be adjusted to tailor content and teaching methods based on users' progress and learning styles, emphasizing the flexibility and personalization of pedagogical approaches.

AI Technologies and Personalization: This category encompasses studies that explore how emerging Artificial Intelligence technologies can be applied in the educational context, focusing on how these tools can be integrated into adaptive digital learning platforms to personalize teaching and learning experiences. The use of AI to adapt content according to users' needs is a central theme in this category.

Implementation Challenges and Opportunities: This category refers to articles that discuss the challenges faced in implementing adaptive and personalized digital learning platforms, as well as the opportunities that arise from the use of AI in this process. It also addresses issues related to overcoming technical, pedagogical, and ethical barriers in the adoption of these technologies in education.

Adaptive Digital Learning Platforms

Adaptive digital learning platforms stand out as one of the most promising applications of AI in education. The seven studies identified as relevant to this category analyze how such systems can personalize teaching based on the individual needs of users. Studies like that of Kumar et al. (2024) demonstrate that the use of AI in adaptive environments can significantly enhance academic performance, particularly in STEM and mathematics courses, by providing immediate feedback and adapting content in real time. These results reinforce the need to include self-assessment as a central element in evaluating the effectiveness of these platforms.

On the other hand, the reviewed studies also indicate that the effectiveness of adaptive digital learning platforms is not always guaranteed. Factors such as data quality, user and teacher resistance to personalization, and the lack of a clear pedagogical framework can limit the benefits achieved. Bhatt et al. (2024) analyzed AI-based adaptive learning tools, highlighting their effectiveness in increasing user engagement. Er-Rafyg et al. (2024) explores the barriers and benefits of adaptive teaching, emphasizing the need for effective technological integration. Pradeep et al. (2024) propose a personalized AI-based learning platform that monitors individual user progress and adapts the content accordingly. Vashishth et al. (2024) investigate the application of data analytics in adaptive digital learning platforms to deliver personalized feedback. Wang et al. (2020) compare adaptive systems with traditional teaching methods, demonstrating greater effectiveness of the former. Finally, Rani and Senthil (2024) examine the impact of the pandemic on the adoption of adaptive digital learning platforms, highlighting their effectiveness in remote learning scenarios.

AI Technologies and Personalization

Personalized learning has emerged as one of the greatest benefits of Artificial Intelligence in education, with a growing number of approaches focused on adapting content according to the individual needs of users. As noted by Naseer et al. (2024), key strategies include machine learning algorithms, such as artificial neural networks, which dynamically adjust teaching based on user progress. The use of AI tools in personalization is transforming the educational environment, offering new possibilities for developing unique and effective learning pathways.

Recent studies, such as that of Alam and Mohanty (2023), explore the application of virtual avatars that, by interacting with users in online environments, create a more immersive and personalized learning experience. In these contexts, user-teacher interaction is adjusted by AI-based tools, promoting more fluid and adaptive communication. This approach is particularly effective in distance learning environments, where personalization can be difficult to achieve through conventional means.

Additionally, Fu et al. (2020) investigated how automated scoring tools, which assess user performance in real time, influence motivation for continuous learning.

The use of AI-based assessment systems can provide immediate feedback, allowing rapid adjustments to each user's learning path, thereby enhancing motivation and engagement. Conversely, Gupta et al. (2024) examine the impact of generative AI tools on personalization and educational assessment. By enabling the creation of tailored instructional materials and continuously evaluating user progress, these tools may contribute to more learner-centered education, fostering more effective and personalized learning.

Moorhouse et al. (2023) propose guidelines for the use of AI tools in educational assessment, emphasizing ethical and effective practices. They highlight the importance of using AI responsibly, ensuring that assessment systems are fair and transparent, protecting user privacy, and avoiding algorithmic bias.

Naseer et al. (2024) integrated deep learning techniques to create personalized learning paths in higher education. By using complex models to analyze individual user performance, these tools are capable of adapting content in real time, offering a unique and efficient learning experience.

Finally, Zhai and Nehm (2023) analyzed the impact of AI on formative assessment, highlighting how AI technologies can enhance the consistency and objectivity of feedback. This allows teachers to provide precise and clear information about user performance, facilitating the identification of areas needing improvement.

Challenges and Opportunities in Implementation

The implementation of Artificial Intelligence in education faces a series of challenges that must be addressed to ensure its effective and ethical integration. The absence of a clear theoretical-methodological framework is one of the main difficulties, with many studies highlighting the need for models adapted to the local educational context. Although AI has the potential to enhance learning, few studies offer a systematic approach to how these technologies should be practically and sustainably implemented in educational institutions. In many cases, AI implementation has been treated as an isolated technical solution, without adequately considering integration with existing pedagogical methodologies.

The six studies analyzed in this category address a variety of obstacles but also highlight opportunities to overcome them. Deeva et al. (2021) classified automated feedback systems, identifying significant issues such as a lack of transparency in algorithmic processes and potential biases in automated decisions. These challenges raise crucial questions about user trust in AI technologies, making it necessary to ensure that systems are transparent and fair.

Delello et al. (2024) explores how AI is transforming education, emphasizing issues of accessibility and equity. The authors noted that although AI has the potential to democratize access to education, there are also risks of increasing inequality if implementation is not carefully planned to include all users, regardless of their circumstances.

Kalota (2024) introduces basic concepts about generative AI and its practical applications in education. While generative AI offers great opportunities to personalize learning, the lack of adequate technological infrastructure and resistance to change may hinder its large-scale adoption.

Lee et al. (2024) propose an ethical framework for AI integration in education, highlighting the importance of transparency and fairness. The authors argue that for AI to be truly beneficial, systems must be designed to ensure that all users have equal learning opportunities, without discrimination or bias.

Naithani et al. (2024) investigate the transformative impact of AI in post-pandemic education, noting how lockdowns accelerated the adoption of educational technologies. Despite the progress, the study reveals that many institutions struggle to effectively integrate AI due to a lack of adequate training and resources.

Lastly, Nguyen et al. (2023) address fundamental ethical principles for applying AI in education, such as user data privacy and fairness in decision-making processes. Protecting privacy and ensuring that algorithmic decisions do not favor certain groups over others are central issues to gaining user acceptance and ensuring AI's success in education.

Conclusions

The results of the analyzed studies demonstrate the growing recognition of Artificial Intelligence's potential for learning personalization. Personalization, by adjusting content to individual user needs, has shown a positive impact on both performance and motivation. Real-time content adaptation is a core feature of AI-based platforms, as highlighted in studies such as Bhatt et al. (2024) and Kumar et al. (2024), which underline the importance of tailoring content based on each user's progress. The ability to dynamically personalize learning materials has been identified as one of the greatest benefits of AI platforms, promoting greater user engagement and, consequently, better outcomes.

However, effective implementation of these adaptive solutions continues to pose a challenge. The analyzed studies point out that although AI-based platforms have high potential, their implementation depends on factors such as available infrastructure, teacher acceptance, and the quality of the data used. The absence of a clear theoretical and methodological framework, as well as a systematic adaptation model, may hinder the large-scale application of these solutions. Moreover, the diversity of educational contexts requires a flexible model capable of adapting to different realities, a gap often mentioned in the literature (Alam & Mohanty, 2023; Deeva et al., 2021).

The use of AI models for personalization has also proven relevant, especially in the context of automated feedback systems. These systems are essential for content adaptation, particularly in intelligent tutoring and assessment platforms (Rani & Senthil, 2024).

Nevertheless, true personalization of education is not achieved solely through tools like ChatGPT, which, although capable of offering dynamic responses, do not deeply tailor learning content. The effectiveness of these models depends mainly on their ability to integrate user data to create personalized experiences that meet their specific needs (Bhatt et al., 2024; Vashishth et al., 2024).

Furthermore, AI implementation in education is often influenced by ethical concerns, as mentioned in the study by Nguyen et al. (2023). Issues related to data privacy, the risk of algorithmic bias, and the need for greater transparency in AI systems are widely discussed. Trust in AI platforms is essential for user acceptance, and the lack of trust can compromise the successful implementation of these technologies. Therefore, it is crucial to ensure that AI-based solutions adhere to ethical standards, minimizing biases and ensuring transparency in algorithmic decision-making (Delello et al., 2024; Kumar et al., 2024).

In terms of impact, the results of the selected and analyzed studies indicate that AI has positive effects on user motivation, especially when associated with personalized learning pathways. Studies such as Moorhouse et al. (2023) and Zhai & Nehm (2023) report that personalized learning paths help increase user engagement, creating a more motivating and collaborative environment that can lead to improved performance and better knowledge retention. Additionally, Er-Rafyg et al. (2024) emphasize that AI's ability to automate assessments and provide real-time feedback helps reduce teachers' administrative tasks, allowing them to focus more on users' pedagogical needs.

Inclusion and equity are also key issues raised in the literature. AI has the potential to reduce educational inequalities by promoting equitable access to learning through adaptive systems that accommodate different learning styles (Lee et al., 2024; Naithani et al., 2024). This ability to inclusively personalize content ensures that all users, regardless of their level or context, have access to learning opportunities tailored to their needs.

However, challenges related to teacher training and the adaptation of pedagogical methodologies to AI-based technologies still represent significant obstacles. Teacher resistance to adopting new technologies is often cited as a barrier to effective AI implementation in schools. The lack of continuous training and support in integrating new pedagogical tools with traditional methods may limit the full potential of AI. Therefore, proper teacher training and the provision of support resources are essential to ensure the effective use of AI-based platforms in teaching practices (Alam & Mohanty, 2023; Bhatt et al., 2024).

Acknowledgements

This work was supported by National Funds through FCT-Portuguese Foundation for Science and Technology, I.P., under the scope of UIDEF — Unidade de Investigação e Desenvolvimento em Educação e Formação, UIDB/04107/2020, https://doi.org/10.54499/UIDB/04107/2020

References

Alam, A., & Mohanty, A. (2023). Facial analytics or virtual avatars: Competencies and design considerations for student-teacher interaction in AI-powered online education for effective classroom engagement. In R. S. Tomar, S. Verma, B. K. Chaurasia, V. Singh, J. H. Abawajy, S. Akashe, P.-A. Hsiung, & R. Prasad (Eds.), *Communication, Networks and Computing* (pp. 252-265). *Communications in Computer and Information Science (CCIS), Vol. 1894.* Springer. https://doi.org/10.1007/978-3-031-43145-6_21

Aromataris, E., & Pearson, A. (2014). The systematic review: An overview. *American Journal of Nursing, 114*(3), 53-58. https://doi.org/10.1097/01.NAJ.0000444496.24228.2c

Bhatt, V., Gupta, D. B., & Chandra, G. (2024). Optimizing classroom teaching with AI-based adaptive learning tools. In *2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)*. IEEE. https://doi.org/10.1109/ACROSET62108.2024.10743688

Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & De Weerdt, J. (2021). A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers and Education, 162*, 104094. https://doi.org/10.1016/J.COMPEDU.2020.104094

Delello, J. A., Watters, J. B., & Garcia-Lopez, A. (2024). Artificial intelligence in education: Transforming learning and teaching. In J. Delello & R. McWhorter (Eds.), *Disruptive technologies in education and workforce* *development* (pp. 1-26). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-3003-6.ch001

Er-Rafyg, A., Zankadi, H., & Idrissi, A. (2024). AI in adaptive learning: Challenges and opportunities. In A. Idrissi (Ed.), *Modern artificial intelligence and data science 2024* (pp. 329-342). *Studies in Computational Intelligence (Vol. 1166)*. Springer. https://doi.org/10.1007/978-3-031-65038-3_26

Fu, S., Gu, H., & Yang, B. (2020). The affordances of AI-enabled automatic scoring applications on learners' continuous learning intention: An empirical study in China. *British Journal of Educational Technology, 51*(5), 1674–1692. https://doi.org/10.1111/BJET.12995

Galvão, T. F., Pansani, T. S. A., & Harrad, D. (2015). Principais itens para relatar Revisões sistemáticas e Meta-análises: A recomendação PRISMA. *Epidemiologia e Serviços de Saúde, 24*(2), 335–342.

Gupta, S., Dharamshi, R. R., & Kakde, V. (2024). An impactful and revolutionized educational ecosystem using generative AI to assist and assess the teaching and learning benefits, fostering the post-pandemic requirements. In *2024 international conference on emerging trends in information technology and engineering, (ICETITE)*. IEEE. https://doi.org/10.1109/ic-ETITE58242.2024.10493370

Kalota, F. (2024). A primer on generative artificial intelligence. *Education Sciences, 14*(2), 172. https://doi.org/10.3390/educsci14020172

Kumar, P., Vashishtha, S., Sharma, P., & Agarwal, E. (2024). Exploring the efficacy of adaptive learning platforms enhanced by artificial intelligence: A comprehensive review. In R. Doshi, M. Dadhich, S. Poddar, and K. K. Hiran, (Eds.), *Integrating generative AI in education to achieve sustainable development goals* (pp. 147-168). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-2440-0.ch008

Lee, J., Hong, M., & Cho, J. (2024). Development of a content framework of artificial intelligence integrated education considering ethical factors. *International Journal on Advanced Science, Engineering and Information Technology, 14*(1), 205–213. https://doi.org/10.18517/ijaseit.14.1.19558

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open, 5*, 100151. https://doi.org/10.1016/J.CAEO.2023.100151

Naithani, K., Raiwani, Y. P., Tiwari, S., & Chauhan, A. S. (2024). Beyond the classroom: AI's impact on the future of education. In R. Doshi, M. Dadhich, S. Poddar, & K. K. Hiran (Eds.), *Integrating generative AI in* *education to achieve sustainable development goals* (pp. 322-340). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-2440-0.ch018

Naseer, F., Khan, M. N., Tahir, M., Addas, A., & Aejaz, S. M. H. (2024). Integrating deep learning techniques for personalized learning pathways in higher education. (2024). *Heliyon, 10*(11), e32628. https://doi.org/10.1016/j.heliyon.2024.e32628

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies, 28*, 4221–4241. https://doi.org/10.1007/s10639-022-11316-w

Pradeep, K. R., Manish, A. S., Adithiyaa, A. S., Sahana, N., & Abhishek, S. T. (2024). Personalized adaptive learning platform empowered by artificial intelligence. *2024 international conference on knowledge engineering and communication systems, ICKECS 2024* (pp. 1-8). IEEE. https://doi.org/10.1109/ICKECS61492.2024.10617075

Ramos, A., M. Faria, P., & Faria, Á. (2014). Revisão sistemática de literatura: contributo para a inovação na investigação em Ciências da Educação. *Revista Diálogo Educacional*, *14*(41), 17. https://doi.org/10.7213/dialogo.educ.14.041.ds01

Rani, L. L., & Senthil, S. T. (2024). COVID-19 adaptive e-learning: Data-driven student engagement analysis. In *2024 International conference on integrated circuits and communication systems, ICICACS 2024.* IEEE. https://doi.org/10.1109/ICICACS60521.2024.10498868

Vashishth, T. K., Sharma, V., Sharma, K. K., Kumar, B., Panwar, R., & Chaudhary, S. (2024). AI-driven learning analytics for personalized feedback and assessment in higher education. In T. V. T. Nguyen & N. T. M. Vo (Eds.), *Using traditional design methods to enhance AI-driven decision making* (pp. 206-230). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-0639-0.ch009

Wang, S., Christensen, C., Cui, W., Tong, R., Yarnall, L., Shear, L., & Feng, M. (2020). When adaptive learning is effective learning: comparison of an adaptive learning system to teacher-led instruction. *Interactive Learning Environments, 31*(2), 793–803. https://doi.org/10.1080/10494820.2020.1808794

Zhai, X., & Nehm, R. H. (2023). AI and formative assessment: The train has left the station. *Journal of Research in Science Teaching, 60*(6), 1390–1398. https://doi.org/10.1002/TEA.21885

Author Details

Ana Pedro
UIDEF — Unidade de Investigação e Desenvolvimento em Educação e Formação.
Institute of Education, University of Lisbon
aipedro@ie.ulisboa.pt

Nuno Dorotea
UIDEF — Unidade de Investigação e Desenvolvimento em Educação e Formação.
Institute of Education, University of Lisbon
nmdorotea@ie.ulisboa.pt

Bárbara Azevedo
UIDEF — Unidade de Investigação e Desenvolvimento em Educação e Formação.
Institute of Education, University of Lisbon
barbara-azevedo@edu.ulisboa.pt

INNOVATIVE TECHNOLOGIES IN ETHICAL EXPERTISE: ARCHITECTURE AND FUNCTIONAL CAPABILITIES OF THE SYSTEM

Gulmira Bekmanova, Assel Omarbekova, Dinara Kabdylova, Aiganym Bessembayeva, Alina Yerbolatova, Altanbek Zulkhazhav, and Lena Zhetkenbay
L.N. Gumilyov Eurasian National University
ASTANA, KAZAKHSTAN

Purpose/Objective

The article aims to develop an intelligent system for ethical expertise in research, enhancing training quality, automating ethical reviews, and improving evaluation transparency through AI-driven learning and assessments.

Abstract

This article presents the development of the intelligent system "Ethical Expertise for Researchers", aimed at automating researcher training and ethical review. The system integrates adaptive learning, speech synthesis, avatars, and expert evaluation. Key modules include video lecture generation, testing, a course builder, and content moderation. Built on a microservice architecture, it ensures scalability and personalization. Innovative technologies such as AI-based testing and automatic video creation enhance learning efficiency and transparency. The platform is designed for national and international use, offering a flexible tool for researcher education and certification.

Introduction

The rapid development of artificial intelligence, the swift growth of scientific research volumes, and the globalization of the academic environment are setting new demands for the ethical responsibility of researchers. In the context of digitalization and open access to data, the risk of violating ethical norms is increasing, including issues of confidentiality, participant consent, and integrity in the use of AI. International standards of research ethics are becoming a benchmark for national educational systems as well, which requires clear mechanisms for independent assessment and support of research activities.

In response to these challenges, the concept of the intelligent system "Ethical Expertise for Researchers" was developed, aimed at ensuring high-quality training of researchers and providing a transparent, objective ethical evaluation of their work. The system is designed to offer convenient access to educational materials, create a transparent expert review process, and improve the objectivity of knowledge assessment. A key feature is the use of automated mechanisms for analysis and evaluation. One of the core aspects is the adaptability of the educational process, allowing it to dynamically adjust to the user's knowledge level and needs.

Ethical expertise of research work requires analyzing texts in terms of compliance with ethical standards. The methods of Bekmanova et al. (2024) for analyzing Kazakh political discourse confirm the importance of developing intelligent systems capable of conducting comprehensive assessments of textual content. The application of such methods can be useful for the automated detection of ethical violations in scientific publications and for assessing the degree of argumentation in research materials.

Analysis of Existing Platforms

During the system design process, existing educational platforms such as Udemy (2025), Coursera (2025), and Microsoft Learn (2025) were analyzed. The study revealed that effective learning requires flexible tools for course creation, knowledge assessment, and educational content adaptation. However, most platforms lack mechanisms for automatic video lecture generation and expert evaluation, which makes the proposed system unique.

Additionally, existing educational program management systems were examined, and key requirements for the new system were identified.

Modern scientific research in this field was also reviewed, particularly the works of Ukenova et al. (2025). Ukenova et al. (2025), published in *Sensors*, explores the enhancement of learning systems through the use of avatars by transitioning from basic language compatibility to emotion-based interactions. Including these articles in the analysis of existing solutions helps justify the chosen methodology and architectural decisions of our system.

Application of Digital Technologies in the Educational System

Digital transformation plays a crucial role in modernizing educational processes and implementing innovative solutions. As demonstrated by Bekmanova et al. (2024), the digitalization of education facilitates the shift from traditional learning models to interactive and adaptive formats. This work analyzes key concepts of digital transformation in higher education, including the use of artificial intelligence, digital platforms, and automated learning management systems.

The integration of digital solutions into the "Ethical Review for Researchers" system enhances the efficiency of educational programs through the automated generation of video lectures, personalized learning pathways, and expert evaluation of results. The incorporation of adaptive algorithms and intelligent data analysis ensures flexibility and a personalized approach to researcher training.

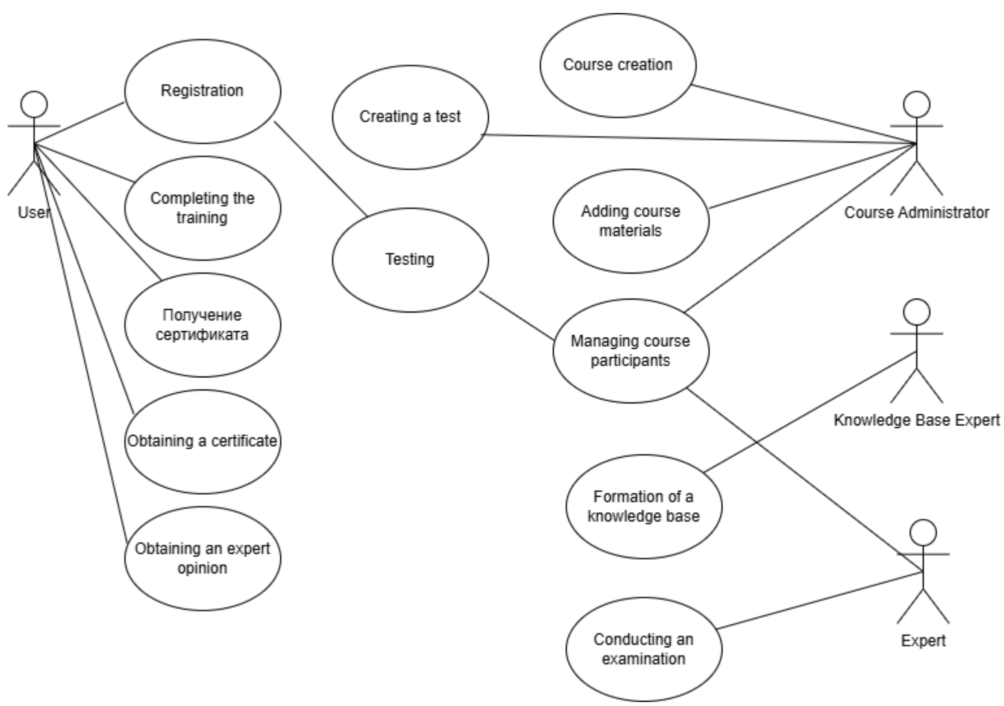
The learning process is structured in a modular format, where each course includes mandatory and optional lectures. To complete a course, the user must pass a test that includes both multiple-choice questions and open-ended assignments assessed by experts.

System Diagram Development

To describe the structure of the system and the relationships between its components, UML diagrams were developed. Initial meetings with users were held to identify both functional and non-functional requirements. Based on the collected data, a Use Case Diagram was created (see Figure 1), illustrating how actors (users or external systems) interact with the system and what functional capabilities it provides.

Figure 1

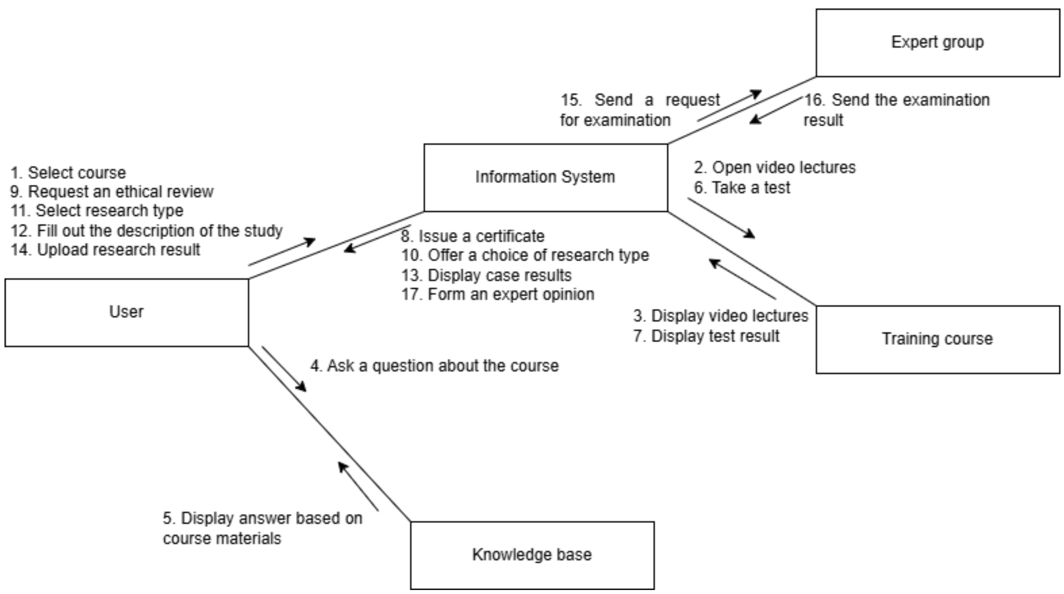
Use Case Diagram of the Intelligent System "Ethical Review for Researchers"



A communication diagram (see Figure 2) was also developed to illustrate the interaction among system components. While similar to a sequence diagram, the communication diagram emphasizes the logical relationships among events rather than their temporal sequence. Interactions among objects are organized using call numbers, providing a clear representation of how information is processed and data is transferred between the system's modules.

Figure 2

Communication Diagram of the Intelligent System "Ethical Expertise for Researchers"



Main Functional Modules of the System

The system includes the following key components:

1. Video Lecture Generation: automatic text processing, speech synthesis, and avatar animation.
2. Flexible Testing System: the ability to personalize tests and assess knowledge.
3. Course Builder: creation of educational programs with various performance evaluation options.
4. Expert Evaluation: the capability to conduct ethical reviews using objective criteria.
5. Content Moderation: quality control of uploaded lectures and user rights management

Video Lecture Generation Technologies

One of the key components of the system is the automatic generation of video lectures. This process includes:

- Converting textual content into video format.
- Speech synthesis for voice-over of the materials.
- Use of animated avatars that convey facial expressions and gestures.
- The ability to manually edit videos before publication.

Neural network-based speech synthesis and gesture generation based on text analysis are used to create the videos. This enables rapid creation of educational content without the need for complex editing.

Knowledge Assessment Algorithms

The system includes several methods for evaluating knowledge:

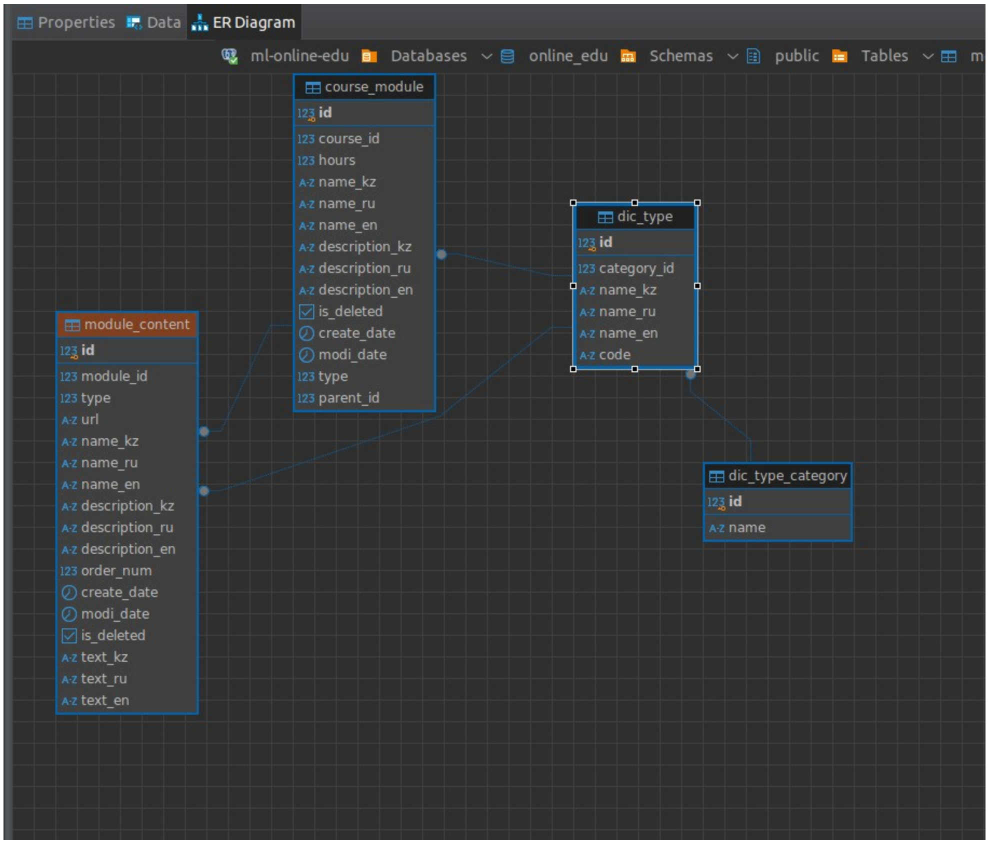
1. Automated Testing: The system randomly generates tests, checking key aspects of the material.
2. Expert Evaluation: Complex tasks with open-ended answers are assessed by experts, who can provide feedback.
3. Hybrid Model: A combination of automated analysis and expert opinion to ensure objectivity.

System Architecture

The system is built on a microservice architecture using Docker Swarm for container management (see Figure 3). This ensures scalability, fault tolerance, and ease of integrating new services. The database is designed with normalization in mind to ensure high performance. Data is structured to enable fast retrieval of information and efficient storage of metadata related to educational programs. A data caching system is implemented to speed up the operation of the interfaces.

Figure 3

Microservice System Architecture



Implementation and Integration

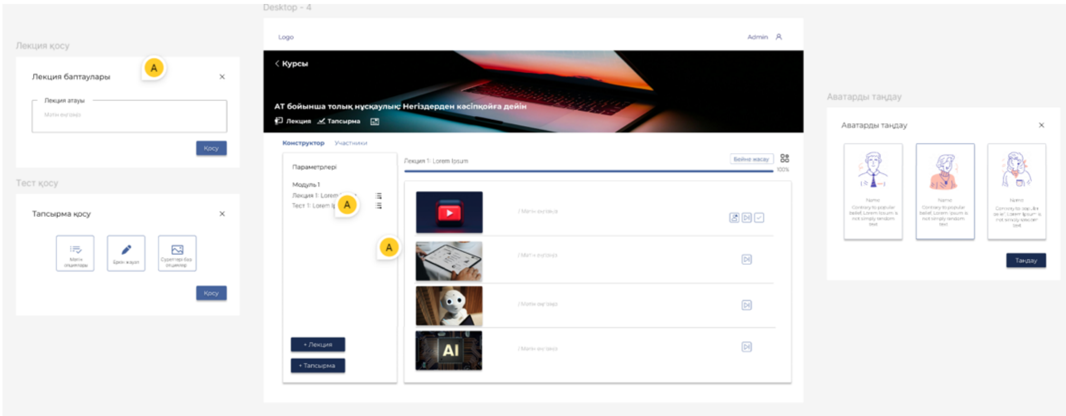
REST APIs were developed for managing courses, lectures, and users. The frontend is implemented using the Vue Composition API and WebStorm. To facilitate working with educational content, technologies like vue3-draggable-resizable are used. Additionally, integration with user management and security systems has been ensured, enabling the configuration of access levels and the personalization of content for different user groups.

Automatic Generation of Video Lectures

One of the innovative features of the system is the creation of video lectures based on text materials. This is achieved through speech synthesis, animation overlay, and the option for manual video editing. The user can upload materials, choose an avatar, and adjust the animation. The generation occurs automatically, after which an editor is available for making changes (see Figure 4).

Figure 4

Prototype of the Automatic Video Lecture Generation Interface

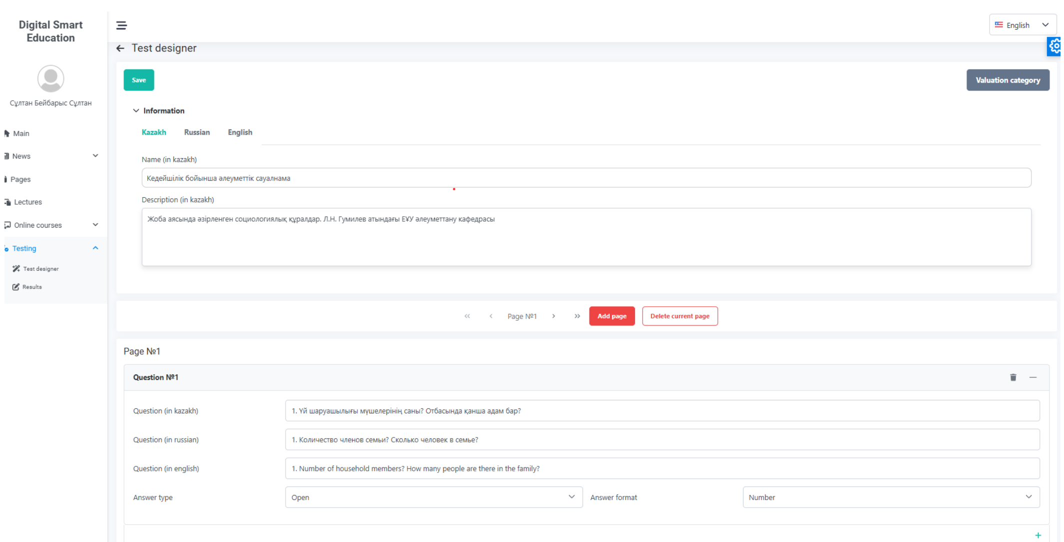


Assessment of Knowledge and Issuance of Certificates

The system supports two learning formats: completing individual courses and a full educational program with a final exam. Customizable assessment criteria are used to conduct objective expert evaluations (see Figure 5). The final exam is generated dynamically, with randomly selected questions covering all key topics of the program. Upon completion of the program, certificates are issued, confirming successful completion of the course.

Figure 5

Question Editing Interface



Conclusion

The developed system represents an innovative tool for automating learning and expert evaluation of research works. The integration of modern data processing technologies, automatic video lecture creation, and microservice architecture enables the creation of a flexible and effective platform. Future work will focus on further adapting the system to meet the needs of researchers and enhancing the level of personalization in the educational process. It is expected that the implementation of this system will significantly simplify the process of ethical review and increase the transparency of research evaluation. In the long term, the system may become a universal platform for training and certifying researchers at an international level.
This research is funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant №. BR21882302 Kazakhstan's society in the context of digital transformation: prospects and risks).

References

Bekmanova, G., Mussina, G., Omarbekova, A., Ospanova, A., Zulkhazhav, A., & Sultan, B. (2024). Aspects of digital transformation of higher education in the Republic of Kazakhstan. In O. Gervasi, B. Murgante, C. Garau, D. Taniar, A. M. A. C. Rocha, & M. N. Faginas Lago (Eds.), *Computational science and its applications – ICCSA 2024* (Lecture Notes in Computer Science, Vol. 14819, pp. 111–123). Springer. https://doi.org/10.1007/978-3-031-65282-0_7

Bekmanova, G., Yergesh, B., Omarbekova, A., Ongarbayev, Y., & Zulkhazhav, A. (2024). Ethical aspects of analyzing Kazakh political discourse. In O. Gervasi, B. Murgante, C. Garau, D. Taniar, A. M. A. C. Rocha, & M. N. Faginas Lago (Eds.), *Computational science and its applications – ICCSA 2024* (Lecture Notes in Computer Science, Vol. 14819, pp. 138–145). Springer. https://doi.org/10.1007/978-3-031-65282-0_9

Coursera. (2025). World-Class Learning for Anyone, Anywhere. https://www.coursera.org/

Microsoft Learn. (2025). Technical Learning Platform for IT Professionals. https://learn.microsoft.com/

Udemy. (2025). Online Learning Platform for Professional Development. https://www.udemy.com/

Ukenova, A., Bekmanova, G., Zaki, N., Kikimbayev, M., & Altaibek, M. (2025). Assessment and improvement of avatar-based learning system: From linguistic structure alignment to sentiment-driven expressions. *Sensors, 25*(6), 1921. https://doi.org/10.3390/s25061921

Author Details

Gulmira Bekmanova
Board Member - Vice-Chancellor for Digitalization - Digital Officer
L.N. Gumilyov Eurasian National University
Astana, Kazakhstan
gulmira-r@yandex.kz

Assel Omarbekova
Department of Digital Development
L.N. Gumilyov Eurasian National University
Astana, Kazakhstan
omarbekova_as@enu.kz

Dinara Kabdylova
Department of Digital Development
L.N. Gumilyov Eurasian National University
Astana, Kazakhstan
dinara.kabdylova@gmail.com

Aiganym Bessembayeva
Department of Digital Development
L.N. Gumilyov Eurasian National University
Astana, Kazakhstan
aiganymbessembayeva@gmail.com

Alina Yerbolatova
Department of Digital Development
L.N. Gumilyov Eurasian National University
Astana, Kazakhstan
alinazhe2000@gmail.com

Altanbek Zulkhazhav
Department of Digital Development
L.N. Gumilyov Eurasian National University
Astana, Kazakhstan
zulkhazhav_a_4@enu.kz

Lena Zhetkenbay
Department of Artificial Intelligence Technologies, Faculty of Information Technologies
L.N. Gumilyov Eurasian National University
Astana, Kazakhstan
jetlen7@gmail.com

THE CHALLENGES THAT ARTIFICIAL INTELLIGENCE BRINGS TO AUSTRALIAN TRANSNATIONAL PROGRAMS

Kathy Michael
Victoria University
AUSTRALIA

Abstract

As offshore delivery of Australian degrees continues to grow, as do the challenges facing universities with the growing use of Artificial Intelligence (AI) technologies, such as Chat GPT, InstructGPT, Chatbox, DeepSeek, Duobao, Poe among others.

This paper explores the multifaceted impact of AI on an Australian university's transnational education (TNE) programs delivered in China. In particular the paper will look at the challenges that AI presents to academics and students involved with offshore programs with a focus on integrity and maintaining academic standards.

Over a period of 12 months, information on student and staff experience with AI was gathered through a combination of focus group discussions and entries in a critical reflective journal. The study further contributes to the rapid growing literature surrounding AI in university settings with a particular focus on transnational programs, a dimension that has received little attention to date. The discussion in the paper will offer strategies to best manage the use of AI without compromising the pedagogy of TNE programs and the relationships of offshore partnerships.

Key Words: Artificial Intelligence (AI), Australian Transnational Programs, Higher Education, Reflective Practices

Introduction

Transnational Education (TNE) refers to institutions crossing borders to provide educational experiences to students located in a country different from the awarding institution. Given TNE course offerings are expected to be a key growth area for Australia, it is appropriate that the concept of AI within the TNE landscape be examined (Australian Government Department of Education, n.d.).

This study will look at three TNE programs delivered by an Australian university in different regions of China, namely, Beijing, Shenyang and Kaifeng. All three programs operate using similar models, allowing Chinese students to pursue a business course directly from China with the option for students to transfer to Australia in their second year and complete their course with the awarding university.

The paper will refer to these students as Offshore International Students (OIS). Likewise, the academics teaching into these programs in China will be referred to as International Local Staff (ILS). Whilst these OIS are required to follow the same curriculum and policies of the awarding university, there are always some slight alterations to ensure that OIS get the most from their Australian degree. The courses are delivered and assessed in English but where appropriate, curriculum development can be adapted to cultural differences in order to bring relevance to the international student audience. In line with Australia's Tertiary Education Quality and Standards Agency (TEQSA), the federal regulator of Australian universities, assessment is moderated by staff from the awarding university.

The popularity of AI tools has inevitably increased to users, offering a range of capabilities from generating codes, images, videos and much more. Whether we like AI or not it is here to stay and it is growing profusely across all aspects of life, in particular the education sector. Whilst AI has dealt the education system with complexity and uncertainty, it has also filled us with excitement with what can be achieved with AI if used correctly in the appropriate landscape.

The increase global hype of how Generated AI has impacted the higher education (HE) sector has resulted in a spike of research (Baidoo-Anu & Ansah, 2023; Hu, 2023; Bearman et al., 2023; Lee et al., 2024; Jin et al., 2025), yet very little on examining and or exploring the concept of AI in TNE programs. Therefore, this paper will gather feedback from those directly involved with TNE programs primarily staff and students.

Research Design

This paper offers a critical reflection from the role of Program TNE Coordinator for China coordinating various degree joint programs across China within the Business faculty, namely Accounting, Banking and Finance and International trade.

Reflective research using focus groups has shown to offer an insightful approach to qualitative research inquiry (Chai et al., 2024). Documenting an educators experience provides an important role with action research. Action research meaning the educator or researcher is participating in the matter that is being researched in this case, navigating the confronting challenges of technologies like AI (Simmons et al., 2021).

The study sought to draw on qualitative data to understand how OIS engage with AI tools and their comprehension of such tools in their studies. Likewise, information was also collected from ILS assessing the challenges they faced with AI at HE. This information was collected using the following instruments: (i) qualitative data collected from focus group discussions and through survey questions; (ii) analytical data collected through a critical reflective journal; (iii) firsthand observations and direct involvement with reported AI cases; and (iv) email correspondence and WeChat messages with staff and students.

A total of thirty OIS and six ILS volunteers were recruited across the three China TNE Sites. Communication was maintained via zoom meetings, the occasional face to face meetings through site visits, emails, and WeChat, a social network platform used in China. Given that the focus of the study is on TNE programs and recognising that quality assurance for such programs requires the awarding university to moderate assessed work, it was fitting to include local Melbourne Staff from the awarding university in the focus groups.

Definition of Artificial Intelligence (AI)

Technology has long played a role in enhancing education, and with it always came challenges, but nothing like the wave of obstacles that the launch of ChatGPT by OpenAI introduced to us in November of 2022. As highlighted by the United Nations Educational, Scientific and Cultural Organisation (UNESCO, 2021), understanding this changing landscape, particularly within a TNE environment, requires us to examine the correlation between technology, language, and pedagogy.

The term AI is very broad and covers an array of technologies. At its core AI is the intersection of a machine and a human. It refers to machines having the ability to take on human intelligence, allowing them to complete human tasks such as problem solving, reasoning, making decisions, and much more. In order to have a grasp on what we are dealing with globally, it is vital that we establish what artificial intelligence (AI) means in today's educational setting. As highlighted by Schuett (2023), having a globally accepted technical and legal definition of AI would help streamline regulation, research, and educational policies.

The concept AI dates back to the 1950's, when Alan Turing had predicted that computers would at some point become thinking machines. This led to the creation of the "Turing Test", a significant tool well-known for determining whether the cognition of a machine was comparable to that of a human being. The term "artificial intelligence" was introduced by John McCarthy in 1955 as the science and engineering of making intelligent machines (Cordeschi, 2007). Copeland (2023) further referenced AI as the capacity of a computer or a computer-controlled robot which could execute tasks typically requiring human intelligence. Lastly, Kaplan & Haenlein (2019) define AI as "a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (p. 5). Irrespective of the definition applied a consistent similarity emerges, that AI is the ability of a computer to perform various intelligent task(s).

Findings

Academic Integrity of Assessment

Academic integrity has been defined as a commitment to the values of "honesty, trust, fairness, respect, responsibility, and courage" in academic work (International Center for Academic Integrity, 2021, p. 4). It is difficult to engage in meaningful research or discussion about AI in HE without addressing the issues it brings to academic integrity. Only by addressing the ethical concerns that accompany the integration and spread of AI within HE are we then able to propose solutions and policies to regulate the way it is used.

Notably since the COVID -19 pandemic, when teaching was propelled to online platforms, academics have generally found a decrease in academic integrity adherence (Eshet, 2024). The rise of AI generated content from tools like ChatGPT, amongst other AI tools, has added another layer of complexity to academic integrity (Elkhatat et al., 2023). Until recently, academic integrity to most of us working in the HE sectors represented either a case of plagiarism, contract cheating, and or collusion found in student work (Newton, 2018; Parkinson et al., 2022). The integrity matters we faced predating the development of AI tools seemed much simpler to deal with.

AI tools have unlocked for students a passage to customised learning but seemingly have also facilitated academic dishonesty. Studies have reported students leveraging AI tools to automate assignment completion, avoid plagiarism detection, prepare presentations, and other shortcuts (Nguyen & Goto, 2024). Furthermore, the ease of access to online paraphrasing tools provides students, particularly OIS, with the opportunity to submit work they have not directly prepared themselves, placing them at risk of not achieving the likely learning outcomes.

The unsettling aspect of all this remains that information collected from OIS for this study suggest they have failed to see how paraphrasing tools can also lead to a breach of academic integrity. This informs me as an educator that despite students having to complete compulsory academic integrity models as part of their first-year studies, they are still not sure how the use of AI paraphrasing tools can lead to a breach of academic integrity. This poses numerous questions such as, are the academic integrity models fulfilling what they were set out to achieve, and are staff (both Melbourne and ILS) explaining the concept and impact of AI clearly to OIS?

When OIS students were asked what they thought a breach involving the use of AI looked like, there was unified consensus that instructing AI to complete an entire assignment piece would be considered an academic breach. However, when asked, "if a student writes their assessment piece entirely in Chinese and then translates the entire piece into English using an AI tool and submits the work as their own, would this constitute an academic breach?" there was complete silence. Furthermore, when OIS were asked whether they have viewed or know what the academic integrity policy of the awarded university contains, the response was "No". One student replied, *"one subject outline indicated that we can use AI so I assumed the same with all subjects" (*Student 14).

Several OIS from one of the China sites completing a second-year unit in a business degree were asked to explain the use of AI in a group assignment after Turnitin AI detector had reported a high AI percentage. The student responses were as follows:

> *Subsequently, I combined these data for analysis, and after drawing preliminary conclusions, I used AI to assist in rewriting to make the article more professional and academic.* (Student 11)
>
> *When I used AI to assist with my assessment 3, I integrated the introductions and main bodies written by other members into one file, and asked AI to help me summarize a conclusion and recommendation.* (Student 8)
>
> *We did not use AI to create arguments or alter analytical substance. We used Doubao and our goal was simply to achieve clearer academic expression.* (Student 2)

The information collated on OIS's understanding of academic integrity signifies more education is required to enhance and support student education on AI and its appropriate use. This feedback led to guidelines prepared requesting Melbourne unit convenors to ensure their unit spaces on the Learning Management System (LMS) had clear instructions on the following: (i) *Use of AI tools*, specifying whether OIS can use AI in their assessment. If allowed, specify how and when AI is appropriate to use; (ii) *Translation tools*. Given feedback from focus groups pointed to many OIS using AI based translation tools, it is crucial to point out to students the level of acceptance and the requirement to acknowledge the use of AI; and (iii) *Academic Integrity Policy*. Provide the link to the academic integrity policy reinforcing that some responsibility needs to be taken up by the student to understand what can and can't be done. Remember that policies are changing at high speed.

How Offshore International Students Engage with Artificial Intelligence

Throughout the focus group conversations, OIS spoke very openly about their views and use of AI tools in their learning journey. They all admitted to using AI tools to various degrees to assist them with the completion of assessment. It became apparent that AI has become embedded into students' daily lives and increasingly integrated into their academic work. In response to questions about how they use AI in their studies, the below three areas were revealed:

1. Improve Writing Skills

AI powered writing tools such as Poe and DeepSeek are used by OIS to help structure and rephrase sentences, polish up language, and check grammar and tone.

> *We used DeepSeek to check the grammar in our article and modify our language, so it reads more smoothly (*Student 4)

> *As non-native English speakers, we want to achieve higher scores by using more formal and professional language. Therefore, we employed DeepSeek to check for grammar errors and to improve the quality of our writing (*Student 9)

2. Translation Support

Noting that these OIS are studying in a second language, it would be appropriate to require support translating terms, unit content, and or assessment requirements.

> *some sentences are borrowed from translation software (*Student 7)

> *We also used translation software Netease Yodao Translator to help us understand some complex parts (*Student 2)

3. Study Aid

Similar to using translators, some students were found to use DeepSeek to have content explained to them differently so that they were better able to understand. This allowed students to resolve their queries instantly without needing to schedule appointments with their tutors for assistance. This also demonstrates how AI is personalising learning for OIS.

*If I email my teachers in Australia or China, they take a couple of days to respond but DeepSeek can help me straight instantly with my questions (*Student 1)

*We used DeepSeek to help us research the assessment topic. We entered the question, "What is the importance of sustainability reporting?" In the AI's response, we found an aspect that we hadn't considered (*Student 21)

*In order to enhance efficiency and reduce workload, we used Poe and DeepSeek to search for references. We also asked the system to give us the references using the Harvard format (*Student 4)

*The reason we use AI is simply to increase efficiency and shorten the completion time of tasks (*Student 15)

Challenges in Accurately Detecting AI

As academics, our concerns for the misuse of AI and the lack of reliable resources to detect AI are not unfounded. There is a range of AI detectors available (Turnitin's AI writing detector, OpenAI's AI text classifier, AI Writing Check, GPTZero, Copyleaks, GPT Radar, Originality.ai, Catch GPT, Winston, Content at Scale, amongst others); however, the accuracy rates have been reported to be unreliable.

Programs such as Turnitin AI were initially embraced as a deterrent to the problems created by AI (Ismail & Jabri, 2023; Elkhatat et al., 2023) but research continues to demonstrate the shortcomings of Turnitin AI (Chaka, 2023; Weber-Wulff et al., 2023) therefore forcing universities to question the use of this and other AI detector tools. Turnitin AI amongst other similar products have been reported to give false readings and, in some cases, unfairly target non-native English speakers (Fowler, 2023; Klee, 2023). It has been further reported that AI detectors are more likely to label text written by non-native English speakers as AI written work (Myers, 2023). The term "nonnative English speakers" generally refers to people who have learnt English as a second language, hence the OIS used in this study would differently be captured under this heading.

Despite the shortcomings of AI detectors, the absence of a more effective solution means that most universities will keep using them in hopes of deterring students and maintaining some level of academic integrity. In light of the shortfalls, universities should consider a more holistic approach when dealing with reported AI academic integrity cases. Simultaneously, effort should be directed to universities to look at redeveloping assessment.

At present the awarding university of the three TNE programs referenced in this study are taking an educative approach to AI identified via detector tools. The shortcomings of the AI detection tools have become known to students, in particular to OIS, and they have not held back in using it as an argument when they are asked to please explain the AI detected percentage in their assessment submissions. The group leader of an assessment (Student 17) replied with the following, when asked to explain to staff how AI was used within the group project, *"…AI is bias towards students like us who have a second language English. We didn't use AI incorrectly…"*

International local staff opinions on AI detection tools were mixed. Whilst staff thought it would be a good prompt to start discussions with OIS, they also shared concerns about their accuracy. ILS also expressed the time required to look into high AI reported cases. Some classes in the TNE programs are made up of 400 plus students, so you can appreciate the concern that echoes through staff. Some other comments put forward by ILS included:

> *Yes, I think it is ok to have the Turnitin AI detector there as a deterrence*" (Staff 2)

> *Everyone, including the students know the detectors are bias against non-natives so I think they are not going to work and waste time* (Staff 7)

> *Yes, I feel it would be a deterrence, but it is a lot of work to look at every high AI especially when we know it is most likely due to translation* (Staff 6)

It should be pointed out that in almost all cases of AI detection and breaches where OIS were required to explain the AI detected, it was referred from the participants as a case of translation, checking English proficiency, and/or correcting grammatical errors.

Challenges with Regulation / Policies / Strategies

As pointed out in the previous section, academic integrity is a cornerstone of HE, and without it there would be no credibility to universities. The principles of academic integrity need to be embedded into institutional policies to promote intellectual growth. Regardless of the benefits associated with AI in education, universities have moved fast to regulate the way AI is used within an academic setting. We need to be mindful that these regulations /policies will continue to evolve as universities better understand how these platforms work and the impact they have on academic integrity.

For instance, the Pro-Vice Chancellor of the University of Cambridge stated the need to recognise ChatGPT rather than not ban it (Olsson, 2023). Along the same thoughts, in 2025 the website of the university indicates that AI is not banned, but students need to be mindful how and where it can be used. Similarly, The University of Oxford also makes no mention of AI being banned but rather encourages the use of AI as part of the learning process and also points out that "…in some instances academic staff, departments and colleges may give more detailed guidance on how they expect AI tools to be used (or not used)…." (University of Oxford, 2025).

Australia's TEQSA, in their ongoing work to carry out and regulate academic standards across Australian institutions, has reminded students to exercise caution in their use of AI in order to ensure that engagement with any AI tools aligns with university policies and academic integrity guidelines (Australian Government TEQSA, 2024).

In 2023, TEQSA further commissioned a document to support university faculties in evaluating the influence which AI has had on assessment practices. The document was intended to offer expert insights on how and why assessment strategies may need to change in this fast-evolving AI educational setting (Lodge et al., 2023). TEQSA has also prepared a document summarising Australian institutional responses to the use of Generative AI (Australian Government TEQSA, 2023). However, given the rapid updates to policies, for the most accurate and up to date information on what each institute is doing with AI it would be recommended to visit each institute's official website.

Some Australian universities have taken a similar approach. Victoria University (2025) outlines the following in a section of its academic integrity webpage entitled *"*Student Guidelines for using text generating tools in assessments": "In your studies with Victoria University (VU), you may find that some assessment tasks explicitly ask you to use such tools, whereas some other assessment tasks will explicitly ask you to not use them... *"*

The University of Melbourne, Australia further outlines their policy as, *"...if a student submits work created and /or significantly modified by AI tools for assessment as if it was their own, then this may constitute academic misconduct and will be subject to the usual academic misconduct procedures of the University"* (n.d.)

Noting that policies of the awarding university will roll out to any established TNE programs. OIS need to be mindful of the policies their awarding university has established around the use of AI. Using AI maybe restricted, banned, and or in some cases compulsory. Most universities have taken a general view on AI policy, advising students that it will depend on the individual units they are enrolled in. It is not difficult to see why students are confused. Student 6, from the focus group reported just that: *"…it is hard to know what is required from us and we are worried about asking or using the word AI with our teachers…"*. As academics we want our students to be open and to engage with us as much as possible, but due to the lack of clarity on the use and appropriateness of AI, OIS are deterring from having a discussion in case their teachers think they are looking to use AI within their assessment. We cannot avoid the use of AI, and our OIS will need to have some level of AI literacy skills upon graduating (Long & Magerko, 2020)

Enhancing AI Literacy Among Local and Partner Academic Staff

As the AI landscape continues to grow, so must the knowledge of staff. The changes in AI tools and concepts are moving faster than the education system can keep up with. It is crucial for staff to have access to ongoing professional development (PD) opportunities to harness an understanding to this evolving development (Bekdemir et al., 2024)PD can take various forms—external courses or conferences, internal workshops, even institutional focus groups. More importantly, we need to ensure that PD isn't forgotten for TNE partner staff. The ILS need to be just as up-to-date with training so they can execute the curriculum and address student matters on a day-to-day basis.

There was a recurring concern emerging through the ILS focus groups, and that was the absence of professional training opportunities made available to them to help them with the application of AI and methods of detection. The sentiment was consistent with both TNE staff and staff in Melbourne from the awarding university. One staff member teaching into one of the TNE China sites expressed, "*I worry, it is becoming difficult to advise and direct students on the use of AI when I don't fully understand how to maximise the benefits of AI but also minimise the downfalls…there appears to be no consistent application of whether students can or can't use AI" (*Staff 5).

Feeling that you have been left behind can be daunting; as academics we need to stay on top of this fast-growing AI plague and the more, we share, discuss, and exchange with colleagues, the more robust we become. More importantly, in the setting of this paper we cannot lose sense of cultural differences and the approach towards AI from the Chinese perspective.

Limitations and Future Direction

This study focused just on TNE programs of one Australian university operating in China. It could further be expanded to encompass additional TNE sites across other geographical areas, allowing for an examination of factors unique to different locations, such as cultural influence, support services, and resource availability.

The study could further expand and draw comparisons with domestic students enrolled in the same course within the same university in Australia. Expanding the study would allow for a more comprehensive understanding of AI across different educational modes.

Conclusion

Through reflections, focus groups, and one-on-one dealings with AI reported cases, this study sought to identify the challenges that generative AI has inflected on TNE programs that mainly consist of Chinese students. As universities and educators navigate this complex and ever-growing use of AI, it is imperative we continue to share and collect information from diverse stakeholders to navigate ethical considerations and develop a culturally responsive but equitable approach to dealing with AI.

It has become evident that universities need to invest in professional development to ensure AI literacy, pedagogical skills, and an understanding of ethical AI practices for both TNE OIS and ILS. The study has identified that a balanced approach should be taken between mitigating inappropriate AI usage while still having an open mind about the benefits it can bring to teaching and learning in a TNE environment, without compromising academic integrity

The findings of this study further provide insight into how OIS and ILS perceive and use AI. The knowledge collected provides a path to improving the way we create assessment, moderate assessment, and report breaches of AI use. Using a reflective and focus group approach allowed for more open-ended responses and collection of information (thoughts and perceptions). Expanding the study across three TNE geographical sites in China allowed for a variety of viewpoints and experiences.

References

Australian Government Department of Education. (n.d.) *Transnational education*. Countering Foreign Interference in the Australian University Sector. https://www.education.gov.au/countering-foreign-interference-australian-university-sector/transnational-education

Australian Government Tertiary Education Quality and Standards Agency (TEQSA). (2024). Artificial intelligence: Advice for students. https://www.teqsa.gov.au/students/artificial-intelligence-advice-students

Australian Government Tertiary Education Quality and Standards Agency (TEQSA). (2023, May). Summary of institutional responses to the use of generative artificial intelligence. Australasian Academic Integrity

Network. https://www.teqsa.gov.au/sites/default/files/2023-05/AAIN-Institutional-Responses-Generative-Artificial-Intelligence.pdf

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52-62. http://dx.doi.org/10.2139/ssrn.4337484

Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education, 86*, 369–385. https://doi.org/10.1007/s10734-022-00937-2

Bekdemir, I., Gungor, N., Pinarbasi, M., & Basdag, S. (2025). Shaping the future of education with artificial intelligence. *International Journal of Social Sciences, Innovation and Educational Technologies*, *6*(21), 19-28. https://www.issjournal.com/DergiPdfDetay.aspx?ID=223

Chai, C. A., Barrios, M., Gómez-Benito, J., Berrío, A. I., & Guilera, G. (2024). Information retrieval in face-to-face and online focus groups: A systematic review. *International Journal of Qualitative Methods, 23*. https://doi.org/10.1177/16094069241286856

Chaka, C. (2024). Reviewing the performance of AI detection tools in differentiating between AI-generated and human-written texts: A literature and integrative hybrid review. *Journal of Applied Learning & Teaching, 7*(1), 115–126. https://doi.org/10.37074/jalt.2024.7.1.14

Copeland, B. J. (2023). Artificial intelligence. In *Encyclopedia Britannica.* https://www.britannica.com/technology/artificial-intelligence

Cordeschi, R. (2007). AI turns fifty: Revisiting its origins. *Applied Artificial Intelligence, 21*(4-5), 259-279. https://doi.org/10.1080/08839510701252304

Eshet, Y. (2024). Academic integrity crisis: Exploring undergraduates' learning motivation and personality traits over five years. *Education Sciences*, *14*(9), 986. https://doi.org/10.3390/educsci14090986

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity, 19*, Article 17. https://doi.org/10.1007/s40979-023-00140-5

Fowler, G. A. (2023, April 3). We tested a new ChatGPT-detector for teachers. It flagged an innocent student. *The Washington Post*. https://www.washingtonpost.com/technology/2023/04/01/chatgpt-cheating-detection-turnitin

Government of New South Wales. (n.d.). A common understanding: simplified AI definitions from leading standards. Digital NSW. https://www.digital.nsw.gov.au/policy/artificial-intelligence/a-common-understanding-simplified-ai-definitions-from-leading#anchor-what-is-ai

Hu, G. (2023). Challenges for enforcing editorial policies on AI-generated papers. *Accountability in Research*, *31*(7), 1–4. https://doi.org/10.1080/08989621.2023.2184262

International Center for Academic Integrity. (2021). *The fundamental values of academic integrity (3rd ed.)*. https://academicintegrity.org/aws/ICAI/pt/sp/values

Ismail, I., & Jabri, U. (2023). Academic integrity: Preventing students' plagiarism with Turnitin. *Edumaspul: Jurnal Pendidikan, 7*(1), 28–38. https://doi.org/10.33487/edumaspul.v7i1.5392

Jin, Y., Yan, L., Echeverria, V., Gašević, D., & Martinez-Maldonado, R. (2025). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence*, *8, Article 100100348.* https://doi.org/10.1016/j.caeai.2024.100348

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Klee, M. (2023, June 6). She was falsely accused of cheating with AI — and she won't be the last. *Rolling Stone*. https://www.rollingstone.com/culture/culture-features/student-accused-ai-cheating-turnitin-1234747351

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence, 6*, Article 100221. https://doi.org/10.1016/j.caeai.2024.100221

Lodge, J. M., Bearman, M., & Dawson, P. Assessment reform for the age of artificial intelligence. Australian Government Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Myers, A. (2023, May 15). *AI-detectors biased against non-native English writers*. Stanford University Institute for Human-Centered Artificial Intelligence. https://hai.stanford.edu/news/ai-detectors-biased-against-non-native-english-writers

Newton, P.M. (2018). How common is commercial contract cheating in higher education and is it increasing? A systematic review. *Frontiers in Education*, *3,* Article 67. https://doi.org/10.3389/feduc.2018.00067

Nguyen, H.M., & Goto, D. (2024). Unmasking academic cheating behavior in the artificial intelligence era: Evidence from Vietnamese undergraduates. *Education and Information Technologies, 29*, 15999–16025. https://doi.org/10.1007/s10639-024-12495-4

Olsson, E. (2023, January 23). *Pro-Vice-Chancellor for Education: Bans on AI software like ChatGPT are not "sensible". Varsity*. https://www.varsity.co.uk/news/24892

Parkinson, A. L., Hatje, E., Kynn, M., Kuballa, A. V., Donkin, R., & Reinke, N. B. (2022). Collusion is still a tricky topic: Student perspectives of academic integrity using assessment-specific examples in a science subject. *Assessment & Evaluation in Higher Education, 47*(8), 1416–1428. https://doi.org/10.1080/02602938.2022.2040947

Simmons, M., McDermott, M., Eaton, S. E., Brown, B., & Jacobsen, M. (2021). Reflection as pedagogy in action research. *Educational Action Research, 29*(2), 245–258. https://doi.org/10.1080/09650792.2021.1886960

Schuett, J. (2023). Defining the scope of AI regulations. *Law, Innovation and Technology, 15*(1), 60–82. https://doi.org/10.1080/17579961.2023.2184135

The University of Melbourne. (n.d.). University policy and actions. *Academic integrity at the University of Melbourne.* https://academicintegrity.unimelb.edu.au/staff-resources/artificial-intelligence/university-policy-and-actions

Turing, A. M. (1950). I.-Computing machinery and intelligence. *Mind, 59*(236), 433-460. https://doi.org/10.1093/mind/LIX.236.433

UNESCO. (2019). *Beijing consensus on artificial intelligence and education*. UNESDOC digital library. https://unesdoc.unesco.org/ark:/48223/pf0000368303

University of Oxford. (2025). *Use of generative AI tools to support learning.* https://www.ox.ac.uk/students/academic/guidance/skills/ai-study

Victoria University. (2025). Student Guidelines for using text generating tools in assessments. *Academic integrity Guidelines*. https://policy.vu.edu.au/document/view.php?id=412#section7

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity, 19*, Article 26. https://doi.org/10.1007/s40979-023-00146-z

Author Details

Dr Kathy Michael
Business College – Victoria University
Australia
Kathy.michael@vu.edu.au