

2025

# EJEL Volume 23, Issue 1



## Editors

Heinrich Söbke and Marija Cubric

Published by Academic Publishing  
International Limited

Curtis Farm, Kidmore End, Nr Reading, RG4  
9AY, United Kingdom

[karen.harris@academic-publishing.org](mailto:karen.harris@academic-publishing.org)

eISSN: 1479-4403

## EJEL Volume 23, Issue 1

|                                                                                                                                                                                                                                              |         |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| Open-Source Gamification Plug-Ins: A Study on Usability and User Preferences<br><i>Ludmila Walaszczyk, Sylvester Arnaba</i>                                                                                                                  | 01-14   |
| Determinants of Student Adoption of Generative AI in Higher Education<br><i>Hanadi Aldreabi, Nisreen Kareem Salama Dahdoul, Mohammad Alhur, Nidal Alzboun, Najeh Rajeh Alsalhi</i>                                                           | 15-33   |
| Quality Assurance in Distance Higher Education: A Bibliometric Study of Scopus-Indexed Publications Between 1993 and 2024<br><i>Huu-Bich Nguyen, Nguyen Quang-Duy Vu, Duc-Tai Dinh, Hiep-Hung Phamn</i>                                      | 34-52   |
| From Twitch to YouTube Live: A Systematic Literature Review of Streaming in Higher Education<br><i>Jorge Oceja, Carmen Álvarez-Álvarez</i>                                                                                                   | 53-65   |
| An Analysis of Factors Impacting Users' Choice of Freemium or Premium Services in a Mobile-Assisted Language Learning App<br><i>Farah Dita Ashilah, Nurul Hanina Efendi, Yelda Faizah Havara, Putu Wuri Handayani, Nabila Clydea Harahap</i> | 66-80   |
| Beyond Face Recognition: A Multi-Layered Approach to Academic Integrity in Online Exams<br><i>Aivar Sakhipov, Islam Omirzak, Alexey Fedenko</i>                                                                                              | 81-95   |
| Evaluating ChatGPT's Reliability in Second Language Acquisition (SLA): Insights on Language Skills and Technology's Role<br><i>Albatool Ahmad Alhazmi, Muneera Muftah</i>                                                                    | 96-112  |
| Blended Learning and Math Achievement: A Meta-Analytic Review Highlighting the Effectiveness and Heterogeneity<br><i>Amelia Defrianti Putri, Dadang Juandi, Turmudi, Suparman</i>                                                            | 113-128 |
| Virtual – Augmented Reality (VAR) for Science Learning: Development and Impact on Students' HOTS Skills<br><i>Iwan Maulana, Siswandari, Gunarhadi, Agus Efendi</i>                                                                           | 129-142 |
| Quality of e-Learning in Nepalese Universities During the COVID-19 Pandemic<br><i>Ratna Mani Nepal, Shyam Guragain, Jiwnath Ghimire, Bimal Khadka</i>                                                                                        | 143-156 |

# Open-Source Gamification Plug-Ins: A Study on Usability and User Preferences

Ludmila Walaszczyk<sup>1</sup> and Sylvester Arnab<sup>2</sup>

<sup>1</sup>Sieć Badawcza ŁUKASIEWICZ – Instytut Technologii Eksploatacji, Poland

<sup>2</sup>Centre for Postdigital Cultures, Coventry University, UK

[ludmila.walaszczyk@itee.lukasiewicz.gov.pl](mailto:ludmila.walaszczyk@itee.lukasiewicz.gov.pl)

[aa8110@coventry.ac.uk](mailto:aa8110@coventry.ac.uk)

<https://doi.org/10.34190/ejel.23.1.3618>

An open access article under [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](#)

**Abstract:** In the current educational environment, e-learning and online education have gained significant prominence, especially highlighted during the COVID-19 pandemic when their importance increased dramatically. Empirical evidence highlights the undeniable benefits of online learning, with learners globally appreciating the flexibility to access course materials at their convenience, free from geographic or time constraints. Despite many companies' efforts to support online education, some educators are hesitant to adopt this new approach fully. Their concerns revolve around whether online instruction can match traditional classroom teaching's effectiveness, coupled with worries about the financial costs involved in transitioning and the extensive work needed to digitise course materials. Despite these challenges, the integration of plug-ins such as gamification plugins emerges as a transformative force in online education. Gamification plug-ins play a crucial role in infusing gamified elements and mechanics into online courses or digital content, thereby enhancing engagement, motivation, and overall learning experiences. Gamification involves applying game design principles, such as points, badges, leaderboards, challenges, and rewards, to non-game contexts like education or business. In contemporary application, gamification has evolved to include explicit game elements such as quizzes, puzzles, narratives, and mini-game techniques as complementary components in enhancing existing online learning environments. By integrating gamification plug-ins into e-learning platforms or content, educators can create a more interactive and enjoyable learning environment, ultimately leading to improved retention and participation. Thus, in the face of the challenges posed by the transition to online learning, strategically incorporating gamification plug-ins presents a promising approach to addressing concerns regarding efficacy and learner engagement in virtual classrooms. This paper aims to explore the efficacy of specific gamification plug-ins through the evaluation of open-source gamification plug-ins from a user-centric perspective, focusing on their usability and appropriateness. By employing both qualitative and quantitative data collection methods, the research seeks to uncover the reasons behind user preferences and prioritise the plug-ins that best support engagement and functionality in their intended applications. Specifically, the study includes an in-depth review of the relevant literature on gamification in education to position the current investigation within a broader academic context. The gamification tools are assessed by VET trainers, learners, and other stakeholders across five European countries—Poland, Italy, Greece, Cyprus, and the United Kingdom (n=258)—to address the need for practical, effective, and open-source gamification solutions compatible with popular LMS/CMS platforms. This study advances the field by offering a systematic analysis of widely used gamification plug-ins and providing new insights into how these tools can be applied to foster learner engagement. By situating the findings in the context of prior research, this paper highlights how gamification contributes to learning outcomes in ways that have not been sufficiently explored, filling an existing gap in the literature. The research questions aim to identify the most user-friendly open-source plug-ins for gamification and determine their usefulness in creating interactive learning content. Ultimately, these insights can inform the design of more engaging e-learning platforms, improving retention and participation rates.

**Keywords:** e-Learning, Interactive content, Gamification, Plug-ins, Online tools, LMS/CMS platforms

## 1. Introduction

In the ever-evolving landscape of adult and vocational education and training (VET), game-based tools have emerged as crucial elements towards enhancing learning outcomes, engagement, and skills development (Anderson, Anderson and Taylor, 2009; Tay et al., 2022; Chang, 2023). This significance holds, particularly within the European context, where continuous professional development and upskilling are key priorities in nurturing a competitive and adaptable workforce. While Industry 4.0 focuses heavily on technology and digitalisation, the emerging Industry 5.0 suggests emphasis on human-centricity, sustainability, and resilience for ensuring a workforce is upskilled (Kolade and Owoseni, 2022).

As the demands of the workforce rapidly change and new skills become essential, traditional instructional approaches often struggle to captivate and motivate adult learners (Huang & Soman, 2013; Barneva, 2017). Here, game-based methodologies offer a transformative solution, creating dynamic and immersive learning environments that align with the digital competencies and preferences of today's learners. By integrating game

elements, mechanics, and interactive features, these interventions breathe life into training programmes, fostering motivation, knowledge acquisition, and skill development. World Economic Forum (2022) emphasises that investing in education and training could add \$2.54 million to the global economy and a playful approach that also includes game-based learning is key.

Game-based approaches have emerged as powerful tools in educational design, significantly influencing learner engagement, motivation, and learning outcomes. By simulating real-world scenarios, employing scenario-based learning, and leveraging virtual environments, game-based interventions provide learners with opportunities to practise and refine their skills in a safe and controlled setting (Arnab, 2020). Game-based pedagogy bridges the gap between theoretical knowledge and practical application, facilitated through a more active and experiential nature of games towards developing practical competencies, problem-solving abilities, and decision-making skills essential for professional roles.

Game-based approaches operate at the intersection of various disciplines, including education, psychology, computer science, design, and human-computer interaction (Arnab, 2020). This interdisciplinary nature is pivotal in addressing the complex challenges of adult and vocational education and training (VET). By integrating diverse perspectives and expertise, game-based approaches explore innovative pedagogical concepts, effective learning design, user experience optimisation, and assessment methodologies. This holistic understanding of learners' needs enables the development of contextually relevant and impactful interventions that resonate with diverse target groups.

Introducing game-based approaches such as serious games, game-based learning, and gamification is essential for enhancing learning experiences. Among these approaches, gamification as a technique, wraps game elements around serious contexts and activities for nudging engagement and sustain motivation to participate in activities, which is often non-content specific. Gamification leverages game techniques to facilitate, observe, and measure learning in non-game contexts (Deterding et al., 2011; Seaborn and Fels, 2015; Arnab, 2020), scaffolding and enhancing learners' experiences. It involves the use of video game elements to enhance user experience and engagement (Landers, 2014), as well as the incorporation of motivational affordances to evoke gameful experiences and behavioural outcomes (Hamari, Koivisto and Sarsa, 2014; Koivisto and Hamari, 2019). This approach focuses on two key aspects: the experiential aspect, which emphasises the gameful experience, and the game design aspect, which explores the design principles employed in gamification (Högborg, Hamari and Wästlund, 2019).

Gamification has gained significant traction in the education domain, including online learning and training contexts (Panigrahi and Srivastava, 2018; Khaldi, Bouzidi and Nader, 2023). Through interactivity, competition, rewards, and progression, gamification captivates learners' attention and sustains their engagement. The increasing digitalisation of education, accelerated by the COVID-19 pandemic, has highlighted the urgent need for innovative tools to enhance learner engagement and motivation in online learning environments. Traditional instructional methods often struggle to meet the expectations of modern learners, particularly in VET contexts, where practical, skills-based learning is crucial. Gamification plug-ins offer a compelling solution by embedding gamification characteristics—such as points, badges, challenges, and leaderboards—directly into existing Learning Management Systems (LMS) and Content Management Systems (CMS). This seamless integration allows educators to enhance the interactivity and functionality of established platforms without the need for extensive redesigns. Despite their potential, there is a significant gap in resources and evidence-based guidance to help educators and institutions make informed decisions about selecting and implementing these tools. This study addresses this critical gap by evaluating the usability, relevance, and effectiveness of gamification plug-ins, providing actionable insights to support their adoption and unlock their potential to transform digital learning experiences.

The use of plug-ins in gamification has historically been a double-edged sword, offering the potential to enhance user engagement while introducing significant challenges. In the past, technical compatibility issues, limited customisation, and resource-intensive designs often hampered their effectiveness, making gamification efforts feel clunky or irrelevant. Security vulnerabilities and poor user experiences further undermined trust and adoption. Today, these problems have evolved with the digital landscape, as privacy concerns, interoperability challenges, and the high cost of subscription-based plug-ins create new barriers. These persistent issues highlight the critical need for robust, seamless, and ethical plug-in solutions to fully unlock gamification's potential in engaging and motivating users.

In the past, plug-ins for gamification often suffered from technical compatibility issues. Many were not universally compatible across platforms, causing glitches or outright failures on certain devices or browsers.

Another common problem was the lack of user engagement; poorly designed plug-ins failed to align with user motivations or game mechanics, making them feel forced or irrelevant. Security vulnerabilities were also a significant concern, as some plug-ins introduced risks like data breaches or malware. Additionally, limited customisation options made it difficult for developers to tailor experiences to diverse audiences or unique use cases. Plug-ins were also frequently resource-intensive, slowing down websites or applications, which frustrated users.

Today, while some of these issues persist, new challenges have emerged. For instance, privacy concerns are more prominent now, with users wary of gamification systems tracking or misusing personal data. Interoperability across advanced systems remains a hurdle, especially as ecosystems become more complex with AI and machine learning integration. Lastly, the rise of subscription-based plug-ins has introduced cost barriers, limiting access for smaller developers or businesses. These evolving challenges necessitate a more thoughtful approach to gamification plug-ins.

Despite their challenges, plug-ins remain essential in gamification because they provide scalable solutions to implement game mechanics quickly and efficiently. They enable developers to add features like leaderboards, rewards, and progress tracking without building these systems from scratch, saving time and resources. Plug-ins also help standardise gamification elements across platforms, ensuring a consistent user experience. Moreover, many plug-ins come with pre-built analytics, allowing organisations to gather insights into user behaviour and optimise engagement strategies. Their adaptability to integrate with various tools and platforms makes them invaluable for creating customised and dynamic gamification solutions. While challenges like privacy and compatibility persist, the benefits of plug-ins in enhancing engagement and driving user motivation make them a critical component of gamified systems.

Plug-ins offer a myriad of features and functionalities tailored to enrich existing e-learning platforms, providing educators with the tools necessary to create dynamic and engaging learning environments. Among these plug-ins are gamification plug-ins, which serve as integral tools for improving modern e-learning platforms, i.e. Learning Management Systems (LMS) or Content Management Systems (CMS), by infusing them with elements of interactivity, motivation, and gamified learning experiences. Introducing gamification into the learning process not only creates an attractive environment for learners but also fosters skill acquisition, problem-solving, and group collaboration, resulting in a more effective educational process. In the present context, the use of gamification plug-ins opens up new perspectives for educational course developers, providing innovative tools to develop learners' skills in a way that is engaging and adapted to contemporary educational expectations. Gamification plug-ins compatible with LMS/CMS offer a promising avenue for enhancing the educational experiences of existing online learning platforms. However, their effective implementation requires a balanced approach that prioritises meaningful engagement, intrinsic motivation, and the educational objectives at hand.

With these perspectives, this paper explores and discusses the features of fifteen (15) gamification plug-ins, which can enrich the learning-teaching process. The authors understand a *gamification plug-in* as a software component designed to integrate gamification features into an existing system, platform, or application. The investigation was based on the following research questions:

- Why is it beneficial to integrate gamification plug-ins into educational material development?
- Which open-source gamification plug-ins are perceived as the most effective and user-friendly by VET trainers, learners, and stakeholders, and what features contribute to their perceived effectiveness within popular LMS/CMS platforms?

This study adopts a novel approach by reviewing gamification plug-ins with educators from five diverse European countries, offering a unique opportunity to examine how cultural, educational, and technological differences shape the perceived relevance and usability of these tools. Additionally, the focus on VET professionals introduces an innovative dimension, as most gamification research centers on traditional academic settings. This emphasis underscores the potential of gamification to address the specific needs of skills-based education effectively.

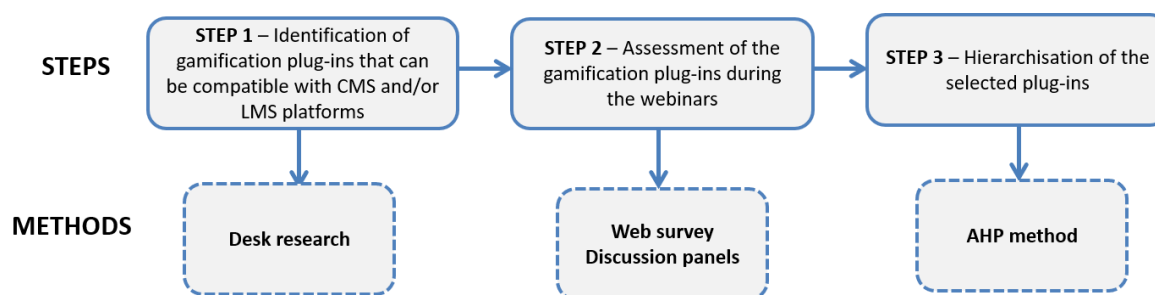
The study involved a total of 258 participants, recruited from five European countries: Poland, Italy, Greece, Cyprus, and the United Kingdom. The participants consisted of VET trainers, learners, and individuals with an interest in gamification in educational contexts. The selection process was based on purposive sampling, targeting stakeholders who have experience or interest in online learning and gamification. Recruitment took place through professional networks, educational institutions, and e-learning communities, ensuring that the

sample represented a diverse cross-section of individuals with practical insights into gamification in e-learning environments.

Data were collected using a mixed-methods approach, incorporating both quantitative and qualitative tools to gain comprehensive insights into the efficacy of gamification plug-ins. The primary instrument was a structured survey, which included both closed-ended questions to gather quantifiable data and open-ended questions for qualitative insights. The survey was supplemented by interviews with a subset of participants (n=50) to explore their experiences in greater depth.

## 2. Methodology

The investigation consisted of the following steps as illustrated in Figure 1.



Source: Authors.

**Figure 1: A flow diagram for the research methodology used**

*Step 1:* The initial phase involved systematically identifying gamification plug-ins compatible with CMS and/or LMS platforms. Organisations from Poland, Italy, Greece, Cyprus, and the United Kingdom were tasked with this endeavour, aiming to pinpoint at least 6 plug-ins per country. Leveraging a combination of literature review, available online documents, and website exploration, this month-long independent pursuit sought plug-ins meeting specific criteria: low-cost, free, or open-source. Out of the initial pool of thirty (30) plug-ins identified, only 15 selected were taken into account for further research, with duplicates eliminated across various countries.

*Step 2:* Subsequently, a thorough evaluation of the identified gamification plug-ins was conducted through a structured series of webinars. These sessions engaged a total of 258 participants from five countries, systematically designed to facilitate comprehensive plug-in assessments. The webinars were segmented into smaller cohorts, each accommodating approximately twenty (20) attendees. The process unfolded methodically: Participants with a keen interest in gamification and its educational applications were invited to the events. Plug-in presentations were delivered, featuring succinct overviews by representatives from each country, highlighting key features, functionalities, and unique benefits. Subsequent Q&A sessions allowed for nuanced exploration, enabling participants to seek clarification on plug-in capabilities and implementation details.

Participants conducted evaluations using pre-defined criteria (see Table 4) covering various aspects such as plug-in features, strengths, and weaknesses, installation guidelines, tutorials, and compatibility with LMS and CMS platforms, employing web surveys for assessment. Feedback and discussion panels further enriched the assessment process by facilitating an exchange of insights and opinions.

The categories from Table 4 were selected based on the analysis of the data collected from participants. During the survey and interview process, several themes emerged as key factors influencing the participants' evaluation of gamification plug-ins. These themes were consistently mentioned by multiple respondents, which led to the categorisation of the plug-ins under the following five areas:

- *How-to/Tutorial:* Many participants expressed the need for clear and accessible guidance on how to use the plug-ins. This category was included to assess the availability and quality of tutorials or how-to guides provided by the plug-ins, as this directly impacted usability.
- *Guidelines for Installation and Usage:* Ease of installation and initial configuration were highlighted as critical factors in the adoption of a plug-in. Participants emphasised the importance of having step-by-step instructions to streamline the implementation process.

- *Compatibility with LMS and CMS Systems*: One of the primary objectives of the study was to identify plug-ins that work seamlessly with LMS and CMS. This category was included to evaluate whether the plug-ins were easily integrated into existing educational platforms, a key concern for participants.
- *Strong and Critical Aspects*: During the interviews, participants often discussed both the strengths and limitations of the plug-ins they had experience with. This category allowed for a balanced evaluation of each tool, reflecting the participants' nuanced perspectives on what worked well and what could be improved.
- *Features*: Participants were particularly focused on the functionality offered by each plug-in. This category examines the specific features that differentiate one plug-in from another, helping to identify which tools provided the most value for enhancing learner engagement.

These categories were not arbitrarily selected; they emerged naturally from the data as key factors influencing participants' decisions and preferences.

During the discussions, the focus on 'Strong and Critical Aspects' aided the target groups in understanding the plug-in's key benefits and functionalities, while also addressing potential concerns or limitations transparently. 'Clear Guidelines for Installation and Usage' were provided to ensure a smooth user experience, guiding target groups through the process of installing and configuring the plug-in effectively. Detailed 'How-to/Tutorial sections' were offered to enhance participants' understanding and proficiency, providing practical resources for learning how to utilise the plug-ins' features effectively.

Furthermore, discussions on 'Compatibility with LMS and CMS platforms' seek to assess seamless integration and interoperability within existing digital ecosystems. This discussion enabled target groups to confidently deploy the plug-in within their preferred platforms, demonstrating its versatility and adaptability.

Additionally in Step 2, participants ranked the top three plug-ins they deemed most relevant and necessary for their work and practice, providing valuable insights into their preferences and needs.

*Step 3*: Building upon the results of Step 2, the features of the plug-ins underwent hierarchical prioritisation employing the Analytic Hierarchy Process (AHP) method. This methodological approach enables the delineation of crucial elements when selecting plug-ins for the incorporation of interactive content into educational materials. AHP's structured framework, hierarchical representation, pairwise comparisons, mathematical consistency, flexibility, adaptability, and sensitivity analysis capabilities render it a highly apt method for hierarchisation in decision-making processes across diverse domains, including project management, resource allocation, risk assessment, and strategic planning.

### 3. Findings and Discussions

In this section, we present the culmination of an extensive review process aimed at identifying and evaluating gamification plug-ins compatible with CMS and LMS platforms. Through the multi-step process (desk research, web survey, discussion panels, and the use of the AHP method), fifteen gamification plug-ins have been identified, assessed, and hierarchised. The chosen plug-ins offer unique features and functionalities aimed at promoting learner engagement, facilitating communication between educators and learners, and fostering a gamified learning environment. Among these plug-ins are *Class Dojo*, a behaviour management system designed to promote positive behaviours and communication between teachers and parents, and *WordPress (WP) Achievements*, a WordPress plugin enabling the creation and management of user achievements, quests, and ranks. Additionally, plug-ins like *Level Up!* and *GamiPress* introduce elements of gamification directly into course content, allowing teachers to reward learner progress and track their achievements. Other notable plug-ins include *Kahoot!* for interactive quizzes and games, *Bookwidgets* for creating custom interactive exercises, and *HD Quiz* for building professional questionnaires and quizzes. Each plug-in offers unique functionalities, catering to diverse educational needs and preferences, and holds promise in transforming online learning environments into engaging and interactive spaces for learners and educators alike. Table 1 further describes the gamification plug-ins.

The commonality in the research lies in the overarching goal of enhancing the teaching and learning process through gamification technologies. By focusing on LMS/CMS-compatible plug-ins, the study prioritises tools that align with institutional infrastructures while still addressing the broader aim of fostering engagement, interactivity, and learning outcomes across various educational contexts.

Table 1: The descriptions of the selected gamification plug-ins

| No | Name            | Short description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|----|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | Class Dojo      | An online behaviour management system intended to foster positive learner behaviours and classroom culture. Learners earn 'Dojo Points' based on their classroom conduct. Class Dojo's primary goal is to encourage positive learner behaviours and to provide a means for teachers and parents to communicate frequently and effectively about learner development. Teachers use Class Dojo to promote positive behaviours in their classroom and to strengthen lines of communication between school and home. This is done primarily through Class Dojo's 'Dojo Point' system and messaging system that connects teachers and parents. Class Dojo is fully accessible on computers, tablets, smartphones, and can even be used on interactive whiteboards and projectors.                                                                                                                  |
| 2  | WP Achievements | <p>WP Achievements is a powerful WordPress Achievements, Quests &amp; Ranks Plugin. It is a perfect extension for WordPress-powered website to improve the user's experiences and increase user interactivity. With WP Achievements it is possible to create and manage user achievements, quests, and ranks with ease.</p> <p>WP Achievements provides: to add Achievements &amp; Quests to website for a wide range of activities; to reward users with points when they gain Achievements &amp; Quests; to restrict content by gained achievements, solved quests, or user ranks; to publish to a user's BuddyPress components when they gain Achievements &amp; Quests; to add and manage ranks to your website; to limit Achievements to specific Ranks; to publish to a user's BuddyPress components when they gain a new Rank; to lock content so only specific Ranks can view it.</p> |
| 3  | Level Up!       | Level up! is a customisable block that a teacher can add to a course to give experience points to learners as they progress through a course. It displays their current level and progresses towards the next level. It adds an element of gamification to a course. Exemplary features: automatically attributes points to learners for their actions; block that displays current level and progress towards next level; report for teachers to get an overview of their learners' levels; notifications to congratulate learners as they level up; a leaderboard to display the ranking of the learners; ability to customise the number of levels, the points they require, and their appearance.                                                                                                                                                                                         |
| 4  | GamiPress       | GamiPress is a free WordPress gamification plugin. It enables the users to incorporate features common in gameplay into websites, such as point systems and competition between users. GamiPress is the easiest and the most effective WordPress plugin to gamify whatever is needed on the website within a few minutes. It allows the users to reward your users with different kinds of awards and badges for interacting on the website. It is also possible to easily illustrate the achievements, organise the requirements, and select from a wide range of customisable assessment options to check if each requirement has been achieved successfully or not. Thanks to GamiPress it is possible to use gamification to incentivise any kind of interaction with the site. GamiPress enables the users to create three types of digital rewards – points, achievements, and ranks.   |
| 5  | Kahoot!         | Kahoot!'s platform is used by educators around the world to make learning fun, vibrant, inclusive and engaging. With Kahoot!, anyone can effortlessly create, host, and participate in dynamic games and quizzes from their computers or mobile devices. Whether in a group setting or through the app's versatile player modes, Kahoot! serves as a conduit for interactive learning. Users can swiftly create and share quizzes, fostering an environment where learning is both enjoyable and accessible.                                                                                                                                                                                                                                                                                                                                                                                  |
| 6  | Bookwidgets     | Bookwidgets is a tool that helps to develop interactive educational resources for the classroom. Bookwidgets enables the users to easily create custom material that will work for the class. It allows teachers to create interactive exercises and automatically grade tests in minutes. It is possible to choose from digital exercise templates that work on smartphones, tablets, and computers; transform paper tests to interactive quizzes or worksheets that work on iPads, Chromebooks or any other device.                                                                                                                                                                                                                                                                                                                                                                         |
| 7  | HD Quiz         | HD Quiz is a highly user-friendly plugin that empowers users to effortlessly create an infinite array of quizzes and seamlessly embed them into any page or post. Whether crafting professional questionnaires or entertaining BuzzFeed-style quizzes, HD Quiz offers unparalleled versatility. Its key features include the ability to generate an unlimited number of quizzes, each with customizable options. Users can enhance questions by adding featured images and tips, or even incorporate animated gifs for added visual appeal. Moreover, the plugin boasts a quiz Timer feature, enabling users to impose time limits for completing quizzes. Additionally, basic translation features facilitate global accessibility, while the option to add links or images to quiz results provides enhanced customisation.                                                                 |
| 8  | Quiz Cat        | Quiz Cat is the easiest way to build viral and engaging quizzes for WordPress sites. It takes just a few minutes to set up a knowledge test, trivia quiz or viral personality quiz. Main features: multiple choice answers can be added to each question; ability to add text and images to questions and answers, which makes quizzes more engaging; quiz                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |

| No | Name           | Short description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|----|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                | results are displayed on quiz completion, together with the 'share' button enabling them to be posted to social media.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| 9  | Jigsaw Planet  | Jigsaw Planet is a website that contains free digital jigsaw puzzles that can be assembled using any device with a web browser. With the website, users can upload their own images and create a digital jigsaw puzzle that can be shared with others. Main features: ability to select a jigsaw puzzle format and the number of pieces; Jigsaw puzzles can be published on the website or blog without the necessity to register; ability to upload own images and create digital jigsaw puzzles that can be used for educational purposes; ability to share with learners the link to previously created digital jigsaw puzzles.                                                                                                                                                                                                                                                                          |
| 10 | Sticky Notes   | Sticky Notes is a Moodle plugin designed to facilitate various tasks, including brainstorming, pedagogic exercises, ranking, meetings, and idea organisation. This user-friendly activity enables users to construct a virtual post-it wall where they can create, update, and reposition sticky notes across different columns, reminiscent of a kanban board. Additionally, users have the option to customise the background color of sticky notes and participate in voting processes to prioritise or evaluate ideas.                                                                                                                                                                                                                                                                                                                                                                                  |
| 11 | StudentQuiz    | StudentQuiz enables students to collaboratively create their own question pools in Moodle. Students can filter these questions into quizzes, and they can rate and comment questions while working through the quizzes. StudentQuiz computes each student's contribution and learning progress and compares this with the community. The created questions become part of the Moodle question bank and can be reused in other Moodle quizzes.                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 12 | Stash          | The Stash block serves as a highly effective tool for fostering increased engagement with course activities, proving indispensable for educators seeking to gamify their teaching approach. This plugin introduces a block that displays items acquired by learners throughout the course, incentivising their participation and exploration. Teachers have the flexibility to create and distribute items within various activities and resources, encouraging learners to actively seek and collect them. These items can be configured for one-time collection, stimulating exploration of course content. Furthermore, educators can designate certain items to have an unlimited supply from specific locations, encouraging learners to revisit these areas for further collection, thereby promoting sustained interaction with course materials.                                                    |
| 13 | GAME           | This is a moodle plugin that gets input from quiz, glossary or questions and converts them into games automatically. The games are: Hangman; Crossword; Cryptex; Millionaire; Sudoku; Snakes and Ladders; The hidden picture; Book with questions.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| 14 | Real Time Quiz | The Real Time Quiz allows the users to create a quiz where each question is presented at the same time to learners. The teacher or administrator can, among others: create a quiz with a list of questions; set the time given to answer a question (30 seconds is the default); view statistics about the percentage class / individual correct answers.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| 15 | Block Game     | The Block Game plugin aims to seamlessly integrate gamification techniques into the Moodle platform with ease and efficiency. This plugin offers a range of gamification features, including the ability to customise avatars, whereby users can select or change their avatar at any time. Additionally, it provides access to player information, showcasing an overview of a user's progress across various courses within the platform. Users can also view their ranking relative to other players and engage with a scoring system that determines their rating and current level. Moreover, the plugin offers the option to display a level system, allowing administrators to configure the number of levels and the score required to progress through each level. Furthermore, a progress bar indicates the user's advancement towards the next level, enhancing their overall gaming experience. |

Source: IO3 - i-CONTENT Gamification A Toolkit for the gamification of an online course.

In response to the research question, which aimed to identify the most appropriate, user-friendly, and effective open-source gamification plug-ins, we collected both qualitative and quantitative data from participants. The analysis of this data revealed the specific reasons behind the selection and prioritisation of the plug-ins listed below.

Rather than presenting only technical descriptions, we explored the reasons why participants—VET trainers, learners, and educational stakeholders—favored certain plug-ins over the others. These insights were derived from both survey responses and in-depth interviews, where participants articulated their preferences based on criteria such as usability, learner engagement, compatibility with existing LMS/CMS platforms, and ease of integration.

Throughout the interviews (n=50), participants explained their rationale for prioritising these plug-ins. For example, *Kahoot* gamifies learning, making lessons fun and engaging for students. Its user-friendly interface allows educators to create interactive quizzes and access a vast library of ready-made resources, saving time on preparation. Additionally, Kahoot fosters active participation, collaboration, and real-time feedback, enhancing the overall learning experience.

*Sticky Notes* are a versatile tool for teachers, allowing them to quickly jot down ideas, reminders, or feedback in an organised and accessible way. They can be used for lesson planning, tracking student progress, or highlighting key points during instruction. Their simplicity and flexibility make them ideal for fostering creativity, collaboration, and efficient classroom management.

Below are response samples:

*"I really enjoyed how Level Up! added a sense of progression to the tasks. Earning experience points made routine activities feel more rewarding. However, I sometimes found it hard to balance the rewards with the actual goals—it's easy to chase badges and lose sight of priorities."*

*"I appreciated how BookWidgets allowed me to create interactive assignments. It made learning feel hands-on and fun. That said, it took some time to get used to the interface, and some advanced features weren't as intuitive as I'd hoped."*

*"WP Achievements integrated seamlessly with my existing WordPress setup. It was satisfying to see users engage more with the site because of the reward system. My only critique is that the achievements sometimes felt too generic—I wish there were more tailored options out of the box."*

The plug-ins were positively assessed. Some of them are used more often than the others, but one of the most important aspects for the participants was that they have choice which plug-in to choose for an individual lesson.

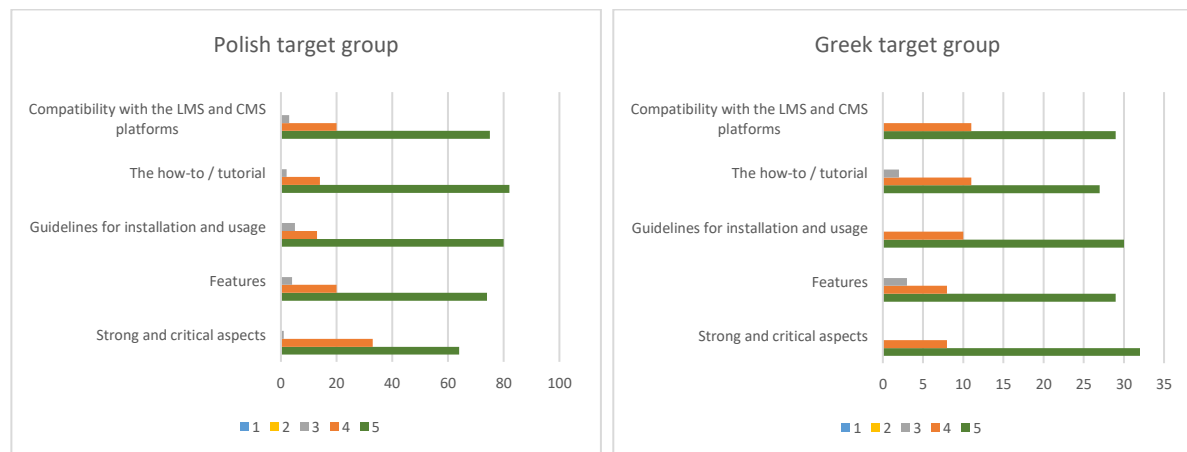
The plug-ins were then reviewed with VET trainers, students (mainly IT and pedagogics faculties), and other interested individuals (especially HR representatives of the companies) across multiple European countries (n=258, Table 2).

**Table 2: Participant demography**

| Country                       | Poland | Greece | Italy | Cyprus | The UK |
|-------------------------------|--------|--------|-------|--------|--------|
| Number of respondents         | 98     | 40     | 40    | 40     | 40     |
| <b>Type of respondents</b>    |        |        |       |        |        |
| VET trainers                  | 44     | 30     | 25    | 21     | 21     |
| Students                      | 47     | 0      | 8     | 15     | 15     |
| Other interested in the topic | 7      | 10     | 7     | 4      | 4      |

Source: Authors.

As depicted in Figure 2, all elements were rated at either 5 or 4 on the scale, indicating their significance in the selection of each individual plug-in.



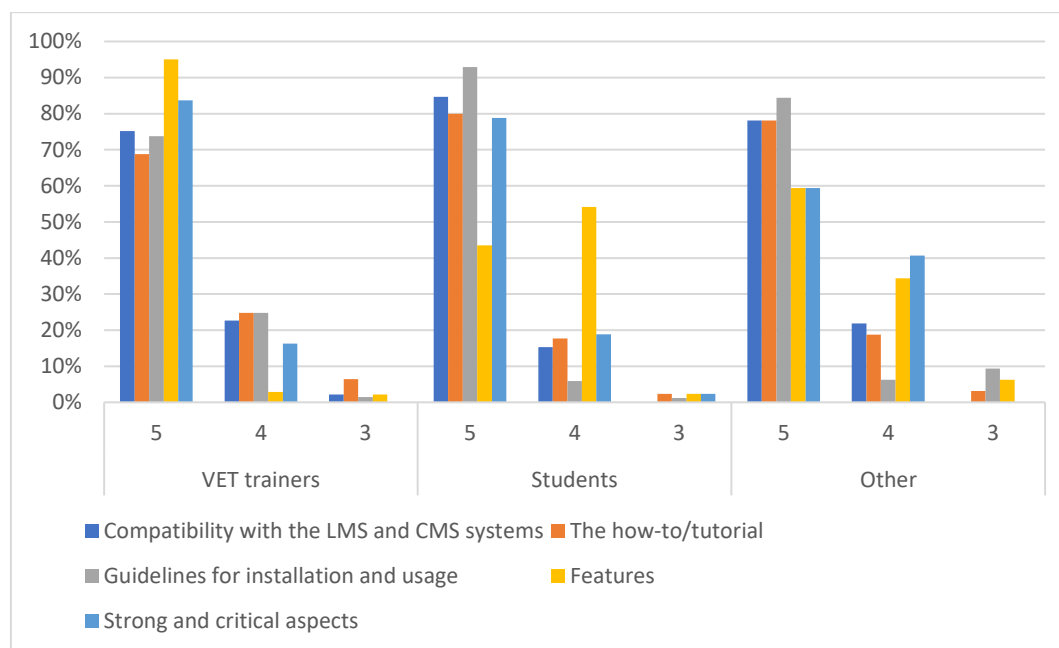


Source: Authors.

**Figure 2: Overall evaluation of selected plug-ins by the participants from five European countries**

The importance of various elements varies depending on the specific needs, goals, and preferences of different target groups. For instance, in Poland, participants placed significant emphasis on aspects related to "how-to/tutorial" (82) and "guidelines for installation and usage" (80). Similarly, the Italian group prioritised "guidelines for installation and usage" (34), while the Greek, UK, and Cypriot groups highly valued "strong and critical aspects" (34). Interestingly, unlike the Polish group, other groups rated "how-to/tutorial" as the least important aspect.

Figure 3 presents a comparative analysis across participant types, including VET trainers, students, and other individuals interested in the topic, represented in relative percentages. VET trainers showed the strongest assessment toward plug-in features and identified both strengths and critical aspects. However, they placed less emphasis on "how-to/tutorial" compared to other aspects. In contrast, students rated "guidelines for installation and usage" and "compatibility with LMS and CMS systems" most highly, although they still valued "features" relatively highly. Overall, examining responses across different target groups reveals consistent high appraisal of elements associated with identified plug-ins.



Source: Authors.

**Figure 3: Evaluation of selected plug-ins across the participants types [in %]**

Participants were also requested to specify their preferred individual plug-ins in a two-step activity. Initially, they rated each plug-in on a scale from 5 (excellent) to 1 (poor) based on its perceived value. With this data collected, the authors determined which plug-ins received an average rating above 4.0: *WP Achievements*, *Level Up!*, *GamiPress*, *Kahoot!*, *BookWidgets*, *GAME*, and *BlockGame*. In the second step, participants were asked to select their top three plug-ins. The results are presented in Table 3.

**Table 3: Combined frequency of times software is collectively ranked in top 3**

| Plug-ins with the rate above 4.0 | Frequency (number of times) | Relative Frequency [%] |
|----------------------------------|-----------------------------|------------------------|
| <b>WP Achievements</b>           | 98                          | 13                     |
| <b>Level Up!</b>                 | 154                         | 20                     |
| <b>GamiPress</b>                 | 190                         | 24                     |
| <b>Kahoot!</b>                   | 137                         | 18                     |
| <b>BookWidgets</b>               | 76                          | 10                     |
| <b>GAME</b>                      | 84                          | 11                     |
| <b>BlockGame</b>                 | 35                          | 4                      |

Source: Authors.

According to the target groups, *GamiPress* emerged as the most relevant plug-in, offering a comprehensive suite of gamification features, including points, achievements, and ranks. Its versatility enables site owners to craft engaging experiences tailored to their specific needs. With active maintenance and regular updates, *GamiPress* ensures compatibility with the latest *WordPress* versions and promptly addresses reported issues. Moreover, a supportive community with forums and resources enhances user experience, while its user-friendly interface accommodates varying levels of technical expertise.

Following closely is *Level Up!*, which encourages increased interaction with websites through task completion, discussions, and content consumption. This heightened engagement fosters higher retention rates and a vibrant community. *Level Up!* seamlessly integrates with other *WordPress* plugins and platforms, ensuring smooth operation with regular updates to maintain compatibility.

*Kahoot!* ranks as the third most preferred plug-in, offering educators a fun and interactive platform for engaging learners in quizzes, surveys, and discussions. Its accessibility across devices facilitates convenient participation

from anywhere with internet access. Real-time feedback during activities aids educators in assessing learner understanding instantly, enabling timely intervention where needed.

The feedback items and elements underwent further analysis using the AHP method to establish a hierarchy. The results revealed that three elements were deemed crucial by the target groups: 'strong and critical aspects,' 'compatibility with LMS and CMS platforms,' and 'guidelines for installation and usage' (see Table 4).

**Table 4: The hierarchisation of the analysed elements related to the gamification plug-ins**

| Preference matrix for criteria                                   | C1      | C2      | C3      | C4      | C5      | $\bar{K}_g$ | Eigenvector |
|------------------------------------------------------------------|---------|---------|---------|---------|---------|-------------|-------------|
| C1: Strong and critical aspects                                  | 1       | 3       | 3       | 5       | 5       | 1.719       | 0.305       |
| C2: Features                                                     | 1/3     | 1       | 1/3     | 4       | 3       | 1.029       | 0.183       |
| C3: Guidelines for installation and usage                        | 1/3     | 3       | 1       | 5       | 1/2     | 1.096       | 0.194       |
| C4: The how-to / tutorial                                        | 1/5     | 1/4     | 1/4     | 1       | 1/4     | 0.562       | 0.100       |
| C5: Compatibility with the LMS and CMS platforms                 | 1/5     | 5       | 2       | 4       | 1       | 1.231       | 0.218       |
| Suma                                                             | 2.067   | 12.250  | 6.583   | 19.000  | 9.750   | 5.637       | 1.000       |
| Method 1 (with calculation of $\lambda_m$ for consecutive lines) |         |         |         |         |         |             |             |
| Matrix product                                                   | 3.02627 | 1.40286 | 1.45126 | 0.30949 | 1.97977 |             |             |
| $\lambda_m$                                                      | 9.9247  | 7.68332 | 7.46414 | 3.10589 | 9.06429 |             |             |
| $\lambda_m$ in                                                   | 7.44847 |         |         |         |         |             |             |
| CI                                                               | -0.2835 |         |         |         |         |             |             |
| CR (RI=1.49 dla n=10)                                            | -0.1903 | <10%    |         |         |         |             |             |
| Metoda 2 (indication of $\lambda_m$ for the whole matrix)        |         |         |         |         |         |             |             |
| $\lambda_m$                                                      | 8.16966 |         |         |         |         |             |             |
| CI                                                               | -0.2034 |         |         |         |         |             |             |
| CR (RI=1.49 dla n=10)                                            | -0.1365 | <10%    |         |         |         |             |             |
| Normalised matrix                                                | C1      | C2      | C3      | C4      | C5      | average C   |             |
| C1: Strong and critical aspects                                  | 0.484   | 0.245   | 0.456   | 0.263   | 0.513   | 0.392       |             |
| C2: Features                                                     | 0.161   | 0.082   | 0.051   | 0.211   | 0.308   | 0.162       |             |
| C3: Guidelines for installation and usage                        | 0.161   | 0.245   | 0.152   | 0.263   | 0.051   | 0.175       |             |
| C4: The how-to / tutorial                                        | 0.097   | 0.020   | 0.038   | 0.053   | 0.026   | 0.047       |             |
| C5: Compatibility with the LMS and CMS platforms                 | 0.097   | 0.408   | 0.304   | 0.211   | 0.103   | 0.224       |             |
| Criterion weight value                                           | 0.392   | 0.162   | 0.175   | 0.047   | 0.224   |             |             |
| Normalized values of weights<br>scope (1 ÷ 5)                    | 5.0     | 2.1     | 2.2     | 0.6     | 2.9     |             |             |
| Proposed weights                                                 | 5       | 2       | 3       | 1       | 4       |             |             |

Source: Authors.

Table 4 presents a structured Analytic Hierarchy Process (AHP) analysis to prioritise five criteria (C1 to C5) for decision-making, with a focus on evaluating preferences and ensuring consistency in judgments. The preference matrix shows pairwise comparisons among criteria, where higher values indicate stronger relative importance. From these comparisons, eigenvectors (priority weights) are calculated, with C1 (Strong and critical aspects) emerging as the most important criterion, followed by C5 (Compatibility with LMS and CMS platforms). Consistency checks using Consistency Index (CI) and Consistency Ratio (CR) confirm that judgments are reliable, as both CR values fall below the 10% threshold, indicating consistent assessments. The normalised matrix further refines each criterion's weight by averaging its relative importance across comparisons. Finally, these weights are scaled to a 1–5 range, producing proposed weights that highlight C1 and C5 as primary decision factors, with secondary emphasis on features (C2) and guidelines (C3), while tutorials (C4) hold minimal influence. This structured weighting guides decision-makers in prioritising essential factors, promoting a balanced and reliable decision outcome.

The eigenvector in the preference matrix is crucial for calculating ranks and determining which options are preferred, which supports the decision-making process. It allows to mathematically establish a hierarchy of preferences and check that the evaluations are consistent, which is invaluable in multi-criteria analyses.

The analysis in this table provides clear insights into the relative importance of each criterion, guiding decision-makers on where to focus their attention. The Strong and critical aspects (C1) are identified as the most crucial factor, reflecting the need to prioritise robust and essential features in whatever solution or tool is being evaluated. This is followed by Compatibility with LMS and CMS platforms (C5), indicating that seamless integration with existing systems is also highly valued. Together, these two criteria carry the most weight, suggesting that they should be the primary focus when evaluating options.

The Features (C2) and Guidelines for installation and usage (C3) hold moderate importance, suggesting that while desirable, they are less critical than C1 and C5. Thus, these factors should influence the decision but can be considered secondary priorities. The how-to/tutorial (C4) is the least significant criterion, implying that the presence of user guides or tutorials, while helpful, should not strongly impact the decision relative to the other factors.

The Consistency Ratios (CR) for both evaluation methods are below the 10% threshold, confirming that the pairwise comparisons made in the preference matrix are consistent. This consistency adds credibility to the results, as it indicates that the judgments used to prioritise criteria are reliable and not arbitrary.

In summary, this analysis suggests a decision-making strategy that prioritises core functionality and compatibility with existing systems (C1 and C5), while still considering features and usability (C2 and C3) as valuable but secondary. Tutorials (C4) should be viewed as a minor factor. This structured approach enables decision-makers to focus on what truly matters, facilitating a well-balanced and effective choice.

The strong and critical aspects of gamification plug-ins play a pivotal role in shaping engaging, effective, and sustainable learning experiences that empower learners and align with educational objectives. Participants consistently identified these elements as crucial, underscoring their significance in understanding the potential benefits and considerations associated with each plug-in. Prioritising these aspects enables educators to unlock the full potential of gamification, thereby enhancing teaching and learning outcomes.

Integration with Learning Management Systems (LMS) and Content Management Systems (CMS) emerges as a vital consideration for seamless implementation within educational environments. Compatibility ensures that gamified elements can seamlessly integrate into existing systems, thereby minimising technical barriers for educators and learners alike. Notably, this aspect garnered particular attention when evaluating individual plug-ins, notably *GamiPress*, *Kahoot!*, and *Level Up!*, which emerged as the most popular choices among participants.

Equally important is the provision of clear Guidelines for Installation and Usage, as educators and administrators require straightforward instructions for installing, configuring, and utilising plug-ins effectively. Accessible documentation, tutorials, and support resources are instrumental in facilitating smooth adoption and utilisation of gamified elements in educational settings. While this aspect received less attention in the evaluation of more obscure plug-ins such as *Stash*, *Sticky Notes*, or *Bookwidgets*, it remains essential for ensuring successful implementation and management.

#### **4. Conclusions**

This study set out to explore the integration of gamification plug-ins into educational material development, addressing two central research questions: (1) Why is it beneficial to integrate gamification plug-ins into educational material development? and (2) Which open-source gamification plug-ins are perceived as the most effective and user-friendly, and what features contribute to their effectiveness within popular LMS/CMS platforms?

The findings underscore the multifaceted benefits of gamification plug-ins. By incorporating game mechanics such as rewards, challenges, and interactivity, these tools significantly enhance learner engagement and motivation. They create immersive learning environments that foster skills such as problem-solving, critical thinking, and collaboration. Importantly, gamification also facilitates continuous feedback and progress tracking, allowing educators to tailor their strategies to the specific needs of learners. This aligns with the needs of vocational education and training (VET) professionals, addressing the skill-based and applied nature of their training environments.

In response to the second research question, the analysis of fifteen gamification plug-ins revealed that features such as ease of use, compatibility with existing LMS/CMS platforms, and clear installation guidelines are critical to their perceived effectiveness. Tools like *GamiPress*, *Level Up!*, and *Kahoot!* were identified as the most effective, owing to their versatility, user-friendliness, and capacity to create engaging and dynamic learning experiences. For example, *GamiPress*'s ability to reward learners with points and badges, coupled with its seamless integration into WordPress platforms, makes it highly adaptable for diverse educational contexts. Similarly, *Level Up!*'s gamified progression system fosters sustained engagement, while *Kahoot!*'s interactive quizzes promote collaboration and real-time feedback.

However, the study also highlights several challenges and areas for improvement. While gamification plug-ins enhance engagement, reliance on extrinsic motivators such as points and badges may risk undermining intrinsic motivation (Nicholson, 2012) if not carefully balanced with meaningful deep learning experiences (Reid, 2011).

Additionally, the study revealed varying priorities among different target groups, with VET trainers emphasising the importance of robust features and learners prioritising compatibility and usability. This suggests a need for more customisable solutions that can cater to diverse user needs.

From a critical perspective, the findings emphasise that the mere inclusion of gamification elements is insufficient to guarantee meaningful engagement or improved learning outcomes. Meaningful gamification requires thoughtful integration of play, challenge, and emotional resonance to sustain engagement and foster deeper learning (Reid, 2011; Nicholson, 2012; Arnab, 2020). Furthermore, the study was limited to the evaluation of plug-ins in controlled settings, and their real-world application within authentic educational contexts remains untested. Future research should focus on longitudinal studies to evaluate the long-term impact of these tools on learner outcomes and explore how gamification can be tailored to diverse cultural and educational contexts.

In conclusion, this study contributes to the growing body of literature on gamification in education by providing a systematic evaluation of gamification plug-ins. It highlights their potential to revolutionise digital learning environments while calling for a more nuanced and context-sensitive approach to their implementation. By addressing both the opportunities and limitations, the findings pave the way for more effective and impactful applications of gamification in education.

**AI Statement:** The authors state that Artificial Intelligence tool was not used in this study.

**Ethics Statement:** Ethical approvals have been obtained with precautions taken to ensure participants' informed consent and confidentiality.

## References

- Anderson, B. O., Anderson, M. N. and Taylor, T.A., 2009. New territories in adult education: Game-based learning for adult learners. *Adult Education Research Conference*. Retrieved May 10, 2024, from <https://newprairiepress.org/cgi/viewcontent.cgi?referer=&httpsredir=1&article=3747&context=aerc>.
- Arnab, S., 2020. *Game Science in Hybrid Learning Spaces*. New York: Routledge.
- Barneva, R.P., Kanev, K., Kapralos, B., Jenkin, M. and Brimkov, B., 2017. Integrating technology-enhanced collaborative surfaces and gamification for the next generation classroom. *Journal of Educational Technology Systems*, 45(3), pp. 309-325, <https://doi.org/10.1177/0047239516671945>.
- Chang, C.C. and Yang, S.T., 2023. Interactive effects of scaffolding digital game-based learning and cognitive style on adult learners' emotion, cognitive load and learning performance. *International Journal of Educational Technology in Higher Education*, 20(1), 16, <https://doi.org/10.1186/s41239-023-00385-7>.
- Deterding, S., Dixon, D., Khaled, R. and Nacke, L., 2011. *From game design elements to gamefulness: defining "gamification"*. In Proceedings of the 15<sup>th</sup> international academic MindTrek conference: Envisioning future media environments (pp. 9-15).
- Hamari, J., Koivisto, J. and Sarsa, H., 2014. *Does gamification work? – a literature review of empirical studies on gamification*. In 2014 47<sup>th</sup> Hawaii International Conference on System Sciences (pp. 3025-3034). IEEE.
- Höglberg, J., Hamari, J. and Wästlund, E., 2019. Gameful Experience Questionnaire (GAMEFULQUEST): an instrument for measuring the perceived gamefulness of system use. *User modeling and user-adapted interaction*, 29(3), pp. 619-660, <https://doi.org/10.1007/s11257-019-09223-w>.
- Huang, W. H. Y. and Soman, D., 2013. *Gamification of education. Report Series: Behavioural Economics in Action*, Rotman School of Management, University of Toronto.
- Khalidi, A., Bouzidi, R. and Nader, F., 2023. Gamification of e-learning in higher education: a systematic literature review. *Smart Learning Environments*, 10, p. 10, <https://doi.org/10.1186/s40561-023-00227-z>.
- Koivisto, J. and Hamari, J., 2019. The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, 45, pp. 191-210, <https://doi.org/10.1016/j.ijinfomgt.2018.10.013>.
- Kolade, O. and Owoseni, A., 2022. Employment 5.0: The work of the future and the future of work. *Technology in Society*, 71, p. 102086, <https://doi.org/10.1016/j.techsoc.2022.102086>.
- Landers, R.N., 2014. Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & Gaming*, 45(6), pp. 752-768, <https://doi.org/10.1177/1046878114563660>.
- Nicholson, S., 2012. *Strategies for meaningful gamification: Concepts behind transformative play and participatory museums*. Meaningful Play 2012. Lansing, Michigan. Retrieved May 10, 2024, from <http://scottnicholson.com/pubs/meaningfulstrategies.pdf>.
- Panigrahi, R. and Srivastava, P.R., 2018. Understanding the motivation in massive open online courses: A Twitter mining perspective. *International Journal of Web Based Communities*, 14(3), pp. 228-248, <https://doi.org/10.1504/IJWBC.2018.094915>.
- Reid, A. (2011). Welcome to badge world. Retrieved May 10, 2024, from <https://profalexreid.com/2011/09/15/welcome-to-badge-world/>.

- Seaborn, K. and Fels, D.I., 2015. Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, 74, pp. 14-31, <https://doi.org/10.1016/j.ijhcs.2014.09.006>.
- Tay, J., Goh, Y. M., Safiena, S. and Bound, H., 2022. Designing digital game-based learning for professional upskilling: A systematic literature review. *Computers & Education*, 184, p. 104518, <https://doi.org/10.1016/j.compedu.2022.104518>.

# Determinants of Student Adoption of Generative AI in Higher Education

Hanadi Aldreabi<sup>1</sup>, Nisreen Kareem Salama Dahdoul<sup>2</sup>, Mohammad Alhur<sup>3</sup>, Nidal Alzboun<sup>4,5</sup> and Najeh Rajeh Alsalihi<sup>6</sup>

<sup>1</sup>Department of Journalism, Media, and Digital Communication, School of Arts, The University of Jordan, Amman, Jordan

<sup>2</sup>Faculty of Arts & Science, Applied Science Private University, Amman, Jordan

<sup>3</sup>Faculty of Business, Al-Zaytoonah University, Amman, Jordan

<sup>4</sup>School of Arts, The University of Jordan, Amman, Jordan

<sup>5</sup>College of Arts, Humanities, and Social Sciences, University of Sharjah, UAE

<sup>6</sup>Department of Education, College of Arts, Humanities, and Social Sciences, University of Sharjah, UAE

[H.aldreabi@ju.edu.jo](mailto:H.aldreabi@ju.edu.jo)

[n\\_dahdoul@asu.edu.jo](mailto:n_dahdoul@asu.edu.jo)

[mohammasalemahmad.alhur@rai.usc.es](mailto:mohammasalemahmad.alhur@rai.usc.es)

[Nal-zboun@sharjah.ac.ae](mailto:Nal-zboun@sharjah.ac.ae)

[nalsalihi@sharjah.ac.ae](mailto:nalsalihi@sharjah.ac.ae)

<https://doi.org/10.34190/ejel.23.1.3599>

An open access article under [CC Attribution 4.0](#)

**Abstract:** The examination of the impact of Generative AI (GenAI) on higher education, especially from the viewpoint of students, is gaining significance. Although prior research has underscored GenAI's potential advantages in higher education, there exists a discernible research gap concerning the determinants that affect its adoption. In the present study, we aim to enhance our comprehension of the factors influencing the willingness of higher education students to adopt GenAI tools. To achieve this, we have developed an extended Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) model incorporating specific GenAI constructs. Our research methodology entailed the selection of a diverse sample of 374 students through random sampling. We then analyzed their data using Structural Equation Modeling (SEM) to gain insights into the complex relationships between various variables. The study found that students are more likely to use GenAI tools when they view them as supplemental resource and effort expectancy. It also revealed that perceived costs negatively impact adoption intentions, highlighting that financial factors are a significant barrier. Interestingly, Factors like information accuracy and hedonic motivation did not significantly affect students' adoption intentions. This study offers key insights for eLearning practitioners on integrating Generative AI (GenAI) tools into educational settings. It emphasizes the significance of resource perception and effort expectancy, demonstrating GenAI's potential to personalize learning experiences. eLearning platforms can utilize GenAI to enhance active learning through engaging methods and streamline course development. Addressing cost barriers is crucial for equitable access and inclusivity. A gradual approach to integration aligned with learning objectives is recommended, along with fostering critical engagement with GenAI tools to enhance digital literacy. Lastly, the study is constrained by its specific context, potential biases in self-reported data, a narrow focus on factors influencing students' intent to use GenAI tools and a cross-sectional design. Future research should encompass a broader range of factors, employ objective measures, and integrate observational data. Longitudinal studies or experimental designs could offer more comprehensive insights into how students' perceptions and intentions develop, thus promoting a more inclusive educational environment for all students.

**Keywords:** Generative AI, Unified Theory of Acceptance and Use of Technology 2 (UTAUT2), Adoption technologies, Higher education

## 1. Study Background

In recent years, the rapid advancement of artificial intelligence (AI), particularly generative AI (GenAI), has revolutionized various fields, including education (Bahroun *et al.*, 2023). The emergence of GenAI has sparked widespread interest among students, educators, researchers, and educational institutions globally due to its significant impact on teaching and learning (Faisal Rashid, Duong-Trung and Pinkwart, 2024). GenAI represents a sophisticated technology that leverages deep learning models to generate content that closely resembles human responses to complex prompts. Its ongoing evolution is expected to drive innovation and improvements in higher education, while also presenting new challenges (Michel-Villarreal *et al.*, 2023).

Multiple types of research have showcased the great potential of GenAI technology in education (for instance, Perera and Lankathilake, 2023; Tafazoli, 2024; Wang *et al.*, 2024). This technology can transform the conventional learning experience by offering personalized learning opportunities and adapting the educational content to cater to student's needs and abilities.

Furthermore, it promotes collaboration and peer interaction by producing contextually relevant prompts and responses, resulting in a dynamic learning environment that enhances student engagement and understanding (Chan and Zhou, 2023a).

GenAI technology can greatly improve personalized learning experiences by leveraging artificial intelligence (AI) and machine learning (ML) techniques to adapt educational activities based on student's preferences, backgrounds, and requirements (Maghsudi *et al.*, 2021; Fernandes, Rafatirad and Sayadi, 2023). By employing GenAI methods, educational platforms can accurately capture students' characteristics, recommend suitable content, develop customized curricula, and facilitate effective learner connections, ultimately enhancing performance evaluation and motivation for learning (Maghsudi *et al.*, 2021).

Additionally, the integration of AI and ML in personalized learning environments enables the continual refinement of unique profiles for individual students through learning data analytics, deep learning, and explainable AI, ensuring a more personalized and effective learning experience (Shawky and Badawi, 2019; Montebello, 2021).

GenAI technology is poised to significantly impact higher education by automating regular tasks, enhancing productivity, and creating new types of work and industries (Chan and Colloton, 2024). While students generally have a positive attitude towards GenAI in teaching and learning, recognizing its potential for personalized support and research capabilities (Chan and Hu, 2023), challenges persist. Universities exhibit significant variation in policies regarding GenAI use, with only a third having implemented specific guidelines (Xiao, Chen and Bao, 2023). Concerns include issues of academic integrity, ethical dilemmas, accuracy, privacy, and the potential transformation or obsolescence of certain jobs due to the continuous evolution of GenAI tools (Chan and Hu, 2023; Alier, García-Peñalvo and Camba, 2024).

In conclusion, effectively addressing these challenges requires a balanced approach leveraging GenAI benefits while mitigating its potential negative impacts on education and society (Arantes, 2024). This study examines factors influencing students' adoption of GenAI tools in higher education using a modified Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) model. The results show that supplemental resource and effort expectancy significantly and positively impact students' intent to use GenAI tools. At the same time, information accuracy and hedonic motivation do not significantly affect students' willingness to use these tools. This research enriches the UTAUT2 model by introducing new variables and provides practical implications for academic institutions.

## **2. Rationale of Study**

To fully leverage the potential of GenAI, it is imperative to shift our academic focus from bemoaning the challenges in education to understanding how students can effectively utilize such tools (Susarla *et al.*, 2023). An essential aspect of this endeavor is comprehending student perceptions and intentions (Chan and Zhou, 2023). Various studies highlighted the importance of exploring student perceptions and their willingness to embrace GenAI. By dissecting the link between these perceptions and usage intentions, we can gain valuable insight into how students interact with GenAI tools and how to tailor them to better meet student needs and preferences (Ivanov *et al.*, 2024).

It is also crucial to delve into the antecedents of adoption intention and actual usage of AI-based teacher bots, including perceived ease of use, usefulness, information accuracy, interactivity, cost, and perceived intelligence (Pillai *et al.*, 2024). This comprehensive exploration sheds light on the elements contributing to student acceptance of AI technologies, which is vital for developing engaging and effective GenAI tools (Alzahrani, 2023).

Ultimately, a profound understanding of these mechanisms can aid in designing and implementing GenAI tools that enhance educational outcomes. Aligning these tools with student needs and preferences can drive more personalized, interactive, and effective learning experiences.

### 3. Study Problems and Aims

Recently, there has been a growing focus on the impact of GenAI in higher education. However, there is a need for a comprehensive exploration of the personal and technological factors that influence users' intentions to utilize GenAI, including hedonism, usefulness, and supplemental resource. Existing research primarily addresses concerns related to academic integrity, potentially limiting student engagement with this transformative technology. Despite students' interest, there is a lack of thorough examination of their perspectives on incorporating GenAI into learning environments (Furze *et al.*, 2024). While previous studies have underscored GenAI's potential in higher education (McDonald *et al.*, 2024), there exists a notable research gap regarding the factors influencing its adoption (Gupta and Yang, 2024).

Understanding these adoption determinants is vital for developing tailored theoretical and practical frameworks to optimize GenAI platforms in education. Given that students are primary beneficiaries, our study aims to explore the diverse factors influencing their adoption of GenAI tools. While the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) model provides valuable insights into technology adoption, its application in educational contexts must be modified to be more suitable.

To address these gaps, our research proposes a modified UTAUT2 model that incorporates GenAI -specific characteristics. This approach aims to elucidate how elements like hedonic motivation, effort expectancy, and behavioral intention influence the adoption of GenAI tools among higher education students. Additionally, variables such as information accuracy, perceived cost, and the role of GenAI as a supplemental resource will be investigated to determine their impact on adoption behavior. By identifying reliable predictors of adoption, this study seeks to provide nuanced insights and practical recommendations for optimizing GenAI integration in higher education settings.

### 4. Study Questions

The current study aims to uncover the key drivers behind higher education students' adoption of Generative AI (GenAI) tools. By extending the UTAUT2 model with GenAI-specific components, we will delve into essential variables influencing adoption behaviors. Accordingly, we pose the following pivotal questions: How do specific factors of the UTAUT2 model, namely hedonic motivation, effort expectancy, and behavioral intention, influence the adoption of GenAI tools by higher education students? In addition, how do additional factors, specifically information accuracy, supplemental resource, and perceived cost, contribute to the adoption of GenAI tools by higher education students? Lastly, among these factors, which is the most dependable predictor of higher education students' adoption of GenAI tools?

### 5. Significance of the Study

This research is paramount for advancing the integration of GenAI tools in higher education. By examining the factors influencing students' overall experience and expanding the user base of GenAI in education, the study aims to enrich students' experience and promote wider adoption of GenAI in education. The anticipated results of this study are expected to bring substantial and far-reaching benefits for the effective implementation of GenAI in education.

The research's model offers a comprehensive understanding of the factors impacting GenAI adoption among higher education students, providing insights into how various factors collectively influence students' acceptance and use of GenAI tools. This study is instrumental in enhancing the localization and adaptation of GenAI design technology specifically for higher education students. By analyzing the factors influencing their adoption of this technology, we aim to improve the user experience and expand the current user base for GenAI tools, leading to a positive and extensive impact on the utilization of GenAI in higher education.

### 6. Literature Review and Theory Development

There has been a surge in the use and popularity of GenAI tools, which are being utilized in various fields, including education (Chan and Zhou, 2023a). Integrating these technologies in educational settings has transformed the learning landscape and revolutionized how students approach their studies (Mishra, Oster and Henriksen, 2024).

Recent research on the integration of GenAI in higher education suggests a generally positive reception among students (Chan and Hu, 2023). They acknowledge the benefits of personalized learning support, writing assistance, and enhanced research capabilities (Akyuz, 2020). However, concerns have been raised regarding

accuracy, privacy, ethical implications, and the potential impact on personal and societal development (Wach *et al.*, 2023). Given that student perceptions significantly influence learning approaches and outcomes; it is important to address their concerns to effectively incorporate GenAI tools in education (Chan and Hu, 2023).

Additionally, students' intention to use GenAI is influenced by information accuracy and cost, highlighting the importance of considering these factors in promoting adoption (Gupta *et al.*, 2024). Educators and students must be involved in assessment reform efforts to emphasize learning processes, critical thinking, and practical applications in the context of the evolving landscape of AI in education. (Pedro *et al.*, 2019; Alam, 2021).

However, for these technologies to be widely adopted, it is crucial to understand students' perceptions and the factors influencing their acceptance (Ivanov *et al.*, 2024). In this regard, The Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) model provides a detailed framework for examining how students adopt GenAI tools. It considers factors such as effort expectancy, social influence, hedonistic motivation, and facilitating conditions. This comprehensive approach allows for a more thorough analysis of the adoption process (Gulati *et al.*, 2024).

Various academics have employed the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) model to comprehend users' inclinations toward accepting AI-based products or technologies. Recent research has demonstrated that these aspects have a significant impact on students' attitudes towards GenAI tools (Wang and Zhang, 2023). For instance, several studies (such as Budhathoki *et al.*, 2024; Sobaih *et al.*, 2024; Wang & Zhang, 2023) have emphasized the significance of performance expectancy. This refers to the degree to which students believe using GenAI tools can improve their academic performance. In addition, the ease of use of these technologies, known as effort expectancy, is a significant factor in determining students' willingness to adopt GenAI tools. Social influence, habit, hedonistic motivation, and facilitating conditions are other factors that have a bearing on students' perceptions of these technologies (Nikolopoulou, Gialamas and Lavidas, 2021; Alhur *et al.*, 2022).

The UTAUT2 model is undoubtedly a valuable framework for understanding technology adoption (Faqih and Jaradat, 2021). Still, it can be challenging to apply in practice, especially in educational settings where resources are often limited (Malatji, VanEck and Zuva, 2023). Additionally, the model overlooks the role of technology characteristics, particularly GenAI, such as information accuracy, in technology adoption. Despite these limitations, modifying the model can make it more useful for education. Researchers have extended the UTAUT2 model, and these modifications show promise for understanding and implementing technology adoption in education (Tamilmani *et al.*, 2021). A recent study conducted by Wang & Zhang (2023) aimed to understand the factors that influence Generation Z's (GenZers) willingness to adopt GenAI technology. To achieve this, the study combined the UTAUT2, Technology Readiness Index (TRI) model, and trait curiosity. The study found that hedonic motivation and effort expectancy are positively correlated to using GenAI. However, no significant correlation was found between performance expectations and the willingness to use GenAI technology.

Despite the presence of these studies, there is a lack of research that thoroughly investigates how GenAI tools' positive and negative aspects can effectively predict the core elements of the UTAUT2 model, such as behavioral intention and use behavior. As a result, this article seeks to address this gap in research by extending the UTAUT 2 model to include GenAI.

One potential modification that could be made to the UTAUT2 model is to simplify the constructs and incorporate GenAI-related constructs. This way, the model can capture the essential factors influencing students' acceptance of GenAI tools, making them more accessible and practical for real-world use (Chan, 2023; Chan and Lee, 2023). By streamlining the model, educators and developers can gain valuable insights into students' perceptions of GenAI tools, which can enhance their design and implementation in educational settings (Budhathoki *et al.*, 2024; Chiu, 2024).

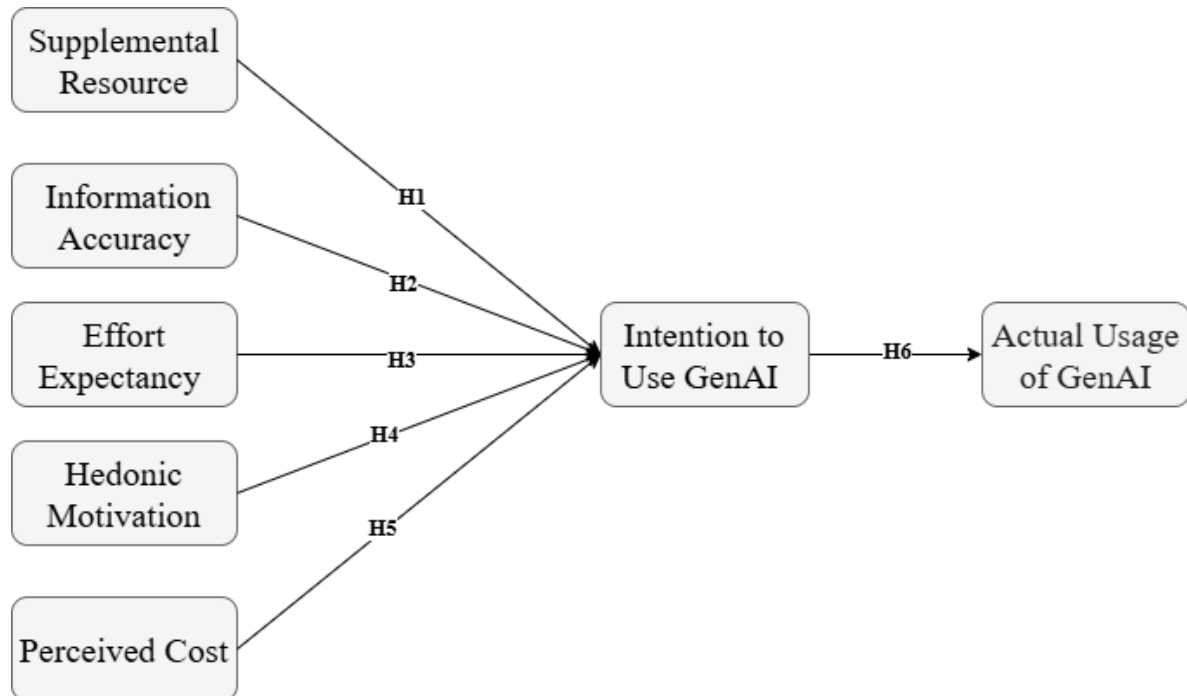
Educators and developers should consider the factors influencing technology adoption when creating and implementing GenAI tools. Integrating GenAI-related concepts, such as information accuracy, and viewing GenAI as a supplemental resource (Michel-Villarreal *et al.*, 2023; AlDreabi *et al.*, 2024) within the UTAUT2 model can help address gaps and improve understanding of technology adoption in educational settings.

## **7. Theoretical Model and Hypotheses**

This study expands on previous research by combining UTAUT2 with GenAI characteristics to create a more comprehensive model for understanding technology adoption. This approach provides deeper insight into the factors influencing higher education students' willingness to use GenAI for learning. The upcoming sections will

explore these research factors and evaluate their implications for educational practice. Additionally, we will identify potential areas for future research in this rapidly evolving field.

Additionally, Figure 1 illustrates the proposed hypotheses of this research and displays six interrelated pathways. Each pathway represents a specific hypothesis, and the model summarizes each component. Overall, the diagram functions as a visual representation of the hypotheses being examined in the study.



**Figure 1: Study Model**

## 8. Supplemental Resource

GenAI serves as a supplemental resource for students by utilizing algorithms to produce customized educational resources such as textbooks, eBooks, quizzes, and other creative materials (Alier, García-Peñalvo and Camba, 2024). This technology adapts content to individual learning preferences, enhancing the learning experience (Borah, T N and Gupta, 2024).

There are numerous promising opportunities for students, educators, and researchers in higher education with the use of GenAI (Chiu, 2024). With the aid of GenAI, students can improve their learning and foster critical thinking skills by receiving personalized feedback, explanations, and recommendations (Michel-Villarreal *et al.*, 2023). Research has shown that GenAI can enhance essay-writing skills and serve as a valuable tutoring tool, encouraging lively student debates and discussions (Dwivedi *et al.*, 2023). Furthermore, when used with traditional course materials, GenAI can help reinforce learning and promote independent research (Mai, Da and Hanh, 2024).

GenAI can greatly help medical teaching, particularly in resource-limited settings. GenAI tools enable students to ask queries about medical ideas and receive customized replies to aid them organize their understanding more effectively (Leng, 2024).

Additionally, GenAI tools can aid research by training students in data organization and location for papers and studies. These same tools can also provide direct feedback on diction and grammar to pupils learning a new language, facilitating their language development (Javaid *et al.*, 2023). Studies by Baidoo-Anu & Ansah (2023) Koraishi (2023), Michel-Villarreal *et al.* (2023) all concur that GenAI is a supplemental resource for higher education students.

Thus, H1: *The perception of GenAI tools as supplemental resource (such as answering queries, generating thoughts, and conducting analyses) has a positive linear impact on higher education students' behavioral intention to use these tools.*

## 9. Information Accuracy

The construct of information accuracy pertains to how students view the dependability and correctness of information given by AI tools (Dahri *et al.*, 2024). Students' readiness to utilize these tools is affected by their trust in the accuracy of the information. A recent study by Dahri *et al.* (2024) emphasizes the significance of the information accuracy concept and how it influences the usage of AI tools. In a different examination, Mizumoto & Eguchi (2023) assessed ChatGPT as an automated tool for scoring essays and discovered that it decreased grading time while ensuring consistency in scoring.

Furthermore, it furnished prompt feedback on the writing skills of students. The effectiveness and reliability of ChatGPT showcase the potential of GenAI to transform the process of teaching, leading to better academic results for college and university students. Nevertheless, it is crucial to remember that the accuracy of AI tools is not always guaranteed, and therefore, they must be used cautiously (Chan and Hu, 2023).

In a study by Ding *et al.* (2023), ChatGPT was used as a virtual tutor to assist in teaching undergraduate-level introductory physics. While it provided an 85% accuracy in answering questions, it occasionally changed its answers from correct to incorrect and vice versa. Students needed clarification about ChatGPT, and almost half trusted its answers regardless of their accuracy.

Thus, H2: *Information accuracy has a positive linear impact on higher education students' behavioral intentions to use GenAI tools.*

## 10. Effort Expectancy

Effort expectancy is an essential factor in deciding whether someone will use a technology. It means how easy or difficult someone thinks it will be to use a technology. If students think it will be easy to use, they will likely use it (Venkatesh *et al.*, 2003). According to UTAUT2, if technology is easy to use, people will think it requires less effort. Some recent studies have found that people are more likely to use AI services if they think they are easy to use (Wang and Zhang, 2023).

Previous inquiries have yielded helpful insights into utilizing GenAI tools in different scenarios, such as education and research, using various theoretical approaches (Ivanov *et al.*, 2024). Specifically, in education, this factor refers to the level of simplicity exhibited by technology that is perceived by students. In case students consider a system or technology to be user-friendly, they are more likely to recognize its benefits and demonstrate deliberate behavior. As a result, this influences their intention to adopt a specific technology (Budhathoki *et al.*, 2024).

Therefore, it is crucial to consider the perception of effort expectancy when introducing new technologies like GenAI in the educational setting. Students who perceive GenAI tools as simple and easy to use are likelier to engage in deliberate behavior and develop an awareness of the benefits. This, in turn, increases their willingness to adopt the technology and utilize it to its full potential.

Consistent with the research mentioned above, we suggest that: H3: *effort expectancy has a positive linear impact on higher education students' behavioral intentions to use GenAI tools.*

## 11. Hedonic Motivation

The Unified Theory of Acceptance and Use of Technology (UTAUT) model has been valuable in understanding technology adoption and use. However, it has been criticized for not accounting for the pleasure and enjoyment that comes with using technology (Budhathoki *et al.*, 2024). To address this, the Unified Theory of Acceptance and Use of Technology (UTAUT2) model was introduced in 2012, which includes hedonic motivation as a factor. As defined by Venkatesh, Thong and Xu, (2012), *hedonic motivation* pertains to the satisfaction and enjoyment individuals experience when using cutting-edge technological systems (Venkatesh, Thong and Xu, 2012). Recent studies have revealed a favorable correlation between hedonic motivation and users' inclination to embrace artificial intelligence assistants. Research has also revealed that hedonic motivation has a positive impact on the inclination to embrace and utilize mobile technology, especially for students who value enjoyable and satisfying user experiences (Al-Azawei and Alowayr, 2020).

Moreover, research has found that teachers' intention to adopt mobile Internet for course instruction is positively influenced by the joy they derive from using it (Nikolopoulou, Gialamas and Lavidas, 2021). Similarly, hedonic motivation has been found to impact the acceptance of mobile technology among secondary school

teachers and students, with perceived enjoyment significantly affecting students' intentions to accept mobile learning (Açıkgül and Şad, 2021).

Our study proposes that hedonic motivation is essential in how higher education students utilize GenAI tools. The interactive and enjoyable environment created by the conversational aspect of GenAI tools enhances the learning experience and stimulates students, ultimately enhancing their knowledge acquisition.

We propose that H4: *hedonic motivation has a positive linear impact on higher education students' behavioral intentions to use GenAI tools.*

## 12. Perceived Cost

Per the UTAUT2 model, perceived cost/price is the rational assessment of the anticipated benefits of utilizing technology for the required financial investment (Wang and Zhang, 2023). Lower costs associated with learning or adopting a new technology result in greater perceived benefits, leading to a stronger intention to use it (Al-Adwan and Al-Debei, 2024).

It is a fundamental principle in technology adoption that the perceived benefits of a technology must outweigh its associated costs, as outlined by Venkatesh, Thong and Xu (2012). Therefore, the financial investment necessary to learn or acquire new technology is a crucial factor, as highlighted by Cecilia Ka Yuk Chan and Zhou (2023). Higher perceived benefits of new technology are associated with lower learning or acquisition costs, ultimately increasing the likelihood of technology use.

In other words, investing in technology can pay off in the long run, especially if we take the time to find affordable options (Wang and Zhang, 2023). An individual's motivation and intention to use a service are significantly influenced by its cost. Students might be less likely to use GenAI if the costs are greater than the advantages for them. Research have shown that students' willingness to use educational technology can be negatively impacted by perceived barriers, such as cost (Chan and Zhou, 2023a).

Thus, H5: *The perceived cost has a negative linear impact on higher education students' behavioral intentions to use GenAI tools.*

## 13. Behavioral Intention

Over the years, researchers in information systems have delved into studying individual behavior and intentions as they relate to technology. This has resulted in the development of various acceptance models for information technology, including the Technology Acceptance Model (TAM), the Unified Theory of Acceptance and Use of Technology (UTAUT), and the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) (Pan and Gao, 2021). The UTAUT2 framework posits that intention is a significant predictor of behavior, influenced by seven fundamental constructs. This theory emphasizes the power of intentions in shaping actions, indicating that individuals are more inclined to act when they genuinely believe their efforts will yield favorable outcomes (Silverman *et al.*, 2016).

In examining the success of information systems, researchers look at actual system usage. The user's willingness to utilize the system can then be understood as their intention to use it. According to experts in technology acceptance, behavioral intention to use directly translates to actual system usage. Most studies aimed at validating technology acceptance models have found this relationship to hold true (Mardiana, Tjakraatmadja and Aprianingsih, 2015).

In this study, "behavioral intentions" refers to students' willingness and determination to integrate GenAI tools into their learning practices. A positive attitude toward these tools indicates students' enthusiasm for incorporating AI technology into their educational endeavors. Previous research has demonstrated that a favorable disposition toward technology usage strongly correlates with its adoption (Ivanov *et al.*, 2024). Similarly, Chatterjee and Bhattacharjee (2020) examined students' behavioral intentions regarding using AI agents or chatbots. Their findings revealed that positive intentions were positively correlated with increased usage of such tools. Consequently, we hypothesize that H6, *the intention to use GenAI has a positive linear impact on higher education students' use of GenAI tools.*

## 14. Methodology

The study utilized a quantitative cross-sectional approach to explore the utilization of GenAI tools by higher education students in Jordanian public universities during the academic year 2023/2024. It specifically targeted students from three prominent governmental universities: The University of Jordan (1631 students), Jordan

University of Science and Technology (JUST) (1214 students), and Al-Balqa Applied University (350 students). Follow random sampling, a total of 374 students participated in a survey conducted via Google Forms between December 10, 2023, and February 5, 2024, with the support of the student affairs deanships of the universities.

The framework of this study was evaluated using Structural Equation Modeling (SEM), a powerful technique designed for analyzing intricate models with multiple variables and their interconnections (Hair *et al.*, 2021). SEM allows for the simultaneous exploration of both direct and indirect relationships among constructs, deepening our insight into how various factors within the study's model influence the adoption of Generative AI tools (Masud *et al.*, 2024). This methodological approach perfectly aligns with the study's aim to investigate behavioral intention and adoption behavior among students, as SEM effectively integrates measurement and structural components, enhancing the reliability and robustness of the findings.

The questionnaire, initially developed in English and later translated into Arabic, consisted of 23 items that assessed various aspects of the research model, including four demographic questions, incorporating demographic variables such as gender and frequency of GenAI tool usage, essential for interpreting the study's findings. Gender influences technology adoption behaviors, and understanding how often students use GenAI, including whether they opt for free or paid versions, provides insights into access and familiarity in adoption intentions (Table (1)). We adapted a previously validated questionnaire from earlier studies, as outlined in Appendix 1. All the scales used in our study have been validated and shown reliability in studies by Chan and Lee (2023), Venkatesh, Thong and Xu (2012), and Dahri *et al.* (2024).

The survey investigated the elements of the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2), initially designed to analyze technology acceptance in consumer contexts. This framework has been refined to integrate characteristics specific to Generative AI (GenAI), highlighting key aspects such as hedonic motivation, effort expectancy, and behavioral intention. Additionally, it considers crucial factors like information accuracy, perceived costs, and the perception of GenAI as a valuable supplemental resource. This enhanced UTAUT2 framework is particularly suited for higher education environments, where varying motivations and perceived utility influence individual attitudes toward technology. Consequently, it provides a robust theoretical foundation for comprehending the adoption of GenAI tools in educational settings.

Prior to analysis, the data underwent thorough scrutiny for any missing information or anomalies. The sample size was deemed suitable for SEM methodology (Kyriazos, 2018), and an exploratory factor analysis was conducted to effectively consolidate the dimensions linked with each construct.

## 15. Sample Characteristics

In Table 1, an analysis of the demographic data from the sample is presented, including gender, age, frequency of GenAI tools usage, and whether participants used the paid version of GenAI tools. The sample included 374 students, 52.7% female and 47.3% male. The majority of participants 43.3% fell within the age range of 18-30 years, followed by 31-41 years 31.8%, and ≥42 years 24.9%. In terms of GenAI tools usage, 12% reported using it only once, 23% used it 2-3 times, and 65% used it three times or more. As for whether participants used the paid version of GenAI tools, 56.7% reported using the free version, while 43.3% used the paid version.

**Table 1: Sample characteristics**

| Demographic Data                  | Categories         | Count | Percentage % |
|-----------------------------------|--------------------|-------|--------------|
| Gender                            | Female             | 197   | 52.7%        |
|                                   | Male               | 177   | 47.3%        |
|                                   | Total              | 374   | 100%         |
| Age                               | 18-30 years        | 162   | 43.3%        |
|                                   | 31-41 years        | 119   | 31.8%        |
|                                   | 42 yrs. and over   | 93    | 24.9%        |
|                                   | Total              | 374   | 100%         |
| How often do you use GenAI tools? | Once times or less | 45    | 12.0%        |
|                                   | 2-3 times in week  | 86    | 23.0%        |
|                                   | More than 3 times  | 243   | 65.0%        |
|                                   | Total              | 374   | 100%         |

| Demographic Data                            | Categories | Count | Percentage % |
|---------------------------------------------|------------|-------|--------------|
| Do you use the paid version of GenAI tools? | No         | 212   | 56.7%        |
|                                             | Yes        | 162   | 43.3%        |
|                                             | Total      | 374   | 100%         |

## 16. Results

The study's data underwent rigorous analysis using IBM SPSS 27 and IBM AMOS 28. As per Hair et al.'s (2019) two-step approach, the researchers conducted confirmatory factor analysis (CFA) to evaluate the measurement model's reliability, validity, and fitness indices. In the second step, they employed robust structural equation modeling (SEM) to examine all hypotheses and comprehensively understand the study's findings.

## 17. The Study's Reliability and Validity

This study presents a novel measurement model for assessing students' adoption of GenAI tools in higher education. Drawing on the Unified Theory of Acceptance and Use of Technology (UTAUT) model and GenAI literature, the model was formulated based on the Hair 2019 guidelines. Measurement theory was employed to determine how the latent variables (constructs) are measured, wherein the reflective measurement approach was used due to its suitability for the current context. This approach can effectively capture the nature and nuances of the constructs and provide more reliable and accurate results (Hair et al., 2021).

The model used in this study was rigorously fitted with data, yielding strong fit indices. Specifically, the findings revealed a chi-square value of  $\chi^2 (180) = 565.552$ , a chi-square to degrees of freedom ratio of  $\chi^2/df = 3.142$ , a Comparative Fit Index (CFI) of 0.924, a Standardized Root Mean Square Residual (SRMR) of 0.051, and a Root Mean Square Error of Approximation (RMSEA) of 0.076, with a P value exceeding 0.05 (Crawford and Kelder, 2019). To further establish the validity and reliability of the instruments employed, Tables 2 and 3 present findings demonstrating a Cronbach's alpha value greater than 0.70, alongside factor loadings that surpass the recommended threshold of 0.50 (Hair et al., 2021). Additionally, the Average Variance Extracted (AVE) was greater than 0.50 (Kline, 2011), as detailed in Table 3.

**Table 2: CFA and descriptive statistics**

| Items | Factor Loadings* | $\alpha^*$ | M(SD)*     | Skewness* | Kurtosis* |
|-------|------------------|------------|------------|-----------|-----------|
| SR1   | .939             | 0.937      | 3.70(.967) | -.179     | -.631     |
| SR3   | .917             |            |            |           |           |
| SR2   | .898             |            |            |           |           |
| IA1   | .935             | 0.925      | 3.03(.947) | -.077     | -.559     |
| IA3   | .840             |            |            |           |           |
| IA4   | .863             |            |            |           |           |
| IA2   | .853             |            |            |           |           |
| HM3   | .923             | 0.859      | 3.31(.917) | -.178     | -.698     |
| HM2   | .864             |            |            |           |           |
| HM1   | .692             |            |            |           |           |
| Int.3 | .911             | 0.851      | 4.01(.946) | -.093     | -.666     |
| Int.2 | .835             |            |            |           |           |
| Int.1 | .693             |            |            |           |           |
| PC2   | .752             | 0.795      | 3.38(.879) | .147      | .516      |
| PC1   | .786             |            |            |           |           |
| PC3   | .738             |            |            |           |           |
| AU2   | .832             | 0.797      | 3.86(.903) | -.134     | -.481     |
| AU1   | .715             |            |            |           |           |
| AU3   | .721             |            |            |           |           |
| EE3   | .745             | 0.745      | 3.94(.917) | -.100     | -.450     |

| Items | Factor Loadings* | $\alpha^*$ | M(SD)* | Skewness* | Kurtosis* |
|-------|------------------|------------|--------|-----------|-----------|
| EE2   | .745             |            |        |           |           |
| EE1   | .625             |            |        |           |           |

Note: SR: Supplemental resource, IA: Information accuracy, HM: Hedonic motivation, Int.: Intention to use GenAI tools, PC: Perceived cost, AU: Actual usage of GenAI tools, EE: Effort expectancy.  $\alpha$ = Cronbach's Alpha coefficient; M(SD)= Mean & Standard deviation. \* These values fall within the thresholds established by Kline (2011) and Hair *et al.* (2019, 2021a)

Examining both convergent and discriminant validity indicated that the research instrument exhibited adequate convergent validity (Hair *et al.*, 2021). Furthermore, Tables 3 and 4 confirm that the measurement items possess sufficient discriminant validity, with Composite Reliability (CR) values going beyond the AVE values (Kline, 2011). The AVE values also exceeded the Average Shared Variance (ASV) and Maximum Shared Variance (MSV) values. At the same time, the correlations among the independent variables remained below the threshold of 0.70 (Almén *et al.*, 2018).

**Table 3: Study model's validity**

| Factors | CR    | AVE   | MSV   | MaxR(H) | SR           | IA           | HM           | Int.         | PC           | AU           | EE           |
|---------|-------|-------|-------|---------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| SR      | 0.941 | 0.843 | 0.158 | 0.944   | <b>0.918</b> |              |              |              |              |              |              |
| IA      | 0.928 | 0.763 | 0.195 | 0.937   | 0.397        | <b>0.874</b> |              |              |              |              |              |
| HM      | 0.869 | 0.692 | 0.165 | 0.906   | 0.186        | 0.348        | <b>0.832</b> |              |              |              |              |
| Int.    | 0.857 | 0.669 | 0.114 | 0.890   | 0.154        | 0.160        | 0.175        | <b>0.818</b> |              |              |              |
| PC      | 0.803 | 0.576 | 0.196 | 0.805   | 0.133        | -0.442       | 0.329        | -0.300       | <b>0.759</b> |              |              |
| AU      | 0.801 | 0.574 | 0.114 | 0.814   | 0.160        | 0.250        | 0.159        | 0.338        | 0.247        | <b>0.758</b> |              |
| EE      | 0.749 | 0.501 | 0.196 | 0.758   | 0.170        | 0.370        | 0.406        | 0.201        | 0.443        | 0.204        | <b>0.707</b> |

Note: SR: Supplemental resource, IA: Information accuracy, HM: Hedonic motivation, Int.: Intention to use GenAI tools, PC: Perceived cost, AU: Actual usage of GenAI tools, EE: Effort expectancy; Composite Reliability = (CR) > 0.70, Average Variance Extracted = AVE > 0.50, Maximum Shared Variance = AVE > MSV and McDonald Construct Reliability = MaxR(H) > 0.7. The square root of the AVE is displayed as diagonal boldface values. These values fall within the thresholds established by Kline (2011), Hair *et al.* (2019) and Almén *et al.* (2018)

**Table 4: HTMT Analysis**

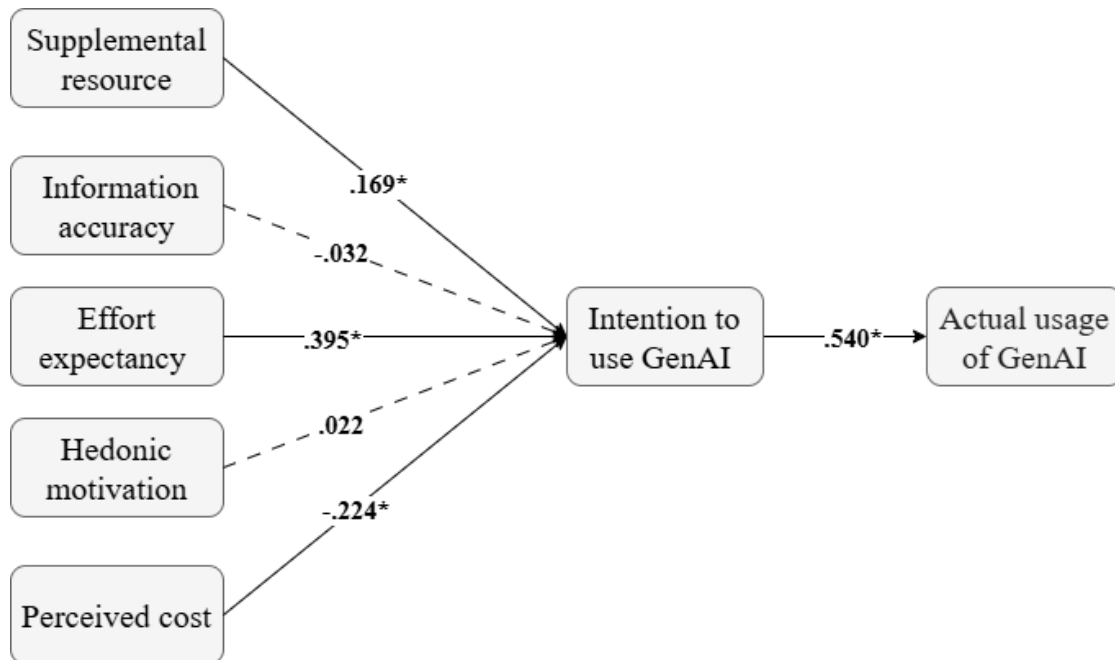
| Factors | SR    | IA    | HM    | Int.  | PC    | AU    | EE |
|---------|-------|-------|-------|-------|-------|-------|----|
| SR      |       |       |       |       |       |       |    |
| IA      | 0.404 |       |       |       |       |       |    |
| HM      | 0.253 | 0.362 |       |       |       |       |    |
| Int.    | 0.167 | 0.178 | 0.202 |       |       |       |    |
| PC      | 0.138 | 0.471 | 0.364 | 0.313 |       |       |    |
| AU      | 0.159 | 0.267 | 0.188 | 0.358 | 0.273 |       |    |
| EE      | 0.178 | 0.354 | 0.455 | 0.212 | 0.468 | 0.210 |    |

Note: SR: Supplemental resource, IA: Information accuracy, HM: Hedonic motivation, Int.: Intention to use GenAI tools, PC: Perceived cost, AU: Actual usage of GenAI tools, EE: Effort expectancy. These values fall within the thresholds established by Almén *et al.* (2018)

## 18. Structural Model

The study utilized Structural Equation Modeling (SEM) to investigate the factors that affect students' intention to use Generative AI (GenAI) tools. The structural model was developed in accordance with the guidelines established by Hair *et al.* (2021). The findings indicated that the model displayed a satisfactory fit, as assessed against the criteria set forth by Crawford & Kelder (2019):  $\chi^2$  (185) = 597.590,  $\chi^2/df$  = 3.230, CFI = 0.939, SRMR = 0.052, and RMSEA = 0.073, with a p-value exceeding 0.05.

Figure 2 illustrates the finalized structural model, depicting the relationships among several key predictors—including supplemental resource, information accuracy, effort expectancy, hedonic motivation, and perceived cost—along with the intention to use GenAI tools and the actual usage of these tools. Each pathway is annotated with its standardized regression weight ( $\beta$ ) and statistical significance level. Significant relationships are marked with solid arrows ( $*p < 0.05$ ), while dashed arrows indicate non-significant relationships.



**Figure 2: Structural Model, \*  $P < 0.05$ .**

The results, as presented in Table 5, show that supplemental resource ( $\beta = 0.169$ ,  $p < 0.01$ ), effort expectancy ( $\beta = 0.395$ ,  $p < 0.001$ ), and perceived cost ( $\beta = -0.224$ ,  $p < 0.05$ ) significantly influence students' intention to use GenAI tools, thereby supporting hypotheses H1, H4, and H5. Additionally, the intention to use GenAI tools is a strong predictor of actual usage ( $\beta = 0.54$ ,  $p < 0.001$ ), supporting hypothesis H6.

In contrast, the factors of information accuracy ( $\beta = -0.032$ ,  $p > 0.05$ ) and hedonic motivation ( $\beta = 0.022$ ,  $p > 0.05$ ) do not have a significant impact on students' intention, failing to support hypotheses H2 and H3. Overall, the model accounts for 39% of the variance in students' intention to use GenAI tools and 29% of the variance in their actual usage.

**Table 5: Hypotheses testing**

| Hypothesis | Predictors                   | Outcomes                     | S.E.* | t-value | Beta    |
|------------|------------------------------|------------------------------|-------|---------|---------|
| H1         | Supplemental resource        | Intention to use GenAI tools | .062  | 2.867   | .169**  |
| H2         | Information accuracy         | Intention to use GenAI tools | .075  | 0.417   | -.032   |
| H3         | Hedonic motivation           | Intention to use GenAI tools | .062  | 0.340   | .022    |
| H4         | Perceived cost               | Intention to use GenAI tools | .114  | 2.572   | .224*   |
| H5         | Effort expectancy            | Intention to use GenAI tools | .093  | 4.651   | .395*** |
| H6         | Intention to use GenAI tools | Actual usage of GenAI tools  | .050  | 9.863   | .540*** |

Note: S.E. = Standard Error, \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$

## 19. Discussion

This research utilized an adapted version of the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) model to examine the factors influencing students' adoption of generative AI (GenAI) tools. The analysis highlighted several key factors: perceived cost, effort expectancy, hedonic motivation, supplemental resource, and information accuracy.

The findings in Table 5 indicate that students are significantly more inclined to use GenAI tools when they view them as offering valuable supplemental resource, being cost-effective, and being easy to use. Furthermore, Figure 2 illustrates the relationships among these factors and their influence on students' behavioral intentions to adopt GenAI tools in higher education.

The results reveal that effort expectancy and behavioral intention significantly influence students' adoption of Generative AI (GenAI). Conversely, the hedonistic value has little effect on students' willingness to embrace Generative AI. These findings align with previous research conducted by Ivanov *et al.* (2024), McDonald *et al.* (2024), which also emphasized the significance of usability and the availability of supportive resources in technology adoption.

Moreover, the results show that intention behavior has a robust and significant effect on willingness to embrace Generative AI. This finding aligns with previous research (Venkatesh, 2022; Li, 2024; Lu *et al.*, 2024), reinforcing the notion that intention is a critical determinant in technology adoption.

The recent shift in focus highlights the importance of educational institutions prioritizing the creation of user-friendly tools that seamlessly fit into students' academic workflows. These institutions must invest in training and support resources that improve students' experiences, enabling them to utilize these tools effectively and navigate their academic tasks with minimal challenges.

The research also highlights vital factors influencing the adoption of Generative AI (GenAI) tools, mainly focusing on information accuracy, supplemental resource, and perceived cost. Findings indicate that students are more likely to use GenAI tools when they view them as cost-effective and offering valuable resources. This aligns with studies by Michel-Villarre *et al.* (2023) and Wang and Zhang (2023) on the importance of considering GenAI as a supplemental resource for technology adoption in education.

While information accuracy is relevant, students prioritize perceived cost and availability of supportive resources (Chan and Zhou, 2023a). This change in priorities suggests that initiatives to promote GenAI adoption in higher education should place less emphasis on refining precision or enhancing the hedonic value of these tools and more on ensuring they are accessible, user-friendly, and accompanied by robust resources (Hmoud *et al.*, 2023; Li, 2024).

Among the various factors influencing the adoption of Generative AI (GenAI) tools by higher education students, effort expectancy and the availability of supplemental resource have emerged as the most reliable predictors. The findings suggest that students are more inclined to adopt GenAI tools when they view them as user-friendly and recognize the presence of supporting resources, such as tailored examples and information that cater to their learning needs. This observation aligns with existing literature on technology adoption in educational settings, underscoring the importance of accessibility and support structures for effective technology integration (Ivanov *et al.*, 2024; Meakin, 2024).

The focus on effort expectancy indicates that institutions should prioritize making GenAI tools intuitive and easy to use. Students who find these tools straightforward to navigate are more likely to engage with them consistently. Furthermore, the availability of supplemental resource is vital in promoting sustained use. Students benefit significantly from resources that enhance their understanding and application of GenAI tools, which can influence their learning outcomes (Granić, 2022).

## **20. Study Contributions About Extending the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) Model**

The UTAUT2 framework is expanded in this study to incorporate new factors specific to GenAI technologies in higher education. These additional factors, such as supplemental resource and effort expectancy, are vital for understanding students' intentions to adopt GenAI tools, even though they are not explicitly covered in the original UTAUT2 model.

The study utilizes the UTAUT2 model to analyze the adoption of GenAI tools in higher education, shedding light on how existing UTAUT2 variables, like effort expectancy, interact with new GenAI-specific variables, such as supplemental resource.

Furthermore, the research confirms the relevance of existing UTAUT2 variables, like perceived cost and effort expectancy, in the context of GenAI tools, signifying their continued significance in technology adoption.

Moreover, the research outlines practical implications for educational institutions, underscoring the importance of providing comprehensive resources and ensuring the affordability and usability of GenAI tools to facilitate their integration into higher education environments.

In conclusion, this study enhances the UTAUT2 model by introducing new variables, validating existing ones, critiquing less influential factors, and providing practical insights for implementation in higher education settings.

## 21. Study Implications

This study aims to explore the factors influencing the utilization of GenAI tools in education, building upon previous studies in this field. Prior research mainly concentrated on identifying the essential variables impacting the use of these tools in various situations. However, this study takes a more comprehensive approach by considering multiple essential elements, such as accuracy of information, supplemental resource, and perceived cost. It seeks to establish a structural model that assesses the most significant factors affecting the adoption of GenAI tools among undergraduate and postgraduate students.

The study also contributes to existing knowledge by extending the UTAUT2 framework to demonstrate the main factors promoting the implementation of GenAI technologies, like Genmini, in higher education institutions. This expansion is important as it highlights the necessity for educational institutions to take various factors into account when making decisions about integrating GenAI tools. By doing so, educational institutions can ensure that the tools are effectively used, enabling their students to benefit from utilizing them in their studies and education.

In contrast, the study's findings present two key practical implications. Firstly, to ensure the effective use of GenAI tools, educational institutions, and developers should prioritize making them more cost-effective, user-friendly, and resourceful. This can be achieved by offering additional resources and support to students, improving the usability of the tools, and ensuring that they are perceived as valuable supplements to traditional learning methods. By addressing these factors, institutions can increase students' willingness to use GenAI tools, promoting their adoption and integration into educational environments.

Secondly, the study highlights the importance of a comprehensive approach to technology adoption in educational settings. While accuracy of information and hedonic motivation are important, they may not be the sole drivers of students' willingness to use GenAI tools. Therefore, institutions should focus on understanding their student's specific needs and preferences, taking into account factors such as perceived ease of use and cost in addition to accuracy and hedonism. By considering the broader context of technology adoption, institutions can develop more effective strategies to encourage the use of GenAI tools among students, ultimately facilitating their integration into the learning process.

## 22. Limitations

It is important to note that this study has limitations due to its specific context (higher education), potential biases in self-reported data, a narrow focus on factors influencing students' intent to use GenAI tools, and a cross-sectional design, which limits establishing causal or temporal relationships between variables.

## 23. Conclusion and Further Research

This study investigates the factors that influence the adoption of Generative AI (GenAI) tools among higher education students by utilizing a modified version of the UTAUT2 model. It examines perceived cost, effort expectancy, hedonic motivation, supplemental resource, and information accuracy. The findings indicate that ease of use and the availability of support resources are critical drivers for students' adoption of GenAI tools. This suggests that the adoption of educational technology is primarily influenced by practical utility and accessibility rather than simply by enjoyment or high accuracy.

The research identifies that supplemental resource and effort expectancy are the strongest predictors of students' intentions to use GenAI tools. This highlights that students tend to favor tools that are user-friendly and come with resources that enhance the learning experience. Conversely, information accuracy and hedonic motivation play a lesser role in adoption, indicating a shift in students' perceptions of what is essential in educational technology. These findings enrich the UTAUT2 model by incorporating context-sensitive variables and provide practical insights for educational institutions looking to implement GenAI tools effectively. By addressing these key factors, institutions can foster environments encouraging GenAI adoption and better supporting the learning process.

For future research, it is recommended that studies broaden their focus to explore additional variables and employ a range of methodologies. Incorporating objective measures and observational data could help alleviate the limitations often associated with self-reported findings. Furthermore, conducting longitudinal or experimental studies could yield deeper insights into how students' perceptions and intentions regarding GenAI tools evolve over time. These approaches would foster a more inclusive and comprehensive understanding of effective integrations of GenAI tools, inspiring the audience with the potential for further exploration and ultimately contributing to a more supportive educational environment for various student populations.

**Ethics Statement:** Participants in the study were thoroughly informed about its nature, purpose, and potential outcomes. Data collection was conducted anonymously, ensuring that no personally identifiable information was gathered or retained. (Ethics approval is not required for our research, and we did not use an AI tool)

**AI statement:** While preparing this work, the authors utilized Grammarly to enhance readability and language. After using this tool, they carefully reviewed and edited the content as necessary and took full responsibility for the final published article.

## References

- Açıkgül, K. and Şad, S.N. (2021) 'High school students' acceptance and use of mobile technology in learning mathematics', *Education and Information Technologies*, 26(4), pp. 4181–4201. <https://doi.org/10.1007/s10639-021-10466-7>.
- Akyuz, Y. (2020) 'Effects of Intelligent Tutoring Systems (ITS) on Personalized Learning (PL)', *Creative Education*, 11(06), pp. 953–978. <https://doi.org/10.4236/ce.2020.116069>.
- Al-Adwan, A.S. and Al-Debei, M.M. (2024) 'The determinants of Gen Z's metaverse adoption decisions in higher education: Integrating UTAUT2 with personal innovativeness in IT', *Education and Information Technologies*, 29(6), pp. 7413–7445. <https://doi.org/10.1007/s10639-023-12080-1>.
- Al-Azawei, A. and 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, pp. 101325. <https://doi.org/10.1016/j.techsoc.2020.101325>.
- Alam, A. (2021) 'Possibilities and Apprehensions in the Landscape of Artificial Intelligence in Education', in *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)*. IEEE, pp. 1–8. <https://doi.org/10.1109/ICCICA52458.2021.9697272>.
- AlDreabi, H., Twahya, F.K.A. Al, Alzboun, N., Anabtawi, M., Ghaboush, R.A., Alhur, M. and Alshurideh, M.T. (2024) 'The role of digital communication in developing administrative work in higher education institutions', *International Journal of Data and Network Science*, 8(2), pp. 1261–1274. <https://doi.org/10.5267/j.ijdns.2023.11.008>.
- Alhur, M.S., Alshamari, S., Oláh, J. and Aldreabi, H. (2022) 'Unsupervised Machine Learning to Identify Positive and Negative Themes in Jordanian mHealth Apps', *International Journal of E-Services and Mobile Applications*, 14(1), pp. 1–21. <https://doi.org/10.4018/IJESMA.313950>.
- Alier, M., García-Peñalvo, F.-J. and Camba, J.D. (2024) 'Generative Artificial Intelligence in Education: From Deceptive to Disruptive', *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5), pp. 5. <https://doi.org/10.9781/ijimai.2024.02.011>.
- Almén, N., Lundberg, H., Sundin, Ö. and Jansson, B. (2018) 'The reliability and factorial validity of the Swedish version of the Recovery Experience Questionnaire', *Nordic Psychology*, 70(4), pp. 324–333. <https://doi.org/10.1080/19012276.2018.1443280>.
- Alzahrani, L. (2023) 'Analyzing Students' Attitudes and Behavior Toward Artificial Intelligence Technologies in Higher Education', *International Journal of Recent Technology and Engineering (IJRTE)*, 11(6), pp. 65–73. <https://doi.org/10.35940/ijrte.F7475.0311623>.
- Arantes, J.A. (2024) 'Redefining Classroom Readiness: How Initial Teacher Training Can Mitigate Risks and Capitalize on the Potential of GenAI', in *Academic Integrity in the Age of Artificial Intelligence*. IGI Global, pp. 78–92. <https://doi.org/10.4018/979-8-3693-0240-8.ch005>.
- Bahrour, Z., Anane, C., Ahmed, V. and Zacca, A. (2023) 'Transforming Education: A Comprehensive Review of Generative Artificial Intelligence in Educational Settings through Bibliometric and Content Analysis', *Sustainability*, 15(17), pp. 12983. <https://doi.org/10.3390/su151712983>.
- Baidoo-Anu, D. and 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), pp. 52–62. <https://doi.org/10.61969/jai.1337500>.
- Borah, A.R., T N, N. and Gupta, S. (2024) 'Improved Learning Based on GenAI', in *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, pp. 1527–1532. <https://doi.org/10.1109/IDCIoT59759.2024.10467943>.
- Budhathoki, T., Zirar, A., Njoya, E.T. and Timsina, A. (2024) 'ChatGPT adoption and anxiety: a cross-country analysis utilising the unified theory of acceptance and use of technology (UTAUT)', *Studies in Higher Education*, pp. 1–16. <https://doi.org/10.1080/03075079.2024.2333937>.
- Chan, C.K.Y. (2023) 'A comprehensive AI policy education framework for university teaching and learning', *International Journal of Educational Technology in Higher Education*, 20(1). pp.20. <https://doi.org/10.1186/s41239-023-00408-3>.

- Chan, C.K.Y. and Colloton, T. (2024) *Generative AI in Higher Education*. London: Routledge. <https://doi.org/10.4324/9781003459026>.
- Chan, C.K.Y. and Hu, W.J. (2023) 'Students' voices on generative AI: perceptions, benefits, and challenges in higher education', *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00411-8>.
- Chan, C.K.Y. and Lee, K.K.W. (2023) 'The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and millennial generation teachers?', *Smart Learning Environments*, 10(1). <https://doi.org/10.1186/s40561-023-00269-3>.
- Chan, C.K.Y. and Zhou, W. (2023a) 'An expectancy value theory (EVT) based instrument for measuring student perceptions of generative AI', *Smart Learning Environments*, 10(1), pp. 64. <https://doi.org/10.1186/s40561-023-00284-4>.
- Chan, C.K.Y. and Zhou, W. (2023b) 'Deconstructing Student Perceptions of Generative AI (GenAI) through an Expectancy Value Theory (EVT)-based Instrument', *Computers and Society* [Preprint]. <https://doi.org/https://doi.org/10.48550/arXiv.2305.01186>.
- Chatterjee, S. and Bhattacharjee, K.K. (2020) 'Adoption of artificial intelligence in higher education: a quantitative analysis using structural equation modelling', *Education and Information Technologies*, 25(5), pp. 3443–3463. <https://doi.org/10.1007/s10639-020-10159-7>.
- Chiu, T.K.F. (2024) 'Future research recommendations for transforming higher education with generative AI', *Computers and Education: Artificial Intelligence*, 6, pp.100197. <https://doi.org/10.1016/j.caeai.2023.100197>.
- Crawford, J.A. and Kelder, J.-A. (2019) 'Do we measure leadership effectively? Articulating and evaluating scale development psychometrics for best practice', *The Leadership Quarterly*, 30(1), pp. 133–144. <https://doi.org/10.1016/j.leaqua.2018.07.001>.
- Dahri, N.A., Yahaya, N., Al-Rahmi, W.M., Vighio, M.S., Alblehai, F., Soomro, R.B. and Shutaleva, A. (2024) 'Investigating AI-based academic support acceptance and its impact on students' performance in Malaysian and Pakistani higher education institutions', *Education and Information Technologies* [Preprint]. <https://doi.org/10.1007/s10639-024-12599-x>.
- Ding, L., Li, T., Jiang, S.Y. and Gapud, A. (2023) 'Students' perceptions of using ChatGPT in a physics class as a virtual tutor', *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00434-1>.
- Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., Baabdullah, A.M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M.A., Al-Busaidi, A.S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., Carter, L., Chowdhury, S., Crick, T., Cunningham, S.W., Davies, G.H., Davison, R.M., Dé, R., Dennehy, D., Duan, Y., Dubey, R., Dwivedi, R., Edwards, J.S., Flavián, C., Gauld, R., Grover, V., Hu, M.-C., Janssen, M., Jones, P., Junglas, I., Khorana, S., Kraus, S., Larsen, K.R., Latreille, P., Laumer, S., Malik, F.T., Mardani, A., Mariani, M., Mithas, S., Mogaji, E., Nord, J.H., O'Connor, S., Okumus, F., Pagani, M., Pandey, N., Papagiannidis, S., Pappas, I.O., Pathak, N., Pries-Heje, J., Raman, R., Rana, N.P., Rehm, S.-V., Ribeiro-Navarrete, S., Richter, A., Rowe, F., Sarker, S., Stahl, B.C., Tiwari, M.K., van der Aalst, W., Venkatesh, V., Viglia, G., Wade, M., Walton, P., Wirtz, J. and Wright, R. (2023) 'Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy', *International Journal of Information Management*, 71, pp. 102642. <https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2023.102642>.
- Faisal Rashid, S., Duong-Trung, N. and Pinkwart, N. (2024) 'Generative AI in Education: Technical Foundations, Applications, and Challenges', in *Artificial Intelligence for Quality Education [Working Title]*. IntechOpen, pp. 1–20. <https://doi.org/10.5772/intechopen.1005402>.
- Faqih, K.M.S. and Jaradat, M.-I.R.M. (2021) 'Integrating TTF and UTAUT2 theories to investigate the adoption of augmented reality technology in education: Perspective from a developing country', *Technology in Society*, 67, pp.101787. <https://doi.org/10.1016/j.techsoc.2021.101787>.
- Fernandes, C.W., Rafatirad, S. and Sayadi, H. (2023) 'Advancing Personalized and Adaptive Learning Experience in Education with Artificial Intelligence', in *2023 32nd Annual Conference of the European Association for Education in Electrical and Information Engineering (EAEIE)*. IEEE, pp. 1–6. <https://doi.org/10.23919/EAEIE55804.2023.10181336>.
- Furze, L., Perkins, M., Roe, J. and MacVaugh, J. (2024) 'The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI supported assessment', *arXiv preprint arXiv:2403.14692* [Preprint]. <http://arxiv.org/abs/2403.14692>.
- Granić, A. (2022) 'Educational Technology Adoption: A systematic review', *Education and Information Technologies*, 27(7), pp. 9725–9744. <https://doi.org/10.1007/s10639-022-10951-7>.
- Gulati, A., Saini, H., Singh, S. and Kumar, V. (2024) 'Enhancing learning potential: investigating marketing students' behavioral intentions to adopt chatgpt', *Marketing Education Review*, pp. 1–34. <https://doi.org/10.1080/10528008.2023.2300139>.
- Gupta, R., Nair, K., Mishra, M., Ibrahim, B. and Bhardwaj, S. (2024) 'Adoption and impacts of generative artificial intelligence: Theoretical underpinnings and research agenda', *International Journal of Information Management Data Insights*, 4(1), pp. 100232. <https://doi.org/10.1016/j.ijime.2024.100232>.
- Gupta, V. and Yang, H. (2024) 'Study protocol for factors influencing the adoption of ChatGPT technology by startups: Perceptions and attitudes of entrepreneurs', *PLOS ONE*. Edited by A. Haldorai, 19(2), pp. e0298427. <https://doi.org/10.1371/journal.pone.0298427>.

- Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2019) *Multivariate Data Analysis*. 8th, illustr edn. Edited by P. Prentice. England: Cengage.  
[https://books.google.es/books/about/Multivariate\\_Data\\_Analysis.html?id=OR9ZswEACAAJ&redir\\_esc=y](https://books.google.es/books/about/Multivariate_Data_Analysis.html?id=OR9ZswEACAAJ&redir_esc=y).
- Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P. and Ray, S. (2021a) 'An Introduction to Structural Equation Modeling', in *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R. Classroom Companion: Business*, Cham, pp. 1–29. [https://doi.org/10.1007/978-3-030-80519-7\\_1](https://doi.org/10.1007/978-3-030-80519-7_1).
- Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P. and Ray, S. (2021b) 'An Introduction to Structural Equation Modeling BT - Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook', in J.F. Hair Jr., G.T.M. Hult, C.M. Ringle, M. Sarstedt, N.P. Danks, and S. Ray (eds). Cham: Springer International Publishing, pp. 1–29. [https://doi.org/10.1007/978-3-030-80519-7\\_1](https://doi.org/10.1007/978-3-030-80519-7_1).
- Hmoud, H., Al-Adwan, A.S., Horani, O., Yaseen, H. and Zoubi, J.Z. Al (2023) 'Factors influencing business intelligence adoption by higher education institutions', *Journal of Open Innovation: Technology, Market, and Complexity*, 9(3), pp. 100111. <https://doi.org/10.1016/j.joitmc.2023.100111>.
- Ivanov, S., Soliman, M., Tuomi, A., Alkathiri, N.A. and Al-Alawi, A.N. (2024) 'Drivers of generative AI adoption in higher education through the lens of the Theory of Planned Behaviour', *Technology in Society*, 77, p. 102521. <https://doi.org/https://doi.org/10.1016/j.techsoc.2024.102521>.
- Javaid, M., Haleem, A., Singh, R.P., Khan, S. and Khan, I.H. (2023) 'Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system', *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 3(2), pp. 100115. <https://doi.org/https://doi.org/10.1016/j.tbench.2023.100115>.
- Kline, R. (2011) *Principles and practice of structural equation modeling*, 3rd edn, Guilford Press, New York.
- Koraishi, O. (2023) 'Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment', *Language Education and Technology*, 3(1), 55–72.
- Kyriazos, T.A. (2018) 'Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General', *Psychology*, 09(08), pp. 2207–2230. <https://doi.org/10.4236/psych.2018.98126>.
- Leng, L. (2024) 'Challenge, integration, and change: ChatGPT and future anatomical education.', *Medical education online*, 29(1), pp. 2304973. <https://doi.org/10.1080/10872981.2024.2304973>.
- Li, W. (2024) 'A Study on Factors Influencing Designers' Behavioral Intention in Using AI-Generated Content for Assisted Design: Perceived Anxiety, Perceived Risk, and UTAUT', *International Journal of Human–Computer Interaction*, pp. 1–14. <https://doi.org/10.1080/10447318.2024.2310354>.
- Lu, H., He, L., Yu, H., Pan, T. and Fu, K. (2024) 'A Study on Teachers' Willingness to Use Generative AI Technology and Its Influencing Factors: Based on an Integrated Model', *Sustainability*, 16(16), pp. 7216. <https://doi.org/10.3390/su16167216>.
- Maghsudi, S., Lan, A., Xu, J. and van der Schaar, M. (2021) 'Personalized Education in the Artificial Intelligence Era: What to Expect Next', *IEEE Signal Processing Magazine*, 38(3), pp. 37–50. <https://doi.org/10.1109/MSP.2021.3055032>.
- Mai, D.T.T., Da, C. Van and Hanh, N. Van (2024) 'The use of ChatGPT in teaching and learning: a systematic review through SWOT analysis approach', *Frontiers in Education*, 9. <https://doi.org/10.3389/educ.2024.1328769>.
- Malatji, W.R., VanEck, R. and Zuva, T. (2023) 'A Review of Theories Utilized in Understanding Online Information Privacy Perceptions', in *Computer Science On-line Conference*. Springer, pp. 54–67. [https://doi.org/10.1007/978-3-031-35311-6\\_7](https://doi.org/10.1007/978-3-031-35311-6_7).
- Mardiana, S., Tjakraatmadja, J. and Aprianingsih, A. (2015) 'Validating the Conceptual Model for Predicting Intention to Use as Part of Information System Success Model: The Case of an Indonesian Government Agency', *Procedia Computer Science*, 72, pp. 353–360. <https://doi.org/10.1016/j.procs.2015.12.150>.
- Masud, A. Al, Ahmed, S., Kaisar, M.T., Hossain, B., Shimu, M. and Islam, M.F. (2024) 'Unveiling brand loyalty in emerging markets: Analyzing smartphone user preferences: Robustness of structural equation modeling (SEM) and simultaneous equation modeling (SEMs)', *Journal of Open Innovation: Technology, Market, and Complexity*, 10(3), pp. 100353. <https://doi.org/10.1016/j.joitmc.2024.100353>.
- McDonald, N., Johri, A., Ali, A. and Hingle, A. (2024) 'Generative Artificial Intelligence in Higher Education: Evidence from an Analysis of Institutional Policies and Guidelines', *arXiv preprint arXiv:2402.01659* [Preprint]. <http://arxiv.org/abs/2402.01659>.
- Meakin, L. (2024) 'Exploring the Impact of Generative Artificial Intelligence on Higher Education Students' Utilization of Library Resources', *Information Technology and Libraries*, 43(3). <https://doi.org/10.5860/ital.v43i3.17246>.
- Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D.E., Thierry-Aguilera, R. and Gerardou, F.S. (2023) 'Challenges and Opportunities of Generative AI for Higher Education as Explained by ChatGPT', *EDUCATION SCIENCES*, 13(9). <https://doi.org/10.3390/educsci13090856>.
- Mishra, P., Oster, N. and Henriksen, D. (2024) 'Generative AI, Teacher Knowledge and Educational Research: Bridging Short- and Long-Term Perspectives', *TechTrends*, 68(2), pp. 205–210. <https://doi.org/10.1007/s11528-024-00938-1>.
- Mizumoto, A. and Eguchi, M. (2023) 'Exploring the potential of using an AI language model for automated essay scoring', *Research Methods in Applied Linguistics*, 2(2), p. 100050. <https://doi.org/https://doi.org/10.1016/j.rmal.2023.100050>.
- Montebello, M. (2021) 'Personalized Learning Environments', in *2021 International Symposium on Educational Technology (ISET)*. IEEE, pp. 134–138. <https://doi.org/10.1109/ISET52350.2021.00036>.

- Nikolopoulou, K., Gialamas, V. and Lavidas, K. (2021) 'Habit, hedonic motivation, performance expectancy and technological pedagogical knowledge affect teachers' intention to use mobile internet', *Computers and Education Open*, 2, pp. 100041. <https://doi.org/https://doi.org/10.1016/j.caeo.2021.100041>.
- Pan, M. and Gao, W. (2021) 'Determinants of the behavioral intention to use a mobile nursing application by nurses in China', *BMC Health Services Research*, 21(1), pp. 228. <https://doi.org/10.1186/s12913-021-06244-3>.
- Pedro, F., Subosa, M., Rivas, A. and Valverde, P. (2019) 'Artificial intelligence in education: Challenges and opportunities for sustainable development'. Unesco. <https://doi.org/https://hdl.handle.net/20.500.12799/6533>.
- Perera, P. and Lankathilake, M. (2023) 'Preparing to Revolutionize Education with the Multi-Model GenAI Tool Google Gemini? A Journey towards Effective Policy Making', *Journal of Advances in Education and Philosophy*, 7(08), pp. 246–253. <https://doi.org/10.36348/jaep.2023.v07i08.001>.
- Pillai, R., Sivathanu, B., Metri, B. and Kaushik, N. (2024) 'Students' adoption of AI-based teacher-bots (T-bots) for learning in higher education', *Information Technology & People*, 37(1), pp. 328–355. <https://doi.org/10.1108/ITP-02-2021-0152>.
- Shawky, D. and Badawi, A. (2019) 'Towards a Personalized Learning Experience Using Reinforcement Learning', in *Machine learning paradigms: Theory and application*. Springer, pp. 169–187. [https://doi.org/10.1007/978-3-030-02357-7\\_8](https://doi.org/10.1007/978-3-030-02357-7_8).
- Silverman, B.G., Hanrahan, N., Huang, L., Rabinowitz, E.F. and Lim, S. (2016) 'Artificial Intelligence and Human Behavior Modeling and Simulation for Mental Health Conditions', in D.D.B.T.-A.I. in B. and M.H.C. Luxton (ed.) *Artificial Intelligence in Behavioral and Mental Health Care*. San Diego: Elsevier, pp. 163–183. <https://doi.org/10.1016/B978-0-12-420248-1.00007-6>.
- Sobaih, A.E.E., Elshaer, I.A. and Hasanein, A.M. (2024) 'Examining Students' Acceptance and Use of ChatGPT in Saudi Arabian Higher Education', *European Journal of Investigation in Health, Psychology and Education*, 14(3), pp. 709–721. <https://doi.org/10.3390/ejihpe14030047>.
- Susarla, A., Gopal, R., Thatcher, J.B. and Sarker, S. (2023) 'The Janus Effect of Generative AI: Charting the Path for Responsible Conduct of Scholarly Activities in Information Systems', *Information Systems Research*, 34(2), pp. 399–408. <https://doi.org/10.1287/isre.2023.ed.v34.n2>.
- Tafazoli, D. (2024) 'Exploring the potential of generative AI in democratizing English language education', *Computers and Education: Artificial Intelligence*, 7, pp. 100275. <https://doi.org/10.1016/j.caeai.2024.100275>.
- Tamilmani, K., Rana, N.P., Wamba, S.F. and Dwivedi, R. (2021) 'The extended Unified Theory of Acceptance and Use of Technology (UTAUT2): A systematic literature review and theory evaluation', *International Journal of Information Management*, 57, pp. 102269. <https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2020.102269>.
- Venkatesh, Morris, Davis and Davis (2003) 'User Acceptance of Information Technology: Toward a Unified View', *MIS Quarterly*, 27(3), pp. 425. <https://doi.org/10.2307/30036540>.
- Venkatesh, Thong and Xu (2012) 'Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology', *MIS Quarterly*, 36(1), pp. 157. <https://doi.org/10.2307/41410412>.
- Venkatesh, V. (2022) 'Adoption and use of AI tools: a research agenda grounded in UTAUT', *Annals of Operations Research*, 308(1–2), pp. 641–652. <https://doi.org/10.1007/s10479-020-03918-9>.
- Wach, K., Duong, C.D., Ejdy, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., Paliszkievicz, J. and Ziemba, E. (2023) 'The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT', *Entrepreneurial Business and Economics Review*, 11(2), pp. 7–30. <https://doi.org/10.15678/EBER.2023.110201>.
- Wang, K., Ruan, Q., Zhang, X., Fu, C. and Duan, B. (2024) 'Pre-Service Teachers' GenAI Anxiety, Technology Self-Efficacy, and TPACK: Their Structural Relations with Behavioral Intention to Design GenAI-Assisted Teaching', *Behavioral Sciences*, 14(5), pp. 373. <https://doi.org/10.3390/bs14050373>.
- Wang, Y. and Zhang, W. (2023) 'Factors Influencing the Adoption of Generative AI for Art Designing Among Chinese Generation Z: A Structural Equation Modeling Approach', *IEEE Access*, 11, pp. 143272–143284. <https://doi.org/10.1109/Access.2023.3342055>.
- Xiao, P., Chen, Y. and Bao, W. (2023) 'Waiting, Banning, and Embracing: An Empirical Analysis of Adapting Policies for Generative AI in Higher Education', *SSRN Electronic Journal* [Preprint]. <https://doi.org/10.2139/ssrn.4458269>.

## Appendix 1: Questionnaire

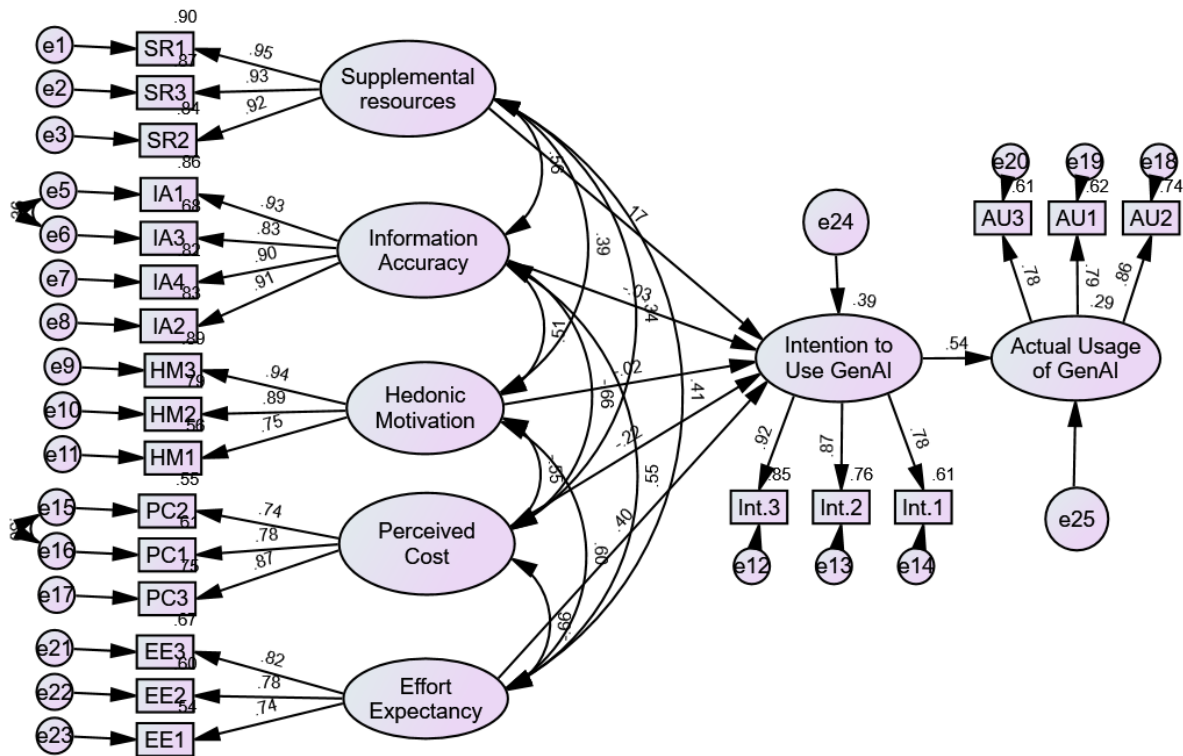
|                                                                                                                                                                                                                                                                                                                                                                          |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Part 1 Demographic Data</b><br><br><b>1- What is your gender?</b><br>Male<br>Female<br><br><b>2- What is your age?</b><br>18-30 years<br>31-41 years<br>42 yrs. and over<br><br><b>3- How often do you use GenAI tools?</b><br>Once times or less<br>2-3 times in week<br>More than 3 times<br><br><b>4- Do you use the paid version of GenAI tools?</b><br>Yes<br>No |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

### Part 2 Items

| Second Part                                                   |                                                                       |
|---------------------------------------------------------------|-----------------------------------------------------------------------|
| <b>Supplemental resource</b> (Chan and Lee, 2023)             |                                                                       |
| SR1                                                           | GenAI is valuable for answering queries.                              |
| SR2                                                           | GenAI helps generate thoughts.                                        |
| SR3                                                           | GenAI helps conduct analyses.                                         |
| <b>Information accuracy</b> (Chan and Lee, 2023)              |                                                                       |
| IA1                                                           | GenAI tools demonstrate biases in their answers                       |
| IA2                                                           | GenAI tools develop factually inaccurate answers                      |
| IA3                                                           | GenAI tools generate answers that are out of context or inappropriate |
| IA4                                                           | GenAI tools generate fake information                                 |
| <b>Effort expectancy</b> (Venkatesh, Thong and Xu, 2012)      |                                                                       |
| EE1                                                           | GenAI tools are easy to use                                           |
| EE2                                                           | Learning how to use GenAI tools is easy                               |
| EE3                                                           | Interaction with GenAI tools is unambiguous and understandable        |
| <b>Hedonic motivation</b> (Venkatesh, Thong and Xu, 2012)     |                                                                       |
| HM1                                                           | GenAI tools are enjoyable.                                            |
| HM2                                                           | Interacting with GenAI is pleasant.                                   |
| HM3                                                           | Using GenAI tools is fun.                                             |
| <b>Perceived cost</b> (Venkatesh, Thong and Xu, 2012)         |                                                                       |
| PC1                                                           | GenAI tools are affordably priced                                     |
| PC2                                                           | They provide good value for the money                                 |
| PC3                                                           | The free plan is better than a paid plan                              |
| <b>Intention to use GenAI</b> (Venkatesh, Thong and Xu, 2012) |                                                                       |
| Int.1                                                         | I intend to use GenAI tools frequently                                |
| Int.2                                                         | I plan to use GenAI tools daily.                                      |
| Int.3                                                         | I intend to continue using GenAI tools in the future.                 |

| Second Part                                           |                                            |
|-------------------------------------------------------|--------------------------------------------|
| Actual usage of GenAI (Venkatesh, Thong and Xu, 2012) |                                            |
| AU1                                                   | The GenAI tools are a pleasant experience. |
| AU2                                                   | I use the GenAI tools currently.           |
| AU3                                                   | I spend a lot of time using GenAI tools.   |

## Appendix 2: Structural Model



# Quality Assurance in Distance Higher Education: A Bibliometric Study of Scopus-Indexed Publications Between 1993 and 2024

Huu-Bich Nguyen<sup>1,3</sup>, Nguyen Quang Duy Vu<sup>2</sup>, Duc-Tai Dinh<sup>3</sup> and Hiep-Hung Pham<sup>4,5</sup>

<sup>1</sup>Ministry of Education and Training, Hanoi, Vietnam

<sup>2</sup>Bachelor of Japanese Studies Program, Faculty of Interdiscipline Social Sciences, Vietnam Japan University, Vietnam National University, Hanoi, Vietnam

<sup>3</sup>The Vietnam Institute of Educational Sciences (VNIES), Hanoi, Vietnam

<sup>4</sup>REDUVATION Research Unit, Thanh Do University, Hanoi, Vietnam

<sup>5</sup>VNU University of Education, Hanoi, Vietnam

[nhbich@moet.gov.vn](mailto:nhbich@moet.gov.vn)

[duyvng@gmail.com](mailto:duyvng@gmail.com)

[taidd@vnies.edu.vn](mailto:taidd@vnies.edu.vn) (corresponding author)

[hiep@thanhdouni.edu.vn](mailto:hiep@thanhdouni.edu.vn)

<https://doi.org/10.34190/ejel.23.1.3632>

An open access article under [CC Attribution 4.0](#)

**Abstract:** This bibliometric study investigates the evolution and trends in quality assurance in distance higher education (QADHE) through a comprehensive analysis of Scopus-indexed publications spanning from 1993 to 2024. The research aims to address several critical aspects, including publication trends, international collaboration networks, influential authors, primary sources, emerging themes, and keywords within the field of QADHE. Utilizing PRISMA methods, a total of 193 relevant documents were identified, reflecting a growing interest in QADHE over the past three decades. The findings reveal distinct phases in the publication trend, indicating an increasing focus on quality assurance practices in distance higher education. Notably, the significant contributions come from Anglophone countries, underscoring the influence of English-speaking nations in shaping the discourse surrounding QADHE. Furthermore, the collaboration networks highlight a global interest in QADHE, with notable contributions from countries in Asia. This collaborative effort illustrates the importance of cross-border partnerships in advancing research and practices related to quality assurance in distance education. The analysis of influential authors and institutions underscores the diverse geographical distribution of research contributions, featuring prominent figures from both developed and developing nations. This diversity enriches the body of knowledge in QADHE and fosters a more inclusive understanding of quality assurance challenges and solutions across different contexts. The study also identifies key themes within QADHE, such as the integration of technology, methodologies for quality assurance (QA), implementation strategies, and regional perspectives. These themes reflect the dynamic nature of the field and the evolving landscape of distance education, which has been significantly impacted by technological advancements and changing educational paradigms. Additionally, an analysis of co-occurrence keywords indicates a growing emphasis on "open" principles and the utilization of Open Education Resources (OER) to enhance distance higher education (DHE). This shift towards openness suggests a transformative approach to education that prioritizes accessibility and collaboration. Overall, this study provides valuable insights into the evolution and current state of QADHE, highlighting areas for future research and the importance of continuous quality assurance efforts. The findings not only contribute to the existing literature but also serve as a foundation for understanding the implications of quality assurance practices in distance education. As the field continues to evolve, ongoing research is essential to address emerging challenges and to promote effective quality assurance frameworks that can adapt to the needs of diverse educational contexts.

**Keywords:** Quality assurance, Distance higher education, Bibliometric analysis, Open education, Online learning

## 1. Introduction

Over the past decades, distance higher education (DHE) programs have gained significant prominence as viable alternatives for students seeking university degrees, moving away from traditional offline learning (Martin, Sun and Westine, 2020). These programs can be offered by brick-and-mortar universities or as fully virtual institutions (Moore, Dickson-Deane and Galyen, 2011). According to Garrett (2020), in 2019, out of approximately 225 million students globally enrolled in higher education, about 17 million were engaged in DHE, with nine million participating in online provision.

Several drivers have contributed to the rise of DHE, with technological advancements being a key factor. The widespread availability of the Internet allows students to enhance their education without needing to attend classes in person (Keržič *et al.*, 2021). DHE offers flexibility that traditional universities often cannot, creating an

ideal learning environment for many (Ferrer *et al.*, 2022). The COVID-19 pandemic has further accelerated this shift toward online education, significantly impacting future educational delivery methods (Zhao and Watterston, 2021). Additionally, innovative technologies like artificial intelligence are enhancing the virtual learning experience (Ouyang, Zheng and Jiao, 2022).

Within DHE in particular, prior scholars have also paid their attention on the topic of quality assurance (Stella and Gnanam, 2004; Belawati and Zuhairi, 2007; Scull *et al.*, 2011; Darajat, 2018; Hannache-Heurteloup and Moustaghfir, 2020; Sarıtaş, Börekci and Demirel, 2022). This is because it encompasses the development and creation of learning materials, the design of academic programs, the provision of services and support, as well as the setting and maintaining of student learning standards (Zawacki-Richter and Anderson, 2014). Nevertheless, to the best of our knowledge, there has yet been any systematic work which may provide a comprehensive view on the state of current literature on QADHE. Therefore, this study aims to address this research gap by attempting to answer the following research questions.

**RQ1:** *What are the publication trends in the field of QADHE over the period from 1993 to 2024?*

**RQ2:** *What does the network of international collaboration in this topic look like?*

**RQ3:** *Who are the most influential authors in terms of the number of publications and citations, and what does the research community in the field of QADHE look like?*

**RQ4:** *What are the primary sources of publications in this field, and what are the most influential documents in the literature of QADHE?*

**RQ5:** *What are the main themes in the field of QADHE, and what new research orientations are emerging?*

Furthermore, it is important to note that this study specifically focuses on DHE, which entails learning exclusively from a remote location, distinct from blended learning models combining online and offline components or online learning as a part of conventional educational programs (Moore, Dickson-Deane and Galyen, 2011).

## 2. Literature Review

Quality assurance in distance higher education has been a topic of growing interest and importance over the past few decades, paralleling the rapid expansion of distance and online learning modalities.

Initial research on QADHE mainly centered on comparing the quality of distance education with traditional in-person instruction. Scholars like Berge and Mrozowski (2001) and Zawacki-Richter, Baecker and Vogt (2009) conducted extensive reviews of distance education studies, underscoring the need for stronger quality assurance practices. As technology advanced, researchers shifted their focus to specific elements of distance education. For instance, Lockhart and Lacy (2002) proposed a framework stressing the significance of student support, course design, and faculty development in distance settings. Meanwhile Rovai (2003) explored factors such as student outcomes, costs, technology, and satisfaction for both students and faculty.

The integration of e-learning technologies into distance education and the surge of open education brought new challenges and opportunities for quality assurance (Ossiannilsson, Williams and Brown, 2015). Stella and Gnanam (2004) addressed concerns from stakeholders, including governments and quality assurance agencies, regarding how to ensure best practices, with differing opinions on how to conduct assurance processes and which criteria should be used in online settings. The rapid growth of Massive Open Online Courses (MOOCs) in the 2010s also sparked new debates and research directions in QADHE. Scholars like Lowenthal and Hodges (2015) and Xiao, Qiu and Cheng (2019) examined quality assurance challenges specific to MOOCs, such as scalability, learner engagement, and credential recognition.

Another key area of research has concentrated on creating and applying quality assurance models specifically designed for distance education. Various methods and tools have been introduced to maintain quality in distance learning programs, such as course evaluations, self-assessments, external reviews, and stakeholder satisfaction metrics (Gaftandzhieva, Doneva and Jagatheesaperumal, 2023). Recent research has also established benchmarks for external QA systems in distance education, with slight adjustments based on local requirements. However, technological advancements may require updates to existing standards and practices (Bukhari, Shah and Arif, 2021).

Although the literature on QADHE has expanded, there are still several gaps and areas that require further exploration. These include the need for more empirical research on the long-term effects of quality assurance

measures on student learning outcomes in distance education; comparative studies of QADHE practices across diverse cultural and institutional settings (Mkwizu and Junio-Sabio, 2024); investigation into quality assurance methods for newer forms of distance education, including competency-based education and micro-credentials (Brown and Duart, 2024). Additionally, the impact of artificial intelligence on QADHE is a particularly important area for future research.

This bibliometric analysis aims to contribute to the existing literature by providing a comprehensive overview of the research landscape in QADHE, highlighting trends, key publications, and emerging themes. By tracing the development of scholarship in this area, the study aims to uncover potential directions for future research in quality assurance for distance higher education.

### **3. Method**

Bibliometric analysis is a powerful method proposed by Pritchard (1969). This method is used for examining the quantitative aspects of research literature, provides a comprehensive overview of the academic landscape, highlighting trends, collaboration networks, key contributors, influential works, and emerging topics. Scopus was chosen due to its extensive coverage of global research literature across various disciplines (Falagas *et al.*, 2008) ensures a more inclusive and representative sample for this study on QADHE.

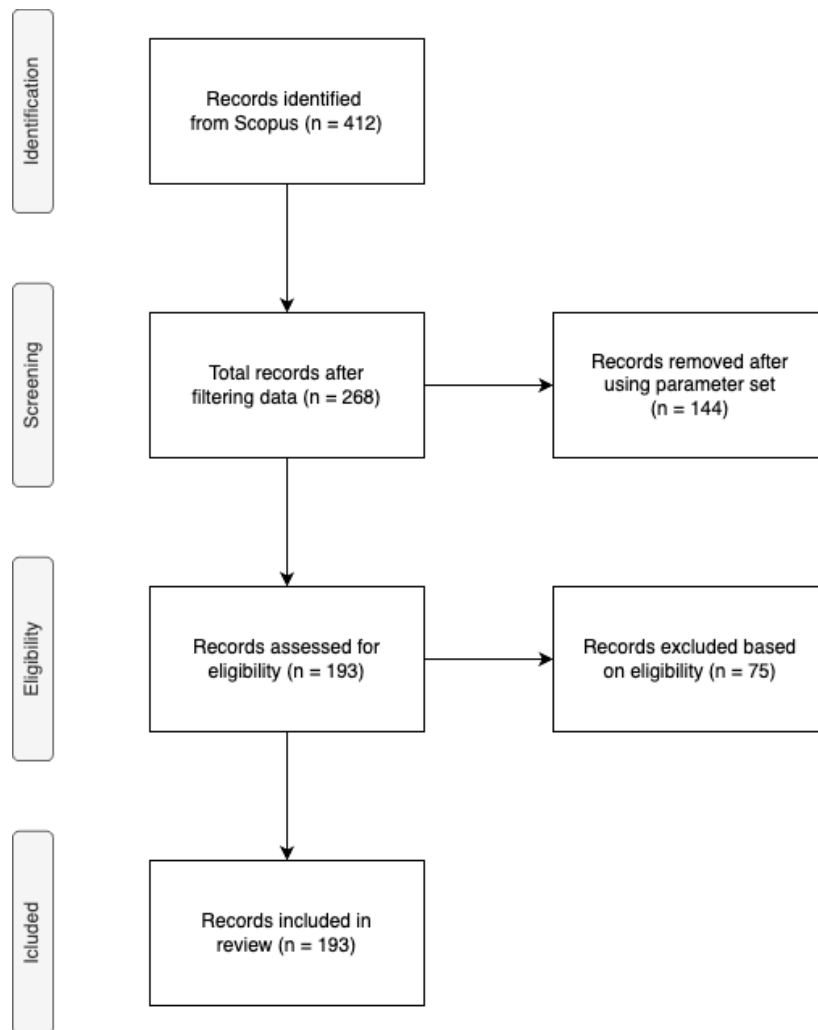
The literature search utilized the keywords "quality assurance," "distance education," and "distance learning." Initially, a total of 378 documents were identified. Subsequently, a more detailed screening process was implemented, which included criteria based on subject area, document types, and publication stage. Papers published in 2023 were excluded due to the incomplete count of papers for that year, as the search query was conducted on June 14, 2023, at 17:00. After applying these criteria, 136 documents were excluded, leaving 242 documents that met all the above criteria. The search query used was as follows:

```
TITLE-ABS-KEY ( "quality assurance" AND ( "distance education" OR "distance learning" ) ) AND ( LIMIT-TO ( PUBSTAGE , "final" ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) OR LIMIT-TO ( DOCTYPE , "cp" ) OR LIMIT-TO ( DOCTYPE , "ch" ) OR LIMIT-TO ( DOCTYPE , "bk" ) ) AND ( EXCLUDE ( PUBYEAR , 2023 ) ) AND ( LIMIT-TO ( LANGUAGE , "english" ) ) AND ( LIMIT-TO ( SUBJAREA , "soci" ) OR LIMIT-TO ( SUBJAREA , "busi" ) OR LIMIT-TO ( SUBJAREA , "agri" ) OR LIMIT-TO ( SUBJAREA , "deci" ) OR LIMIT-TO ( SUBJAREA , "arts" ) OR LIMIT-TO ( SUBJAREA , "econ" ) OR LIMIT-TO ( SUBJAREA , "psyc" ) OR LIMIT-TO ( SUBJAREA , "mult" ) ) )
```

The content of each record was then examined to ensure relevance to QADHE. After eliminating 43 non-relevant documents (e.g., Vrasidas, 2003; Booker *et al.*, 2011; Wyk, 2018), the first dataset for analysis consisted of 173 documents.

To incorporate the most recent data, a second extraction was performed on December 31, 2024, at 19:30, using the same query. This process identified 34 records from 2023 and 2024. After screening, 8 publications were excluded, leaving 26 records. These were cross-checked against the existing dataset, identifying 6 duplicates. Consequently, 20 unique records were added, yielding a final dataset of 193 documents for analysis.

The selection processes adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) method, which is a widely used and transparent approach for conducting systematic reviews and meta-analyses (Moher *et al.*, 2010). The PRISMA flow diagram is shown in Figure 1.



**Figure 1: Selection flow diagram for the studies**

All data management and initial processing were done in Microsoft Excel, including data cleaning, author name standardization, and removal of duplicates. Descriptive analysis was utilized to examine publication trends, document types, and geographical distribution. VOSviewer software, which is a common software in bibliometric analysis, was used to construct and visualize bibliometric networks. Co-authorship data were analyzed to map international collaboration at country and institutional levels, while influential authors and institutions were identified using publication and citation metrics. The analysis of main publication sources considered document quantity and citation impact. Co-citation analysis is also utilized to identify clusters of related documents, and keyword co-occurrence analysis captures evolving trends in QADHE.

## 4. Results and Discussion

### 4.1 Publication Trend and Type

QADHE has emerged as a new research topic in the last three decades. The publication trend of this topic shows a strong growth, but also a high fluctuation in the yearly document counts (see Figure 2). The trend can be divided into three phases:

- Phase 1 (1993-1999): This phase represents the initial stage of research on QADHE, with very few publications. Only one study was published each year in 1993, 1994, and 1995. No publications were recorded in 1996, 1997, and 1999. The only exception was 1998, which had three publications.
- Phase 2 (2000-2015): This phase marks the rapid development of research on QADHE, with an increasing number of publications. The number of documents rose from one in 2000 to two in 2001, three in 2002, and reached a peak of 13 studies in 2004 and 2013. However, the growth was not consistent, as the number of studies varied greatly from year to year, with a low of two studies in 2008, and a range of five to 12 studies per year for the rest.

- Phase 3 (2016-2024): This phase shows a more stable pattern of research on QADHE, with a relatively high number of publications and no sharp declines. The number of documents fluctuated between eight and 12 studies per year. The years 2021 and 2022 saw a slight decrease in the number of publications, which could be attributed to the pandemic that partly affected research activities.

The observed trend aligns with previous findings in distance education (Amoozegar, Khodabandelou and Ebrahim, 2018; Ndibalema, 2022), educational technology (Rodríguez Jiménez, Sanz Prieto and Alonso García, 2019; Shen and Ho, 2020), and quality assurance (Khuram *et al.*, 2023), suggesting an interconnection among these fields. This synergy, especially evident in the early 21st century, likely stems from technological advancements driving DHE growth and necessitating stronger QA mechanisms (Uvalić-Trumbić and Martin, 2021). However, the fluctuation in the number of publications on QADHE highlights a unique characteristic of this field. Compared to the mentioned topics, the number of research in QADHE is relatively low, making even slight changes in the publication count lead to significant fluctuations in the overall trend. As a review by Martin, Sun and Westine (2020), QA remains one of the least researched topics in research on online teaching and learning, highlighting the need for more scholarly focus on QADHE.

In terms of publication types, journal articles comprise the largest portion of the dataset, accounting for 55.44% (107 publications) of the total count. Following closely, book chapters contribute 25.91% (40 publications), while conference papers and books constitute 10.88% (21 publications) and 7.77% (15 publications), respectively. This pattern is not unique to QADHE but can also be observed in other areas of educational research, such as lifelong learning (Do *et al.*, 2021), learning management systems (Thi Phan *et al.*, 2022), education leadership and management (Hallinger and Kovačević, 2021).

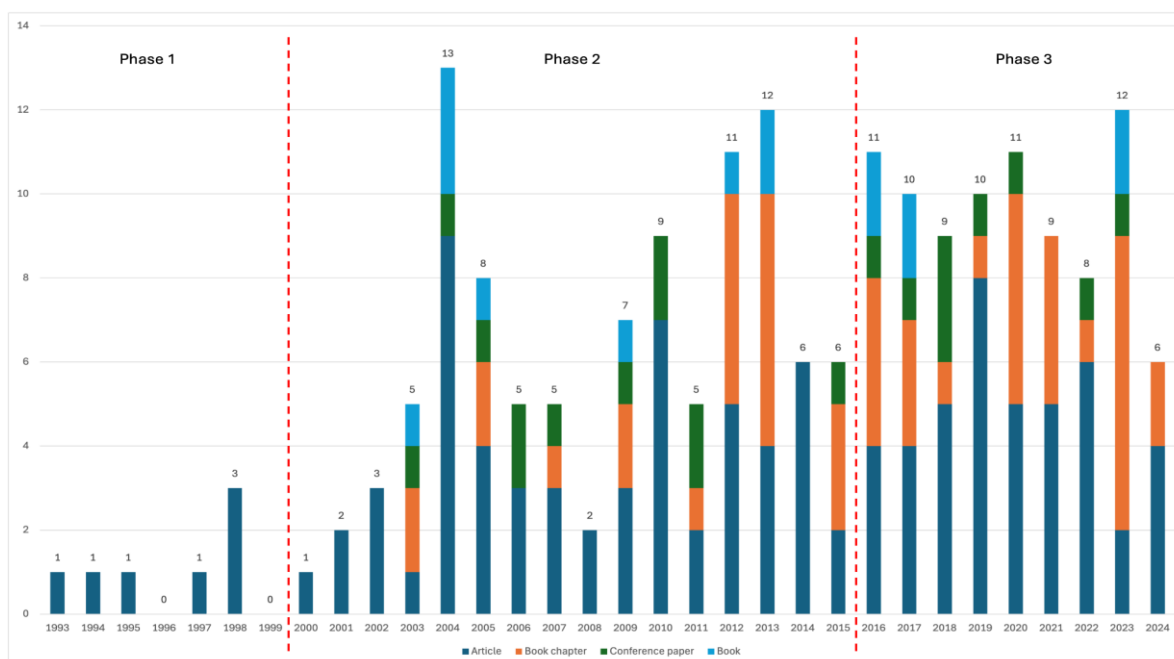


Figure 2: Annual publications related to QADHE from 1993 to 2024

#### 4.2 Network of International Collaboration

Analyzing the landscape of collaboration among countries in the realm of QADHE reveals a global interest in the subject. However, there exist notable disparities in the levels of contribution, impact, interconnectivity, and collaborative endeavors across nations.

The scientific map generated by VOSViewer depicts global scientific collaboration, showcasing clusters of countries with varying extent of research ties (see Figure 3). Larger nodes signify countries with significant research output, while thicker lines connecting nodes indicate closer collaboration and the number of lines reflects the extent of connections with other countries. Node color reflects the average publication year, indicating each country's active research period.

As shown in Figure 3, Anglophone countries like the United States, the United Kingdom, and Australia are significant contributors to QADHE research, with publication counts of 29, 23, and 21 documents respectively

(see Table 1). These nations also lead in citations, indicating their influence and robust international collaboration, reflected in their strong connections on the scientific map. Their central roles stem from established distance learning systems (Brown, 2011), which provide conducive environments for research on DHE and QA. Moreover, they lead in offering numerous MOOCs courses and platforms, such as Coursera, edX, FutureLearn, Open2Study (Baturay, 2015), contributing significantly to research on ensuring quality in these areas.

South Africa, alongside the US, UK, and Australia, is a leading player in QADHE research (23 publications), and holds a central position in the collaborative network, reflecting its strong international connectivity. South Africa is also home to the world's oldest dedicated DHE institution, the University of South Africa (UNISA) (Tait, 2018; UNISA, n.d.), which has significantly shaped the country's academic landscape and contributed to its growing scholarly presence. However, despite its publication volume, South Africa ranks tenth in citations with 61, indicating a gap between output and influence. This could be because South African studies often focus on practical, context-specific challenges (Bornman, 2004; Matlakala and Maritz, 2019; Naidoo and Kemlall, 2019; Zawada, 2019). The emphasis on practicality may result in a narrower scope of applicability beyond the local context, leading to fewer citations from researchers operating in diverse international contexts.

The expanding network of QAHDE, shown by the diverse colors from blue to yellow, includes notable contributors like India, Japan, Canada, Finland, Indonesia, Brazil, Germany, Nigeria, Malaysia, and China, which enrich the field with various perspectives. Many of these countries rank among the top 10 in publications or citations, emphasizing their influence in shaping QADHE. As they increasingly rely on DHE to address infrastructure challenges and diverse student needs (Darojat, Nilson and Kaufman, 2015; Zuhairi, Raymundo and Mir, 2020), the demand for effective QA of these emerging programs intensifies, fueling a surge of interest and investment in this area (Wright, Dhanarajan and Reju, 2009).

In contrast, countries like Turkey, Egypt, and Vietnam appear as isolated nodes on the scientific map, indicating their early-stage involvement in the collaborative network. Vietnam faces challenges such as inadequate infrastructure, lack of program accreditation, and low societal awareness (Nguyen, Ta and Nguyen, 2022). However, these challenges are not unique to Vietnam but are shared by both developing and developed nations (Uvalić-Trumbić and Martin, 2021). Moreover, the recent impetus from the COVID-19 pandemic has heightened attention towards DHE in these countries (Hebebcı, 2021; Pham and Ho, 2020), indicating potential for growth and the opportunity to join the international network, underscoring the need for further integration and collaboration in research efforts.

**Table 1: Top 10 countries sorted by number of documents and number of citations in knowledge base of QADHE from 1993 to 2024**

| Rank | Country            | Number of Documents | Rank | Country            | Number of Citations |
|------|--------------------|---------------------|------|--------------------|---------------------|
| 1    | The United States  | 29                  | 1    | The United States  | 745                 |
| 2    | South Africa       | 23                  | 2    | The United Kingdom | 320                 |
| 3    | The United Kingdom | 21                  | 3    | Australia          | 271                 |
| 4    | Australia          | 16                  | 4    | Japan              | 137                 |
| 5    | India              | 13                  | 5    | India              | 123                 |
| 6    | Canada             | 10                  | 6    | Germany            | 118                 |

| Rank | Country   | Number of Documents | Rank | Country      | Number of Citations |
|------|-----------|---------------------|------|--------------|---------------------|
| 7    | Turkey    | 10                  | 7    | Nigeria      | 86                  |
| 8    | Japan     | 8                   | 8    | Malaysia     | 85                  |
| 9    | Indonesia | 7                   | 9    | Indonesia    | 71                  |
| 10   | Brazil    | 5                   | 10   | South Africa | 61                  |

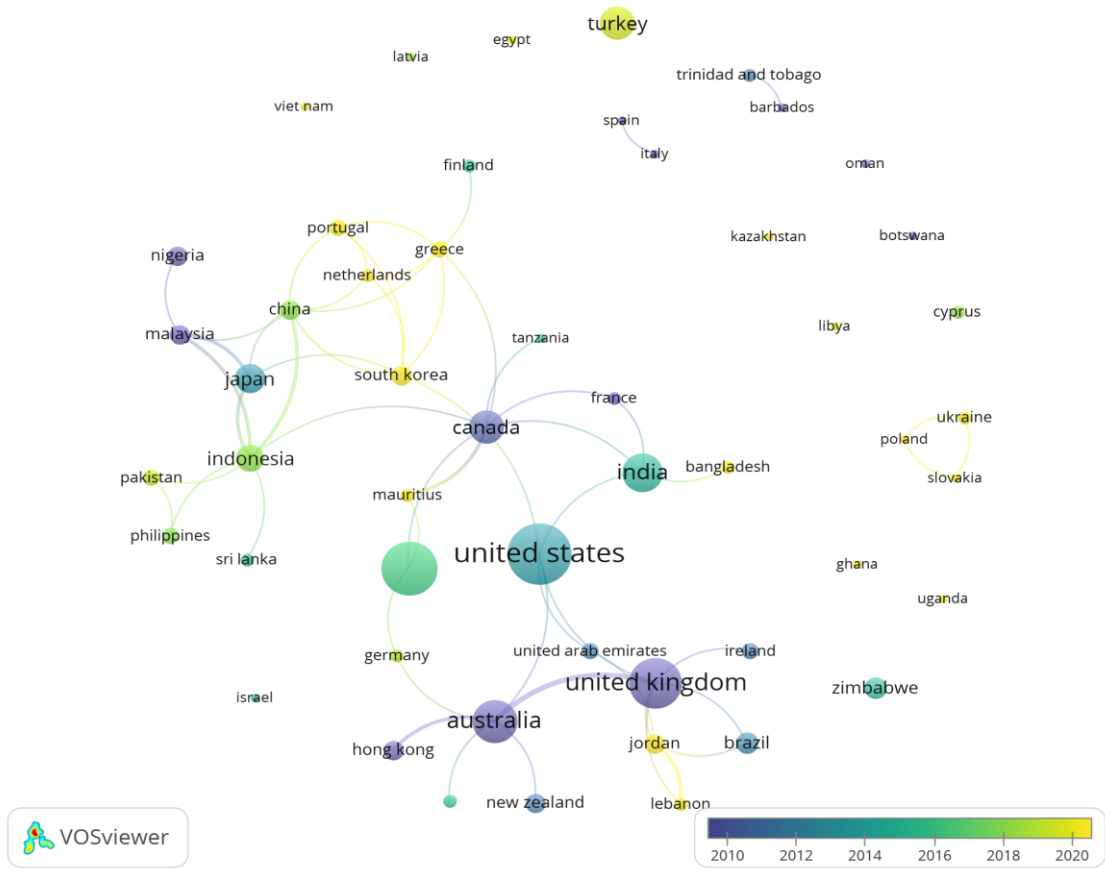


Figure 3: Countries' collaborations in knowledge base of QADHE from 1993 to 2024

Upon closer examination of the institutions spearheading research on QADHE, a consistent pattern emerges. The engagement of numerous distance universities worldwide underscores the global interest in the topic and the pressing need for tailored QA measures to accommodate their unique delivery modes.

UNISA, a longstanding leader in distance education, continues to rank first globally with 13 publications (see Table 2). As the sole distance education university in South Africa, formed through a merger in the 2000s (Prinsloo, 2019), its prolific output reflects its prominent role in advancing DHE development. However, similar to previous assessments, UNISA does not feature among the top institutions by citation count. This discrepancy highlights a persistent challenge: while the institution remains committed to traditional correspondence education—leveraging radio and television for distance learning—it has been slower to integrate contemporary

technologies like the internet. This lag in innovation, as noted by Qayyum and Zawacki-Richter (2019), has contributed to the reduced global influence and relevance of South African research in the field, as evidenced by lower citation numbers.

Significant contributions continue to emanate from universities across the globe, reflecting the widespread interest in QADHE. Notable participants include institutions from Japan, the United Kingdom, New Zealand, and regions in the Global South, such as Nigeria, Indonesia, Zimbabwe, and India. Five of the listed institutions specialize exclusively in distance learning, emphasizing their internal need to improve quality assurance processes, which in turn drives substantial scholarly output on QA in DHE.

Interestingly, institutions from English-speaking countries dominate the list of affiliations with the highest citation counts. American, British, and Australian universities occupy leading positions, suggesting that research originating from these regions may emphasize theoretical contributions that resonate broadly within the academic community. In contrast, institutions with a stronger focus on practical applications and case studies—often situated in the Global South—exhibit fewer citations due to their less theoretical nature.

International Christian University in Japan represents a standout case. Despite producing only seven publications, it achieves the highest citation count (134), underscoring the significant influence and scholarly recognition of its contributions. This phenomenon, along with other notable trends, will be explored further in the subsequent section.

**Table 2: Top 10 affiliations sorted by number of documents and number of citations in knowledge base of QADHE from 1993 to 2024**

| No | Affiliation, Country                              | Documents | No | Affiliation, Country                              | Citations |
|----|---------------------------------------------------|-----------|----|---------------------------------------------------|-----------|
| 1  | University of South Africa, South Africa (*)      | 13        | 1  | International Christian University, Japan         | 134       |
| 2  | International Christian University, Japan         | 7         | 2  | National Open University of Nigeria, Nigeria (*)  | 84        |
| 3  | The Open University, the United Kingdom           | 6         | 3  | Wawasan Open University, Malaysia (*)             | 75        |
| 4  | Universitas Terbuka, Indonesia (*)                | 5         | 4  | Universitas Terbuka, Indonesia (*)                | 70        |
| 5  | Anadolu University, Turkey                        | 5         | 5  | The Open University, the United Kingdom           | 58        |
| 6  | Zimbabwe Open University, Zimbabwe (*)            | 5         | 6  | University of Sydney, Australia                   | 48        |
| 7  | Indira Gandhi National Open University, India (*) | 4         | 7  | Beijing Normal University, China                  | 43        |
| 8  | National Open University of Nigeria, Nigeria (*)  | 3         | 8  | RMIT University, Australia                        | 33        |
| 9  | University of Nottingham, the United Kingdom      | 3         | 9  | Indira Gandhi National Open University, India (*) | 25        |
| 10 | Victoria University of Wellington, New Zealand    | 3         | 10 | University of Nottingham, the United Kingdom      | 23        |

Note: (\*) denotes universities only offering distance learning programs

### 4.3 Influential Authors and Research Community

When examining the list of authors with the highest number of publications and citations, a similar pattern emerges as when analyzing the list of authors or countries with the highest number of publications and citations.

This pattern highlights the contributions of authors from diverse geographical regions worldwide. However, authors from institutions in the United State consistently lead in this regard.

Specifically, Jung I. stands out as the foremost author in terms of document count with eight documents (see Table 3), representing International Christian University in Japan. Additionally, their citation count is noteworthy, ranking third with 134 citations, further highlighting the impact of their research. The author's prominence significantly contributes to the high ranking of both their institution and Japan overall in terms of documents and citations.

South Africa is also well-represented through UNISA, with three authors—Letseka M., Mahlangu V.P., and Belawati T.—among the most prolific contributors. This reflects the country's sustained focus on advancing quality assurance in distance higher education. Additionally, authors from both developed and developing nations such as Australia, Indonesia, and Turkey are featured, highlighting the widespread global engagement in this field. However, the dominance of highly cited authors from the United States is evident. T.C. Reeves and Y. Woo from the University of Georgia topped the citation rankings with 294 citations each, followed by contributors from Regent University and other prominent American institutions. This concentration of highly cited authors highlights the centrality of U.S.-based researchers in shaping the theoretical and conceptual frameworks of QADHE, reinforcing a consistent trend observed across the research network.

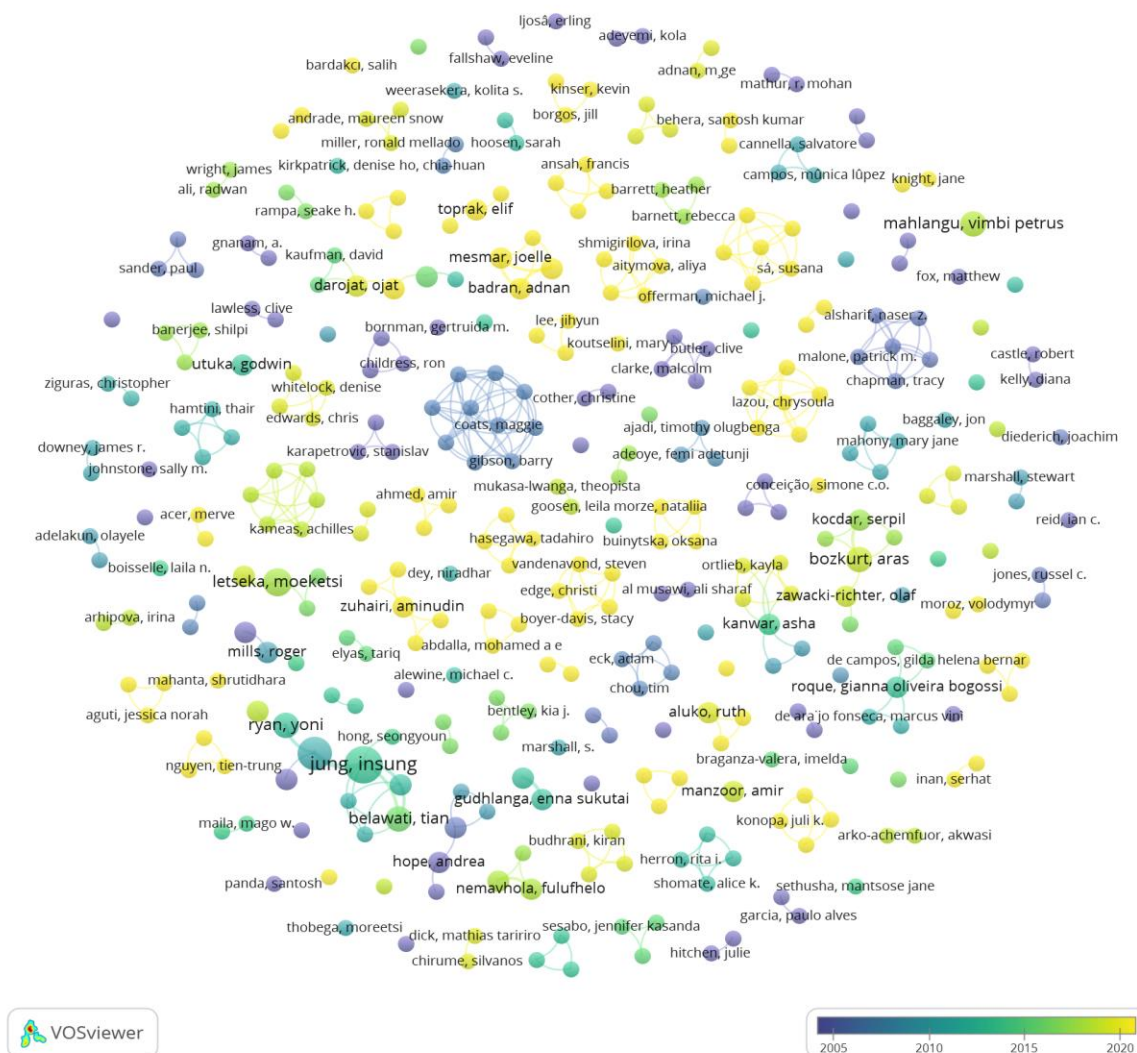
**Table 3: Authors with more than two publications and Top 10 author sorted by number of citations in knowledge base of QADHE from 1993 to 2024**

| No | Author        | Affiliation, Country                      | Number of Documents | No | Author             | Affiliation, Country                                      | Number of Citations |
|----|---------------|-------------------------------------------|---------------------|----|--------------------|-----------------------------------------------------------|---------------------|
| 1  | Jung I.       | International Christian University, Japan | 8                   | 1  | Reeves T.C.        | The University of Georgia, the United States              | 294                 |
| 2  | Latchem C.    | Australian Catholic University, Australia | 6                   | 2  | Woo Y.             | The University of Georgia, the United States              | 294                 |
| 3  | Letseka M.    | University of South Africa, South Africa  | 4                   | 3  | Downey J.R.        | Regent University, the United States                      | 175                 |
| 4  | Belawati T.   | Universitas Terbuka, Indonesia            | 3                   | 4  | Rovai A.P.         | Regent University, the United States                      | 175                 |
| 5  | Bozkurt A.    | Anadolu University, Turkey                | 3                   | 5  | Budhrani K.        | University of North Carolina Charlotte, the United States | 139                 |
| 6  | Mahlangu V.P. | University of South Africa, South Africa  | 3                   | 6  | Kumar S.           | University of Florida, the United States                  | 139                 |
| 7  | Ryan Y.       | Australian Catholic University, Australia | 3                   | 7  | Martin F.          | University of North Carolina Charlotte, the United States | 139                 |
|    |               |                                           |                     | 8  | Ritzhaupt A.       | University of Florida                                     | 139                 |
|    |               |                                           |                     | 9  | Jung I.            | International Christian University, Japan                 | 134                 |
|    |               |                                           |                     | 10 | Zawacki-Richter O. | Carl von Ossietzky University of Oldenburg, Germany       | 118                 |

Stepping back to observe the research community at large, Figure 4 provides an overview of the landscape within the field of QADHE. Each node represents a distinct group of authors who collaborate frequently.

The biggest clusters, which indicate the most prominent group, is led by Jung I., the author with the most published documents in the field. This group also includes or is closely associated with other prolific authors such as Latchem C., Ryan Y., and Belawati T., who are among the top 10 authors with the most documents. The prominence of this group underlines its significant role in shaping the discourse in this field. However, the average publication year for this group, which falls between 2010 and 2015, suggests a decline in its activity in recent years.

The map's significant number of yellow nodes, comprising nearly half, points to a growing cohort of researchers who have entered the field post-2016, marking a key milestone in the domain's development. This influx of newcomers with their fresh perspectives (Zuhairi, Raymundo and Mir, 2020; Mekky, 2021), research inquiries (Kocdar, Okur and Bozkurt, 2017; Littlefield, Rubinstein and Laveist, 2019; Andrade *et al.*, 2020), models and methods (Crisanto, 2018; Mystakidis, Berki and Valtanen, 2019) contribute to the evolving landscape of QADHE.



**Figure 4: Scholars' communication in knowledge base of QADHE from 1993 to 2024**

#### 4.4 Primary Sources and Influential Documents

Considering the primary sources on the topic of QADHE, it is notable that most of the journals listed specialize in distance/online education, and a small portion is exclusively dedicated to QA (see Table 4). Given the rising trend in publications on this topic and the substantial influx of new research groups, there is a growing need for

a dedicated journal focusing exclusively on QADHE to become a primary source for research in this area. Moreover, edited books on this topic stand out as significant sources, with four ranking among the top in document volume and one notable for high citation counts.

**Table 4: Top 10 sources sorted by number of documents and number of citations in knowledge base of QADHE from 1993 to 2024**

| Rank | Source                                                                                                         | Scope                     | Number of Documents | Rank | Source                                                                   | Scope                     | Number of Citations |
|------|----------------------------------------------------------------------------------------------------------------|---------------------------|---------------------|------|--------------------------------------------------------------------------|---------------------------|---------------------|
| 1    | Turkish Online Journal of Distance Education (Journal)                                                         | Online/distance education | 13                  | 1    | Internet and Higher Education (Journal)                                  | Higher Education          | 653                 |
| 2    | International Review of Research in Open and Distance Learning (Journal)                                       | Online/distance education | 11                  | 2    | International Review of Research in Open and Distance Learning (Journal) | Online/distance education | 215                 |
| 3    | Asian Association of Open Universities Journal (Journal)                                                       | Online/distance education | 8                   | 3    | Distance Education (Journal)                                             | Online/distance education | 136                 |
| 4    | Distance Education (Journal)                                                                                   | Online/distance education | 7                   | 4    | Quality Assurance in Education (Journal)                                 | Quality in Education      | 102                 |
| 5    | Quality Assurance and Accreditation in Distance Education and E-Learning: Models, Policies and Research (Book) | N/A                       | 6                   | 5    | Higher Education (Journal)                                               | Higher Education          | 79                  |
| 6    | Handbook of Open, Distance and Digital Education (Book)                                                        | N/A                       | 5                   | 6    | British Journal of Educational Technology (Journal)                      | Educational Technology    | 78                  |
| 7    | Quality Assurance in Education (Journal)                                                                       | Quality in Education      | 5                   | 7    | Distance and Blended Learning in Asia (Book)                             | N/A                       | 65                  |
| 8    | Assuring Institutional Quality in Open Distance Learning (ODL) in The Developing Contexts (Book)               | N/A                       | 4                   | 8    | Asian Association of Open Universities Journal (Journal)                 | Online/distance education | 57                  |

| Rank | Source                                                                                          | Scope            | Number of Documents | Rank | Source                                                     | Scope                     | Number of Citations |
|------|-------------------------------------------------------------------------------------------------|------------------|---------------------|------|------------------------------------------------------------|---------------------------|---------------------|
| 9    | Internet and Higher Education (Journal)                                                         | Higher Education | 4                   | 9    | Open Learning (Journal)                                    | Online/distance education | 52                  |
| 10   | Research Anthology on Preparing School Administrators to Lead Quality Education Programs (Book) | N/A              | 4                   |      | Turkish Online Journal of Educational Technology (Journal) | Online/distance education | 53                  |

(N/A: Not available)

Table 5 showcases the top 10 most influential documents in QADHE, ranked by citation count. At the forefront is the study by Woo and Reeves (2007), offering a new view on online interaction through social constructivism. Following closely is the article authored by Rovai and Downey (2010), cited 175 times, examines factors affecting online program success, including planning, QA, and course design. Additionally, one book by Latchem and Jung (2009) is featured among the top 10, providing a regional perspective to the literature of QADHE. Notably, the majority of these publications are featured in the prestigious journal *Internet and Higher Education*, recognized in both the fields of online education and education in general. It is noteworthy that a significant portion of the highly cited works is authored by Western scholars, which reinforces previous observations regarding publication patterns between Western authors and those from developing countries.

**Table 5: Top 10 documents sorted by number of citations in knowledge base of QADHE from 1993 to 2024**

| No | Document                         | Document Type | Author Keywords                                                                                                                                                   | Source                                | Citation |
|----|----------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------|----------|
| 1  | Woo and Reeves (2007)            | Article       | Authentic task; Meaningful interaction; Online interaction; Social constructivism; Web-based learning environment                                                 | Internet and Higher Education         | 294      |
| 2  | Rovai & Downey (2010)            | Article       | Academic capitalism; Academic globalization; Global learning; Higher education; International strategic enrollment management; Online learning; Quality assurance | Internet and Higher Education         | 175      |
| 3  | Martin et al (2019)              | Article       | Award-winning faculty; Course design; Evaluation facilitation strategies; Online assessment; Online teaching; Qualitative research                                | Internet and Higher Education         | 139      |
| 4  | Zawacki-Richter and Naidu (2016) | Article       | Content analysis; distance education; mapping trends; research and scholarship; text mining                                                                       | Distance Education                    | 108      |
| 5  | Roffe (2002)                     | Article       | Distance learning; Evaluation; Quality assurance                                                                                                                  | Quality Assurance in Education        | 91       |
| 6  | Latchem and Jung (2009)          | Book          | N/A                                                                                                                                                               | Distance and Blended Learning in Asia | 65       |

| No | Document                 | Document Type | Author Keywords                                                                                                                                           | Source                                           | Citation |
|----|--------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|----------|
| 7  | Stella and Gnanam (2004) | Article       | Distance education; Distributed learning; E-learning; New forms of learning; On-line learning; Quality assurance of/challenges of                         | Higher Education                                 | 64       |
| 8  | Ho and Swan (2007)       | Article       | Assessment; Asynchronous learning; Content analysis; Cooperative principle; Direct response; Learning outcome; Online discussion; Socio-cognitive process | Internet and Higher Education                    | 45       |
| 9  | Ajadi et al (2008)       | Article       | Distance education; E-learning; Information and communication technologies; National open university of nigeria; Quality assurance                        | Turkish Online Journal of Educational Technology | 43       |
| 10 | Okada et al (2019)       | Article       | N/A                                                                                                                                                       | British Journal of Educational Technology        | 42       |

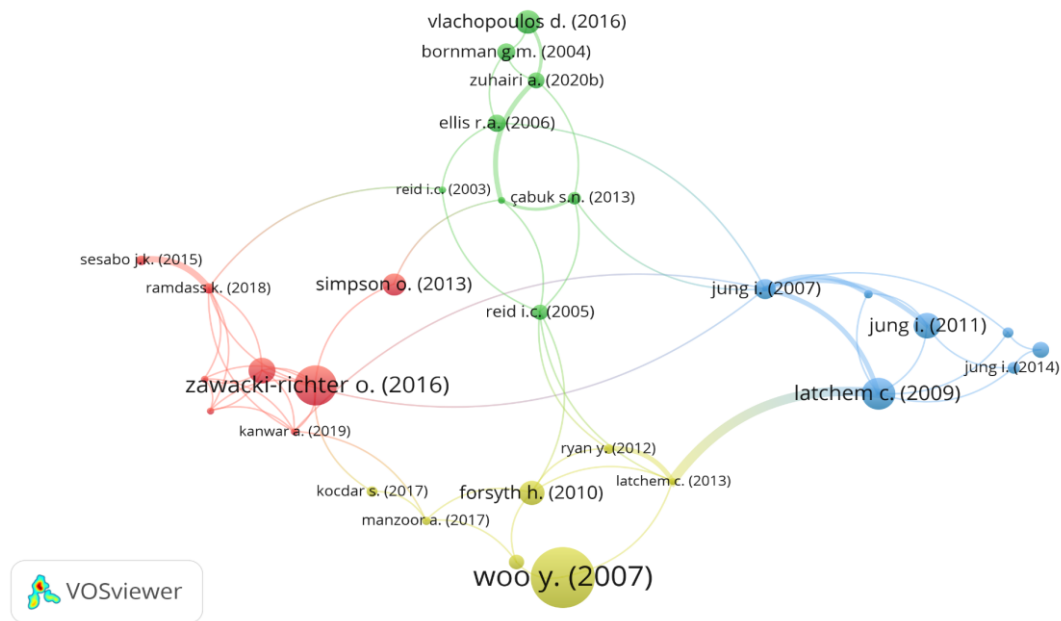
(N/A: Not available)

#### 4.5 Themes and Emerge Topics

The exploration of QADHE themes, using co-citation analysis, underscores the pivotal role of online distance education (ODE). Co-citation analysis operates on the principle that documents frequently cited together are likely to share thematic connections or address similar topics. The analysis reveals four major clusters (Figure 5) ranging from broad concepts to context-specific details.

The yellow cluster investigates the role of technology in enhancing the quality and accessibility of remote learning experiences. As technology continues to play an increasingly prominent role in education, understanding its impact on and how it would change the QA process in distance higher education becomes paramount (Uvalić-Trumbić and Martin, 2021). Meanwhile, the red cluster sheds light on methodologies and frameworks aimed at quality assurance in ODE, addressing the varied needs of learners and guaranteeing the effectiveness of these educational programs.

While these two clusters provide a theoretical contribution to the literature, the green cluster examines strategies and approaches adopted by various institutions. This topic is of particular importance as institutions strive to maintain quality and relevance in an increasingly digital and remote learning landscape (Stella and Gnanam, 2004), also providing examples and insights for other institutions to adopt. Meanwhile, blue cluster zooms in on the landscape of QA in ODL within the Asian context, considering regional perspectives, challenges, and advancements. Asia has experienced significant growth in DHE programs and institutions in recent years (Tait, 2018).



**Figure 5: Science mapping of bibliographic coupling documents in knowledge base of QADHE from 1993 to 2024 (30 documents, each had at least a citation)**

The examination of co-citation patterns reveals important themes, with frequently encountered keywords providing insights into notable topics, especially those that are emerging. Additionally, there is growing interest in newer themes linked to "open" concepts. Table 6 outlines keywords according to their frequency, while Figure 7 visually represents the changing trends and occurrences of these keywords over time, with colors representing their chronological progression.

It's evident that key keywords in this field include "quality assurance," "distance education," and "distance learning," reflecting their longevity and widespread use in the discourse on QADHE. Early keywords such as "assessment," "evaluation," "teaching," "globalization," "administration," and "curriculum development" indicate that the early discourse on QADHE mirrored discussions around traditional programs (Jung and Latchem, 2012). One of the main concerns during this period was whether DHE should be evaluated similarly to traditional programs, and whether their quality was comparable (Jung, 2022). Consequently, the topics prevalent during this stage remain rooted in conventional themes.

Subsequent keywords such as "online learning," "MOOCs," "quality management," and "total quality management" indicate a shift in research dynamics, driven by the rise and proliferation of the internet. While online learning and MOOCs have democratized access to DHE, they have also posed significant challenges to the QA process (Uvalić-Trumbić and Martin, 2021). This may necessitate the introduction of innovative management theories tailored to address emerging complexities.

Recently, keywords highlighting "open" principles, such as "open university," "open education," and "open and distance education," have gained significance, indicating a shift towards more inclusive educational paradigms. Notably, Open Education Resources (OER) are receiving increasing recognition in academic circles, despite being mentioned earlier in the discourse. OER, along with other aspects of "open" education, is a key driver of DHE advancement (UNESCO and Commonwealth of Learning, 2015), alongside digital innovation. However, this advancement also presents notable challenges (Uvalić-Trumbić and Martin, 2021), and underscores the need for further research to understand the frameworks and mechanisms of this new dimension of QADHE.

Table 6: Top 20 keywords in knowledge base of QADHE from 1993 to 2024

| No | Keyword                     | Frequency | No | Keyword                                    | Frequency |
|----|-----------------------------|-----------|----|--------------------------------------------|-----------|
| 1  | quality assurance           | 52        | 11 | assessment                                 | 6         |
| 2  | distance education          | 34        | 12 | evaluation                                 | 4         |
| 3  | distance learning           | 24        | 13 | open learning                              | 3         |
| 4  | higher education            | 19        | 14 | accreditation                              | 3         |
| 5  | open and distance learning  | 17        | 15 | content analysis                           | 3         |
| 6  | e-learning                  | 16        | 16 | course design                              | 3         |
| 7  | online learning             | 16        | 17 | globalisation                              | 3         |
| 8  | quality                     | 10        | 18 | information and communication technologies | 3         |
| 9  | open and distance education | 8         | 19 | learner support                            | 3         |
| 10 | open university             | 7         | 20 | moocs                                      | 3         |

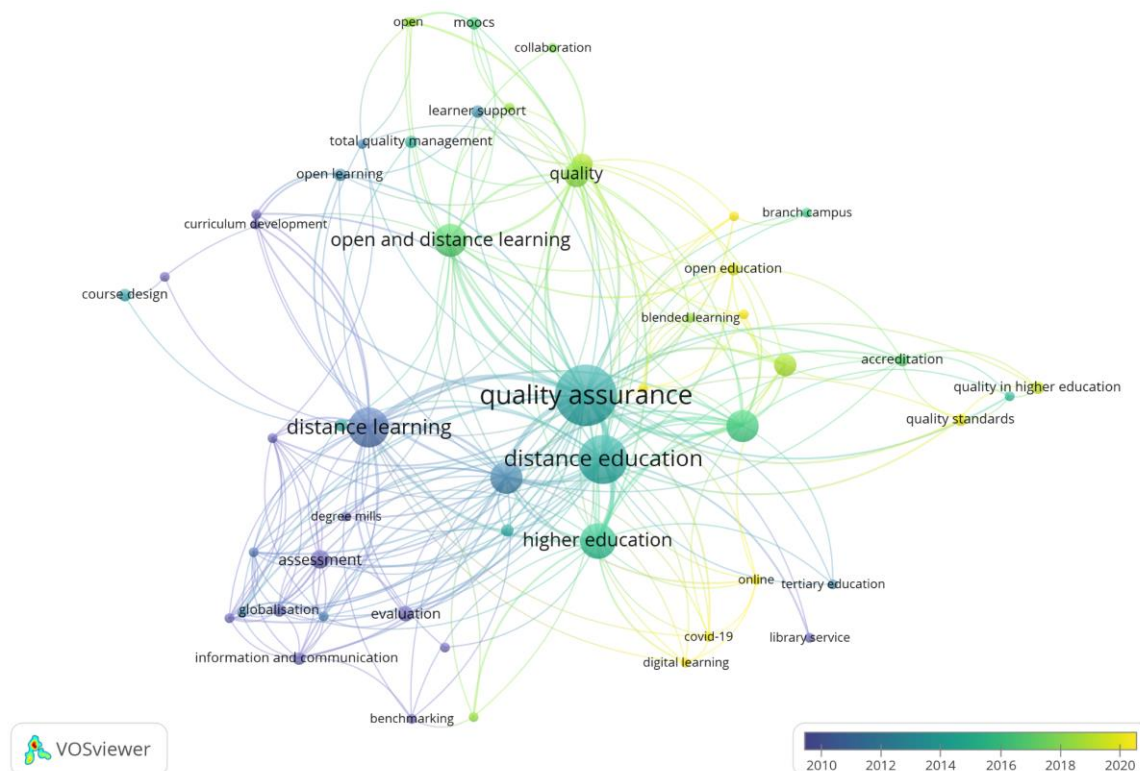


Figure 6: Science mapping of co-occurrence keywords in the knowledge base of QADHE from 1993 to 2024

## 5. Conclusions

In conclusion, this bibliometric study investigates the landscape of QADHE through an analysis of Scopus-indexed publications from 1993 to 2024. The study identified publication trends, international collaboration networks, influential authors, primary sources, thematic focuses, and emerging topics within the field.

The research trend in QADHE has progressed through three phases: an initial stage with few publications in the 1990s, a rapid expansion in the early 2000s with the rise of the internet, and stabilization in the late 2010s, indicating ongoing interest despite fluctuations, particularly in the mid-2000s. While QADHE shares developmental similarities with related fields, it has a lower publication volume, highlighting a significant demand for research in this area.

International collaboration in QADHE is robust, with significant contributions from Anglophone countries and South Africa, forming extensive research networks that highlight the subject's global importance. The emergence of Asian countries like Japan, India, and Malaysia as notable contributors underscores the growing diversity within the field. However, while institutions from English-speaking countries dominate citation counts, there is a strong encouragement for increased contributions from countries in Southeast Asia, Central Asia, and South Africa to further enrich the theoretical frameworks and research diversity within the field of QADHE.

Authors from diverse regions are advancing the discourse, but institutions in English-speaking countries lead in publication counts and citations. Significant contributions also come from International Christian University in Japan and the University of South Africa, highlighting the global impact of QADHE research. The increase in researchers since 2016 reflects a growing interest, particularly due to the shift to online learning. As the field evolves, there is ample opportunity for authors from various backgrounds to engage in the QADHE discussion.

The analysis of primary sources in QADHE indicates a need for a dedicated journal on the topic, complemented by influential edited books that enhance scholarship. Key documents offer insights into theoretical frameworks and practical examples, showcasing both general and specific perspectives. While Western scholars dominate citations, the global impact of QADHE research highlights the importance of diverse viewpoints. This trend further underscores the need for underrepresented regions to contribute to the literature, fostering a more inclusive understanding of QADHE.

The analysis of themes and keywords provides a comprehensive view of the diverse landscape of QADHE research, identifying significant clusters of thematic focus. The study highlights key clusters such as the integration of technology, methodologies, implementation strategies, and the Asian context. Keyword trends reveal the evolving discourse, shifting from traditional assessments to the emergence of "open" principles. Continued research is essential to address the complexities and challenges of online and open education, ensuring the quality and effectiveness of DHE programs globally.

However, several limitations should be considered when interpreting the findings. Firstly, reliance on Scopus-indexed publications may bias the results towards English-language sources, excluding relevant research from non-indexed databases or in non-English languages. Secondly, the study's temporal scope may overlook recent developments in QADHE amid rapid technological advancements. Additionally, while bibliometric analysis provides quantitative insights, it may not fully capture the qualitative nuances of research quality and content. Acknowledging these constraints and embracing future research suggestions can deepen the understanding of QADHE and enhance global educational practices.

## Acknowledgement

This work is the output of the research project no B2024-VKG.06, funded by the Ministry of Education and Training of Vietnam. The authors sincerely thank the Ministry for this support.

**Ethics Approval and AI Usage Declaration:** This research did not involve human participants, animal subjects, or sensitive data collection, and as such, ethics approval was not required. We confirm that all aspects of this study comply with relevant ethical guidelines. Additionally, we affirm that no artificial intelligence tools were used in the drafting, writing, or editing of this manuscript. The content presented herein is solely the result of the authors' independent effort and expertise.

## References

Amoozegar, A., Khodabandelou, R., Ebrahim, N.A., 2018. Major Trends In Distance Education Research: A Combination Of Bibliometric And Thematic Analyze.

- Andrade, M.S., Miller, R.M., Kunz, M.B., Ratliff, J.M., 2020. Online learning in schools of business: The impact of quality assurance measures. *J. Educ. Bus.* 95, 37–44. <https://doi.org/10.1080/08832323.2019.1596871>
- Baturay, M.H., 2015. An Overview of the World of MOOCs. *Procedia - Soc. Behav. Sci.*, International Conference on New Horizons in Education, INTE 2014, 25-27 June 2014, Paris, France 174, 427–433. <https://doi.org/10.1016/j.sbspro.2015.01.685>
- Belawati, T., Zuhairi, A., 2007. The Practice of a Quality Assurance System in Open and Distance Learning: A case study at Universitas Terbuka Indonesia (The Indonesia Open University). *Int. Rev. Res. Open Distrib. Learn.* 8. <https://doi.org/10.19173/irrodl.v8i1.340>
- Berge, Z.L., Mrozowski, S., 2001. Review of research in distance education, 1990 to 1999. *Am. J. Distance Educ.* 15, 5–19. <https://doi.org/10.1080/08923640109527090>
- Booker, J.M., Schluter, J.A., Carrillo, K., McGrath, J., 2011. Quality Improvement Initiative in School-Based Health Centers Across New Mexico. *J. Sch. Health* 81, 42–48. <https://doi.org/10.1111/j.1746-1561.2010.00556.x>
- Bornman, G.M., 2004. Programme review guidelines for quality assurance in higher education: A South African perspective. *Int. J. Sustain. High. Educ.* 5, 372–383. <https://doi.org/10.1108/14676370410561072>
- Brown, M., Duarte, J.M., 2024. Exploring gaps in the quality assurance of micro-credentials: a global mapping review of current practices. *J. Open Distance Digit. Educ.* 1, 1–16. <https://doi.org/10.25619/2bwhvw68>
- Brown, Y.R., Mark, 2011. *Quality Assurance Policies and Guidelines for Distance Education in Australia and New Zealand*, in: *Quality Assurance and Accreditation in Distance Education and E-Learning*. Routledge.
- Bukhari, S.M. ul H., Shah, N.H., Arif, M.H., 2021. Contemporary Trends in Quality Assurance in Distance Education. *Int. J. Distance Educ. E-Learn.* 6, 100–108. <https://doi.org/10.36261/ijde.v6i2.1865>
- Crisanto, M.A.L., 2018. Group reporting as a tool to enhance the quality of courses: The response of database students to online cooperative learning. *Asian Assoc. Open Univ. J.* 13, 73–87. <https://doi.org/10.1108/AAOUJ-01-2018-0006>
- Darojat, O., 2018. How Are the Results of Quality Assurance Programs Used to Inform Practices at A Distance Higher Education? *Turk. Online J. Distance Educ.* 19, 75–88. <https://doi.org/10.17718/tojde.382730>
- Darojat, O., Nilson, M., Kaufman, D., 2015. Perspectives on quality and quality assurance in learner support areas at three Southeast Asian open universities. *Distance Educ.* 36, 383–399. <https://doi.org/10.1080/01587919.2015.1081734>
- Falagas, M.E., Pitsouni, E.I., Malietzis, G.A., Pappas, G., 2008. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J.* 22, 338–342. <https://doi.org/10.1096/fj.07-9492LSF>
- Ferrer, J., Ringer, A., Saville, K., A Parris, M., Kashi, K., 2022. Students' motivation and engagement in higher education: the importance of attitude to online learning. *High. Educ.* 83, 317–338. <https://doi.org/10.1007/s10734-020-00657-5>
- Gaftandzhieva, S., Doneva, R., Jagatheesaperumal, S.K., 2023. Approaches and Tools for Quality Assurance in Distance Learning: State-of-play. *Int. J. Adv. Comput. Sci. Appl. IJACSA* 14. <https://doi.org/10.14569/IJACSA.2023.0140827>
- Garrett, R., 2020. Quality & Borderless Higher Education, Trends and Reflections.
- Hannache-Heurteloup, N., Moustaghfir, K., 2020. Exploring the barriers to e-learning adoption in higher education: a roadmap for successful implementation. *Int. J. Manag. Educ.* 14, 159–182. <https://doi.org/10.1504/IJMIE.2020.105407>
- Hebebcı, M.T., 2021. The Bibliometric Analysis of Studies on Distance Education. *Int. J. Technol. Educ.* 4, 796–817. <https://doi.org/10.46328/ijte.199>
- Jung, I., 2022. Quality Assurance in Online, Open, and Distance Education, in: *Handbook of Open, Distance and Digital Education*. Springer, Singapore, pp. 1–16. [https://doi.org/10.1007/978-981-19-0351-9\\_39-1](https://doi.org/10.1007/978-981-19-0351-9_39-1)
- Jung, I., Latchem, C., 2012. *Quality Assurance and Accreditation in Distance Education and e-Learning*, 0 ed. Routledge. <https://doi.org/10.4324/9780203834497>
- Keržič, D., Alex, J.K., Alvarado, R.P.B., Bezerra, D. da S., Cheraghi, M., Dobrowolska, B., Fagbamigbe, A.F., Faris, M.E., França, T., González-Fernández, B., González-Robledo, L.M., Inasius, F., Kar, S.K., Lazányi, K., Lazăr, F., Machin-Mastromatteo, J.D., Marôco, J., Marques, B.P., Mejía-Rodríguez, O., Prado, S.M.M., Mishra, A., Mollica, C., Jiménez, S.G.N., Obadić, A., Raccanello, D., Rashid, M.M.U., Ravšelj, D., Tomaževič, N., Uleanya, C., Umek, L., Vicentini, G., Yorulmaz, Ö., Zamfir, A.-M., Aristovnik, A., 2021. Academic student satisfaction and perceived performance in the e-learning environment during the COVID-19 pandemic: Evidence across ten countries. *PLOS ONE* 16, e0258807. <https://doi.org/10.1371/journal.pone.0258807>
- Khuram, S., Rehman, Ch.A., Nasir, N., Elahi, N.S., 2023. A bibliometric analysis of quality assurance in higher education institutions: Implications for assessing university's societal impact. *Eval. Program Plann.* 99, 102319. <https://doi.org/10.1016/j.evalprogplan.2023.102319>
- Kocdar, S., Okur, M.R., Bozkurt, A., 2017. An Examination of XMOOCs: An Embedded Single Case Study Based on Conole's 12 Dimensions. *Turk. Online J. Distance Educ.* 18, 52–65. <https://doi.org/10.17718/tojde.340381>
- Latchem, C., Jung, I., 2009. *Distance and Blended Learning in Asia*. Routledge, New York. <https://doi.org/10.4324/9780203878774>
- Littlefield, M.B., Rubinstein, K., Laveist, C.B., 2019. Designing for Quality: Distance Education Rubrics for Online MSW Programs. *J. Teach. Soc. Work* 39, 489–504. <https://doi.org/10.1080/08841233.2019.1658691>
- Lockhart, M., Lacy, K., 2002. An assessment model and methods for evaluating distance education programmes. *Perspect. Policy Pract. High. Educ.* <https://doi.org/10.1080/136031002320634998>
- Lowenthal, P., Hodges, C., 2015. In Search of Quality: Using Quality Matters to Analyze the Quality of Massive, Open, Online Courses (MOOCs). *Int. Rev. Res. Open Distrib. Learn.* 16. <https://doi.org/10.19173/irrodl.v16i5.2348>

- Martin, F., Sun, T., Westine, C.D., 2020. A systematic review of research on online teaching and learning from 2009 to 2018. *Comput. Educ.* 159, 104009. <https://doi.org/10.1016/j.compedu.2020.104009>
- Matlakala, M.C., Maritz, J.E., 2019. Curriculum Development at Institutional Level: Reflections and Lessons Learnt. *Afr. J. Nurs. Midwifery* 21, 11 pages-11 pages. <https://doi.org/10.25159/2520-5293/4781>
- Mekky, S.T., 2021. On-Line Learning Assessment: Egyptian Universities as a Case Study. *Int. J. Inf. Educ. Technol.* 11, 471–478. <https://doi.org/10.18178/ijiet.2021.11.10.1552>
- Mkwizu, K.H., Junio-Sabio, C., 2024. Quality Assurance implementation and application in Distance Education. *Asian J. Distance Educ.* 19.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2010. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Int. J. Surg.* 8, 336–341. <https://doi.org/10.1016/j.ijsu.2010.02.007>
- Moore, J.L., Dickson-Deane, C., Galyen, K., 2011. e-Learning, online learning, and distance learning environments: Are they the same? *Internet High. Educ., Web mining and higher education: Introduction to the special issue* 14, 129–135. <https://doi.org/10.1016/j.iheduc.2010.10.001>
- Mystakidis, S., Berki, E., Valtanen, J., 2019. The Patras Blended Strategy Model for Deep and Meaningful Learning in Quality Life-Long Distance Education. *Electron. J. E-Learn.* 17, pp66-78-pp66-78. <https://doi.org/10.34190/JEL.17.2.01>
- Naidoo, R., Kemlall, R., 2019. A study of the teaching and learning of a QA course in an engineering department at an ODeI institution. *J. Educ. Stud.* 18, 1–11. <https://doi.org/10.10520/EJC-1d87535ec2>
- Ndibalema, P., 2022. The Global Research Trends on the Growth of Remote Learning in Higher Education Institutions: A Bibliometric Analysis. *Int. J. Technol. Educ. Sci.* 6, 218–236. <https://doi.org/10.46328/ijtes.332>
- Nguyen, C.H., Ta, H.T.T., Nguyen, T.-T., 2022. Quality Assurance of Distance Education in Vietnamese Higher Education, in: Tran, T., Nguyen, C.H., Nguyen, L.T.M. (Eds.), *Educational Innovation in Vietnam: Opportunities and Challenges of the Fourth Industrial Revolution*. Routledge, London. <https://doi.org/10.4324/9781003202424>
- Ossiannilsson, E., Williams, K., Brown, M., 2015. Quality models in online and open education around the globe: State of the art and recommendations. *DISTANCE Educ.*
- Pham, H.-H., Ho, T.-T.-H., 2020. Toward a ‘new normal’ with e-learning in Vietnamese higher education during the post COVID-19 pandemic. *High. Educ. Res. Dev.* 39, 1327–1331. <https://doi.org/10.1080/07294360.2020.1823945>
- Prinsloo, P., 2019. South Africa, in: Qayyum, A. (Ed.), *Open and Distance Education in Asia, Africa and the Middle East: National Perspectives in a Digital Age*, SpringerBriefs in Education. Springer Singapore, Singapore. <https://doi.org/10.1007/978-981-13-5787-9>
- Pritchard, A., 1969. Statistical Bibliography or Bibliometrics. *J. Doc.* 25, 348–349.
- Qayyum, A., Zawacki-Richter, O., 2019. The State of Open and Distance Education, in: *Open and Distance Education in Asia, Africa and the Middle East: National Perspectives in a Digital Age*, SpringerBriefs in Education. Springer Singapore, Singapore. <https://doi.org/10.1007/978-981-13-5787-9>
- Rodríguez Jiménez, C., Sanz Prieto, M., Alonso García, S., 2019. Technology and Higher Education: A Bibliometric Analysis. *Educ. Sci.* 9, 169. <https://doi.org/10.3390/educsci9030169>
- Rovai, A.P., 2003. A practical framework for evaluating online distance education programs. *Internet High. Educ.* 6, 109–124. [https://doi.org/10.1016/S1096-7516\(03\)00019-8](https://doi.org/10.1016/S1096-7516(03)00019-8)
- Rovai, A.P., Downey, J.R., 2010. Why some distance education programs fail while others succeed in a global environment. *Internet High. Educ.* 13, 141–147. <https://doi.org/10.1016/j.iheduc.2009.07.001>
- Sarıtaş, M.T., Börekci, C., Demirel, S., 2022. Quality Assurance in Distance Education through Data Mining. *Int. J. Technol. Educ. Sci.* 6, 443–457. <https://doi.org/10.46328/ijtes.396>
- Scull, W.R., Kendrick, D., Shearer, R., Offerman, D., 2011. The Landscape of Quality Assurance in Distance Education 75.
- Shen, C., Ho, J., 2020. Technology-enhanced learning in higher education: A bibliometric analysis with latent semantic approach. *Comput. Hum. Behav.* 104, 106177. <https://doi.org/10.1016/j.chb.2019.106177>
- Stella, A., Gnanam, A., 2004. Quality assurance in distance education: The challenges to be addressed. *High. Educ.* 47, 143–160. <https://doi.org/10.1023/B:HIGH.0000016420.17251.5c>
- Tait, A., 2018. Open Universities: the next phase. *Asian Assoc. Open Univ. J.* 13, 13–23. <https://doi.org/10.1108/AAOUJ-12-2017-0040>
- UNESCO, Commonwealth of Learning, 2015. A Basic guide to open educational resources (OER). UNESCO Digital Library.
- UNISA, n.d. About [WWW Document]. About. URL <https://www.unisa.ac.za/sites/corporate/default/About> (accessed 2.25.24).
- Uvalić-Trumbić, S., Martin, M., 2021. A new generation of external quality assurance dynamics of change and innovative approaches: new trends in higher education. UNESCO, Paris.
- Vrasidas, C., 2003. The Design, Development, and Implementation of LUDA Virtual High School. *Comput. Sch.* [https://doi.org/10.1300/J025v20n03\\_03](https://doi.org/10.1300/J025v20n03_03)
- Woo, Y., Reeves, T.C., 2007. Meaningful interaction in web-based learning: A social constructivist interpretation. *Internet High. Educ., Special Section of the AERA Education and World Wide Web special Interest Group (EdWeb/SIG)* 10, 15–25. <https://doi.org/10.1016/j.iheduc.2006.10.005>
- Wright, C.R., Dhanarajan, G., Reju, S.A., 2009. Recurring Issues Encountered by Distance Educators in Developing and Emerging Nations. *Int. Rev. Res. Open Distrib. Learn.* 10. <https://doi.org/10.19173/irrodl.v10i1.608>
- Wyk, M.M. van, 2018. Blog Phenomenology: Student Teachers’ Views of Learning to Teach Economics. *Int. J. Web-Based Learn. Teach. Technol. IJWLTT* 13, 62–77. <https://doi.org/10.4018/IJWLTT.2018040105>

- Xiao, C., Qiu, H., Cheng, S.M., 2019. Challenges and opportunities for effective assessments within a quality assurance framework for MOOCs. *J. Hosp. Leis. Sport Tour. Educ.* 24, 1–16. <https://doi.org/10.1016/j.jhlste.2018.10.005>
- Zawacki-Richter, O., Anderson, T., 2014. *Online Distance Education: Towards a Research Agenda*. Athabasca University Press. <https://doi.org/10.15215/aupress/9781927356623.01>
- Zawacki-Richter, O., Baecker, E.M., Vogt, S., 2009. Review of distance education research (2000 to 2008): Analysis of research areas, methods, and authorship patterns. *Int. Rev. Res. Open Distrib. Learn.* 10, 21–50. <https://doi.org/10.19173/irrodl.v10i6.741>
- Zawada, B., 2019. From functional quality apparatus to meaningful enactment: UNISA as example. *Qual. Assur. Educ.* 27, 384–400. <https://doi.org/10.1108/QAE-03-2019-0031>
- Zhang, W., Li, W., 2019. Transformation From RTVUs to Open Universities in China: Current State and Challenges. *Int. Rev. Res. Open Distrib. Learn.* 20. <https://doi.org/10.19173/irrodl.v20i4.4076>
- Zhao, Y., Watterston, J., 2021. The changes we need: Education post COVID-19. *J. Educ. Change* 22, 3–12. <https://doi.org/10.1007/s10833-021-09417-3>
- Zuhairi, A., Raymundo, M.R.D.R., Mir, K., 2020. Implementing quality assurance system for open and distance learning in three Asian open universities: Philippines, Indonesia and Pakistan. *Asian Assoc. Open Univ. J.* 15, 297–320. <https://doi.org/10.1108/AAOUJ-05-2020-0034>

# From Twitch to YouTube Live: A Systematic Literature Review of Streaming in Higher Education

Jorge Oceja<sup>1</sup> and Carmen Álvarez-Álvarez<sup>2</sup>

<sup>1</sup>Universitat de les Illes Balears, Spain

<sup>2</sup>Universidad de Cantabria, Spain

[jorge.oceja@uib.es](mailto:jorge.oceja@uib.es) (corresponding author)

[carmen.alvarez@unican.es](mailto:carmen.alvarez@unican.es)

<https://doi.org/10.34190/ejel.23.1.3702>

An open access article under [CC Attribution 4.0](#)

**Abstract:** Although the term streaming is polysemic, nowadays it is generally understood as real time communication between broadcasters of information (or streamers) and their community, through popular platforms such as Twitch or YouTube Live. As there is a lack of knowledge about the educational impact of this kind of streaming on higher education, the main objective is to explore how educational experiences have been documented in scientific literature. The specific research questions are (1) How may streaming educational practices in the university context appear in scientific literature, (2) What types of scientific production (conference proceedings, book chapters, scientific articles, etc.) have documented these experiences? (3) What elements of this culture (specific platforms, communication codes, presence of informal elements, etc.) are present in these practices allowing us to consider them streaming experiences? (4) To what extent have these projects been evaluated and what are their educational results? The study is grounded in theoretical perspectives that challenge the existing gap between educational research and impactful learning practices in both face-to-face and eLearning environments. It explores whether, as we hypothesize, there is a lag between the consolidation of new media in popular culture and various disciplines, and its integration into educational contexts. Following PRISMA declaration guidelines, a systematic literature review (SLR) was carried between June 1st and June 9th, 2024 in English language both in Web of Science and Scopus. Starting with a wide collection of results (Scopus n = 426 and Web of Science n = 354) several exclusion criteria were progressively applied before reaching the final sample. Results show that these experiences have been rarely documented (n = 10) with most of them appearing either on journals or conference proceedings. Most experiences still have an instructional bias, not considering the social and cultural dimension of this phenomenon. Although results seem promising, assessment strategies can be improved. Based on these results, researchers emphasize the need to adapt teaching methodologies to leverage the affordances and cultural dynamics of streaming platforms, providing insights for future research and pedagogical innovation. Additionally, they advocate for enhancing teachers' media literacy to better navigate and utilize the distinct features and cultural codes of this medium effectively.

**Keywords:** Higher education, Streaming, Twitch, Youtube, Facebook, Systematic literature review

## 1. Introduction

Streaming has grown as a dominant mode of digital media consumption, impacting not only entertainment but also general educational practices (Steinbeck et al., 2021) and, more specifically higher education (Pozo-Sánchez et al., 2021).

The Merriam-Webster dictionary defines streaming as “the transfer of data (such as audio or video material) in a continuous stream for immediate processing or playback” (Merriam-Webster, 2022). The term, due to its polysemic nature, is used in different ways. The most common is the broadcasting of audiovisual content from servers to a final user, who stores it to reproduce it almost immediately, without completely downloading this content to a device. In this sense, it is common to speak about the well-known on-demand audiovisual platforms (Netflix, Spotify, etc.) as streaming services.

However, given that technology allows the continuous flow of data, this *broadcasting* is often associated with the idea of *live-streaming*. In fact, in English language, there is a certain terminological confusion as, on many occasions, we find an indiscriminate use of the concepts *streaming* (which should be exclusively associated with the idea of broadcasting), and *live-streaming*, which explicitly indicates that the broadcast is real-time, or live. For example, when performing a synchronous videoconference, it is common in the English-speaking context to use any of these two terms.

Also, in the last years, the terms *streaming* and *streamer* have acquired a new connotation in popular culture, due to the upsurge of platforms that allow users to broadcast live programs easily. This represents the dissolution of the barrier between consumers and content creators, which led Toffler (1980) to coin the term prosumer in the 1980s. Although some authors are skeptical with this apparent democratization of content

(Castañeda *et al.*, 2020), it is true that media, dominated for many years by television, radio, and written press, has been transformed, in part, due to this phenomenon.

This creative democratization affects text (the consolidation of blogs and wikis are an example), sound (podcasts), and audiovisual productions (rise of video repositories such as YouTube or Vimeo). However, in the last years, we have seen changes in the way we consume audiovisual content, in particular video. Although video on-demand (VoD) (the possibility of immediately accessing content stored in a repository) seemed to be the most practical option (particularly with free services like YouTube), streaming platforms such as Twitch are increasingly capturing more users. Some examples are the 23 billion hours of content consumed in 2021, the more than 9 million streamers who habitually generate content, or the almost 7 million concurrent viewers in the platform in June, 2022 (Twitch statistics & charts, 2018). These streaming platforms facilitate immediate interaction between streamers and viewers, fostering a sense of community and engagement that is less pronounced in VoD services (Wulf, Schneider & Beckert, 2020). Twitch also offers monetization avenues such as subscriptions, advertisements, and donations, attracting content creators who can engage with their audience in real-time.

This economic model encourages the production of diverse live content, contributing to the platform's growth.

Twitch growth and the consolidation of other live streaming platforms such as YouTube Live or Facebook Live was associated, originally, to videogames. While playing, streamers commented on games while they interacted with their audience through specific communication codes and language, which led some authors to talk about streaming culture (Burroughs & Rugg, 2014; Woodcock & Johnson, 2019) and even Twitch culture (Sjöblom *et al.*, 2019). The phenomena have progressively expanded to other disciplines such as literature ([https://www.twitch.tv/entre\\_tulipanes](https://www.twitch.tv/entre_tulipanes)), politics (<https://www.twitch.tv/nanisimo>)

or arts (<https://www.twitch.tv/sebaguidobono>).

### **1.1 Previous Research on Streaming and Education**

Although the incorporation of teachers to live streaming is slow, we can find some interesting education channels such as “Entre profes” (Among teachers) (<https://www.twitch.tv/entreprofes>), spaces that aim at working with games and gamification of classrooms (<https://www.twitch.tv/gamificaitor>), or channels oriented towards the creation of educational materials (<https://www.twitch.tv/paraprofes>).

Even though some authors (Catá, 2019) have detected the risk of disregard for specific collectives, most works have underlined the positive aspects of streaming and its educational possibilities (Pirker, Steinmaurer & Karakas, 2021). Among those, authors have mentioned the empowerment of communities, the interactivity between creators and spectators (Flores-Saviaga, *et al.*, 2019) or the ability to co-create contents between both parts (Sjöblom *et al.*, 2019). However, despite these benefits and the potential of using streaming platforms and their communication codes, no systematic exploration of educational projects on higher education has been conducted.

The penetration of streaming in society and the participation of students in this phenomenon, makes reasonable to explore the possibilities of the medium to broaden the conversations outside the classroom, generate dialogic learning, and ultimately, consolidate students and teachers personal learning environments (PLEs) through experiences based in informal and no-formal learning. Works by Salinas have confirmed that students PLEs, which in most cases include streaming channels, allow “to integrate the institutional virtual environment associated to formal learning, with a more informal environment” (Salinas, 2013, p.53).

### **1.2 Problem, Purpose and Research Questions**

Thus, the problem guiding this work is the lack of knowledge about the educational projects that use streaming in the university context. Particularly, we want to understand how these projects have been documented, and by extension, the way in which their implementation is assessed. The study is significant as it will confirm or refute the existence of a gap between academic production and cutting-edge pedagogical practices. Furthermore, it will serve to identify successful projects that appropriate streaming as a cultural phenomenon, enabling an analysis of its potential and limitations. Ultimately, it will assess the capacity of face-to-face and hybrid education to integrate effective communication practices.

Thus, the purpose of this work is to discover the impact of streaming on higher education according to scientific literature. The specific research questions are (1) How may streaming educational practices in the university context appear in scientific literature, (2) What types of scientific production (conference proceedings, book

chapters, scientific articles, etc.) have documented these experiences? (3) What elements of this culture (specific platforms, communication codes, presence of informal elements, etc.) are present in these practices allowing us to consider them streaming experiences? (4) To what extent have these projects been evaluated and what are their results?

At this point there has not been systematic reviews targeting this topic and the scarce attempts have focus on other educational stages (Garrigos et al., 2022).

## 2. Methodology

The study is grounded in theoretical perspectives (Miller, 1999; Reeves, 2000) that highlight the existing gap between educational research and impactful learning practices in both face-to-face and eLearning environments. These perspectives highlight the historical lack of applicability of some educational research, and on some occasions, the impossibility of generalizing results from a specific project to other contexts. Thus, it explores whether, as we hypothesize, there is a lag between the consolidation of this new media both in popular culture and diverse disciplines, and its integration into educational contexts.

To confirm this a systematic literature review (SLR) was carried. SLRs are characterized by a rigorous approach based on a specific methodology that allows other researchers to replicate the results (Purssell & McCrae, 2020). Jesson et al. (2011) mentions the phases of mapping, planning and establishing a protocol, documentation, definition of inclusion and exclusion criteria, search and filtering of data, review of the quality, extraction of data, and summary.

In the last few years, diverse initiatives have been created to consolidate these approaches, with PRISMA (Moher et al., 2009) being one of the most successful models. This proposal provides a guide to improve the quality of the reviews, facilitating a clear and common procedure. Following PRISMA declaration guidelines, our search strategy comprised the period between June 1st and June 9th, 2024.

The two researchers jointly reviewed the PRISMA documentation and conducted the searches independently to minimize errors following the next steps:

- Consensus on the search string: Both researchers agreed on the formulation of the search terms.
- Joint determination of inclusion and exclusion criteria: Criteria were collaboratively defined to ensure alignment.
- Independent execution of the search: Each researcher carried out the search separately, confirming the number of results obtained.
- Export of results to independent spreadsheets: The results were exported to two separate files for analysis.
- Autonomous application of inclusion and exclusion criteria: Each researcher applied the criteria independently at each phase.
- Meetings after each phase: After the phases of identification, screening, eligibility, and inclusion, meetings were held to resolve conflicts and disagreements.

The databases utilized were the core collections of the Web of Science (WoS, Clarivate Analytics) and Scopus. Although WoS is currently the most important corpus of scientific searches and academic information (Li, Rollins & Yan, 2018), some authors (Falagas, et al., 2008; Kulkarni, et al., 2009) have questioned its exclusive use, stating that other databases such as Scopus could be more comprehensive, as they include a larger number of sources.

The search string with its corresponding Boolean operators was the same in both cases, although due to the syntax peculiarities of each system, in the case of WOS, the field "TS" was used (equivalent to title, abstract, and keywords), while in Scopus we used the equivalent "TITLE-ABS-KEY". Being aware about the habitual use of real-time video broadcasts in the area of medicine, two operators were included to eliminate the results associated to this area: "medical" and "surgery".

The search, centered on articles written in English, used the following strings:

Scopus:

(( TITLE-ABS-KEY ( "live stream\*" ) OR TITLE-ABS-KEY ( "live-stream\*" ) OR TITLE-ABS-KEY ( twitch ) ) AND ( TITLE-ABS-KEY ( "higher education" ) OR TITLE-ABS-KEY ( university ) ) AND NOT ( TITLE-ABS-KEY ( surgery ) OR TITLE-ABS-KEY ( medical ) ))

Web of Science:

(TS="live stream\*" OR TS="live-stream\*" OR TS=twitch) AND (TS="higher education" OR TS="university") NOT (TS="medical" NOT TS="surgery")

We opted for a greater inclusivity in the first approach, with a subsequent application of other inclusion/exclusion criteria. Exclusion criteria focused on selecting articles directly related to the problem, purpose, and research questions. They were structured around aspects such as accessibility (e.g., fully published articles, excluding in-press works), a focus on educational issues (e.g., excluding disciplines others than education), and focus on the university context (e.g., excluding non-university settings).

Starting with the first results (Scopus n = 426) (WOS n = 354), articles written before 2000 were eliminated, given that the technology for live broadcasting and, more importantly, what is known today as streaming, did not exist then (Scopus n = 293) (WOS n = 311). After the elimination of redundancies, a total of 312 articles were obtained. Despite the elimination of medical results, the list kept included many articles associated to this area. After the elimination of 196 articles, exclusion criteria were applied to the resulting sample (n = 116). As the final objective was to find studies that documented education experiences based on streaming as a contemporary cultural phenomenon, either through didactic formats or the use of platforms associated to this medium, we applied 13 exclusion criteria: disciplines other than education (n = 27), exclusively technical papers (n = 18), conference proceedings introduction texts (n = 11), articles that could not be accessed (n = 4), non-university contexts (n = 3), sociological analysis (n = 2), streaming as leisure (n = 1), projects with videos exclusively consumed in a deferred manner (n = 1), documents about project funding (n = 1), non-educational streaming studies centered on videogames (n = 1), articles in press or not published (n = 1), other technologies besides streaming (n = 1), and university management (n = 2). This led us to total of 10 articles (n = 10), which were analyzed intensively to provide answers to the research questions. The following figure (Figure 1) provides a visual summary of the process undertaken to reach the final selection of works.

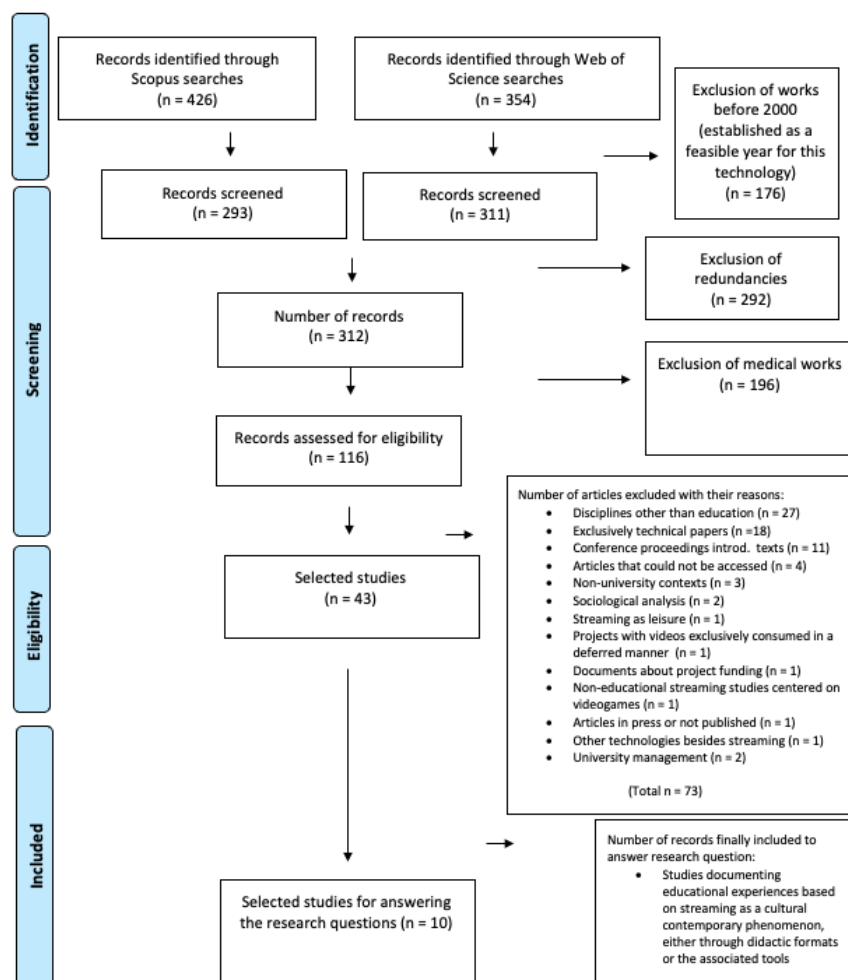


Figure 1: Graphical representation of the process followed in the PRISMA methodology leading to the final sample of works

### 3. Results

Below, the results for the four research questions formulated are presented. Different tables were created, showing and analyzing the different elements in each case.

#### *Research question 1. Documentation of experiences*

To answer the first question, a table was created summarizing projects documented in the scientific literature that have utilized streaming, understood as a contemporary cultural phenomenon, in educational practices in higher education (Table 1). The table shows the number of the article, the authors, the publication date, and its title.

**Table 1: Authors, date and title**

| N. | Authors                                 | Date | Title                                                                                                                                                                   |
|----|-----------------------------------------|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | Liu, I. F., Hung, H. C., & Liang, C. T. | 2023 | A study of programming learning perceptions and effectiveness under a blended learning model with live streaming: comparisons between full-time and working students    |
| 2  | Pirker J., Steinmaurer A., Karakas A.   | 2021 | Beyond Gaming: The Potential of Twitch for Online Learning and Teaching                                                                                                 |
| 3  | ChanLin L.J.                            | 2020 | Engaging university students in an ESL live broadcast                                                                                                                   |
| 4  | ChanLin L.J.                            | 2019 | ESL live broadcast teaching service in a university library                                                                                                             |
| 5  | Calatayud, L., Mireia, J., Monsalve, L. | 2023 | ICT in the University Environment: Twitch and Instagram as Learning Tools                                                                                               |
| 6  | Steinmaurer, A., Gütl, C.               | 2023 | Implementation and Experiences of a Flipped Lecture Hall – A Fully Online Introductory Programming Course                                                               |
| 7  | Aniroh, K; Hanum, L; Ariyanto, AAG      | 2018 | The Effectiveness of YouTube Live Streaming as Digital Learning Media in Tourism and Guiding Subject                                                                    |
| 8  | Wu C. C., Chao H.-W., Tsai C. W.        | 2021 | The effects of Facebook live-stream teaching on improving students' dance skills: Impacts on performance, learning motivation, and physical activity class satisfaction |
| 9  | Hertzog P.E.                            | 2018 | The use of live streaming in design-based modules for open distance learning                                                                                            |
| 10 | Wang S., Pradhan S., Cousins K.         | 2021 | Toward a game-based dialogical pedagogy: Insights from massively multiplayer online role-playing games                                                                  |

In first place, and related to the first research question, it is striking the scarcity of projects, as only 10 of them were aligned with the idea of streaming present in this work. Also, articles 3 and 4, are based on the same educational experience which is divided into two different branches.

Most articles were based on other definitions of the term (see the introduction section), focusing mainly on technology for broadcasting or live classes under classical instruction approaches. This helped us to confirmed that, although videoconference is a common technology in education, the use of tools, communication codes, and language associated to today's streaming culture has been scarcely explored in higher education.

#### *Research question 2. Type of scientific production*

To provide an answer to the second research question, an analysis was performed to determine what kind of scientific production was used in each case. The following table shows, in different columns, the format of this production, the name of the specific journal of medium, some of their peculiarities, as well as the databases in which they appeared (Table 2).

**Table 2: Sources in which the selected projects appeared: type of production, medium, characteristics, and databases**

| N. | Type    | Medium                                                                                           | Character                                                                                                                                                                                                                                                  | DB             |
|----|---------|--------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| 1  | Article | Interactive Learning Environments                                                                | Journal focusing on the design and use of interactive learning environments, focusing on knowledge sharing, adaptive systems, pedagogy                                                                                                                     | WOS and Scopus |
| 2  | Conf.   | Annual Conference on Innovation and Technology in Computer Science Education, ITICSE             | Conference on innovation, technology, and computer science organized since 2016. It takes place in different European and Latin American countries. Organized by SIGCSE, a prestigious group interested in education and computer science.                 | WOS and Scopus |
| 3  | Article | Electronic Library                                                                               | Journal in the area of information management and the ICT. Since 1983, it is centered on matters such as the content creation, its organization, and access to it.                                                                                         | WOS and Scopus |
| 4  | Conf.   | Asia Pacific Information Technology Conference                                                   | Benchmark reference in the Pacific-Asian area about ICT. The works are published in the ACM International Conference Proceeding Series, indexed in Scopus.                                                                                                 | WOS and Scopus |
| 5  | Article | Revista Iberoamericana de Tecnologías del Aprendizaje                                            | Journal of the IEEE Education Society focused on instructional technology in the Ibero-American context                                                                                                                                                    | WOS and Scopus |
| 6  | Conf.   | Proceedings of the 25th International Conference on Interactive Collaborative Learning (ICL2022) | An interdisciplinary conference aims to focus on the exchange of trends, results and practical experience in Interactive Collaborative Learning and Engineering Pedagogy.                                                                                  | Scopus         |
| 7  | Article | Arab World English Journal                                                                       | Journal in the area of Language and linguistics. It is in English, but also directed to the Arab world. Created in 2010, it also includes articles related to communication and education.                                                                 | WOS            |
| 8  | Article | International Journal of Mobile and Blended Learning                                             | Journal in the area of education created in 2009. Very focused on trends in the area of e-learning and m-learning.                                                                                                                                         | WOS and Scopus |
| 9  | Article | World Transactions on Engineering and Technology Education                                       | Journal in the area of engineering and education technology, published by WIETE. It was originally associated to the International Center for Engineering Education of the UNESCO. Its aim is to be a benchmark on teaching of engineering and technology. | Scopus         |
| 10 | Conf.   | Proceedings of the European Conference on Games-based Learning                                   | The most important event in Europe about gamification and game-based learning with a history of almost 20 years. Aside from the academic part, it organized annual contests to award educational games.                                                    | WOS and Scopus |

As for the second research question, sources were of different types, including articles (the format considered to be the most prestigious within scientific literature) (n = 6), conference proceedings, a more agile, brief format, associated with participation in these events (n = 4).

### *Research question 3. Elements present in the documents*

The third research question aimed to understand what characteristic elements of the streaming culture were present in each experience. The table below summarizes, in different columns, the following: specific aspect that allowed us to consider them streaming experiences, where they were found within the formal-informal continuum, and lastly, in what academic context were they are developed (Table 3).

**Table 3: Today's streaming culture elements present in the experiences, place within the formal-informal continuum, and context within which they are developed**

| N. | Reason for inclusion                                                                                                                                  | Formal-informal continuum                                                                                                                                            | Context and dynamics                                                                                                                                                                                                                                                                   |
|----|-------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | - Streaming platforms (Facebook live)                                                                                                                 | Formal but taking into consideration students that could not attend classes in a regular way when building the model and scheduling the broadcasts.                  | Teaching of coding to students from different degrees in a Taiwanese university combining f2f teaching, asynchronous learning and live streaming.                                                                                                                                      |
| 2  | - Streaming platforms (Twitch and Discord)<br>- Levels of interactivity and participation characteristic of streaming                                 | Formal environment with respect to schedules and classes, although the sessions are open to the community.                                                           | Training on programming and game design in Master's. Very interactive classes before the pandemic, so Twitch was used to try to maintain the dynamics, in combination with active methodologies: flipped classroom, Q&A, moderated debates, etc.                                       |
| 3  | - Streaming platforms and tools (OBS, YouTube Live and Facebook Live)<br>- Anyone can be a streamer                                                   | Some sessions take place in regulated schedules, beyond the formal to create learning communities.                                                                   | Creation of a learning community to improve the use of English of university students. Its peculiarity was that students could act as streamers.                                                                                                                                       |
| 4  | - Streaming platforms and tools (OBS, YouTube Live and Facebook Live)                                                                                 | Formal with respect to broadcasting in specific schedules, but re-utilization of materials within the informal learning community manages by the university library. | Re-interpretation of the library as a non-intrusive learning space. Videos are added to the social channels broadcasted in streaming mode that are hosted in their servers. The intention is for the community to be autonomous and that all the users improve their level of English. |
| 5  | - Streaming platforms (Twitch)<br>- Used of Twitch connected with other instagram activities                                                          | Formal but also implemented due to COVID-19                                                                                                                          | Subject "History of the School" taught in the Primary Education Teaching Degree at the University of Valencia. Streaming used to share the results of a case study implemented through role-playing in Instagram.                                                                      |
| 6  | - Streaming platforms (Twitch)<br>-Focus on improving the community interaction                                                                       | Formal but also implemented due to COVID-19 to setup a flipped classroom experience                                                                                  | Course Introduction to Programming within a computer science-degree at University of Graz.                                                                                                                                                                                             |
| 7  | - Streaming platforms (YouTube Live)                                                                                                                  | Formal. Use of YouTube as an alternative to other more academic tools.                                                                                               | University courses of Tourism and Orientation II in the philology degree. The only element imported is the platform, but the development of the classes is regulated.                                                                                                                  |
| 8  | - Streaming platforms (Facebook Live)<br>- An important part of the experience is developed through mobile devices                                    | Formal (university context) is combined with the non-formal, as it is an optional course worth 0 credits.                                                            | Training in the culture of physical education and dance for 1st to 4th-year students Although the synchronous sessions were regulated, the rest of the Facebook elements are utilized to create a sense of community.                                                                  |
| 9  | - Streaming platforms (YouTube Live)<br>- Dynamic interactions through the chat are an important part                                                 | Formal environment, although within an eminently practical course. Professionals from the sector are brought in.                                                     | Training on the Industrial Projects IV in electrical and mine engineering. Professionals in the sector (in charge of the course) act as streamers before handing in the projects.                                                                                                      |
| 10 | - Streaming platforms (Twitch)<br>- Cultural processes and codes of streaming as it is an original experience that takes place in the platform itself | Informal setting based on dialogic learning and the acquisition of competences through game-based learning.                                                          | Analysis of the perception of a group of players who are at the same time streamers, on the education competences acquired when using Twitch to train other players.                                                                                                                   |

The above-mentioned scarcity of experiences had not to do with the exigency of the criteria, as projects including any element of the streaming culture were considered, even if this was just the use most-common platforms. Among those, Twitch (n = 4), YouTube Live (n = 4), or Facebook Live (n = 2) were the tools used in the selected projects. Also, one case mentioned the use of the open software Open Broadcaster Software (OBS), which is used to produce streams instead of broadcasting them.

*Research question 4. Measurement strategies and results*

Lastly, to answer the last research question, the strategies utilized to collect and evaluate results were assessed. The following table shows, in different columns, the methodology of each study, the main variables or phenomena evaluated, as well as the results obtained (Table 4).

**Table 4: Analysis of the evaluations performed on the selected projects**

| N. | Methodology                                                                                                                                                          | Evidence evaluated                                                                                                                                    | Results                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | Quantitative.<br><br>26 working and 28 full-time students. Technique: Self-report a survey and comparison of the two groups through t-test                           | -Course perceptions<br><br>-Digital platform usage, and                                                                                               | -Working students found code annotations more helpful than full-time students<br><br>-Full-time students found class recordings to be more helpful<br><br>-working students exhibited a superior understanding of data analysis<br><br>-full-time students found the teacher-student interactions during live streaming to be more helpful for programming learning than the working<br><br>Students<br><br>-Similar (and high) level of learning                                                                                                                                                                                        |
| 2  | Quantitative. Sample: 145 students.<br><br>Technique: Student survey.                                                                                                | - Advantages perceived<br><br>- Disadvantages perceived.<br><br>- Preferences with respect to other formats<br><br>- Number of visits to the content. | -Advantages: willingness to participate and interact in a live broadcast, ability to reproduce the sessions as a delayed broadcast, levels of usability, openness to the community and knowledge of the platform.<br><br>-Disadvantages: a certain impersonality due to the nicks, lack of a voice channel, problems with connection, elements that remove the "seriousness" of the work, distractions of the chat, limits when accessing recorded videos, trolls.<br><br>-This experienced is preferred over exclusive face-to-face classes and the consumption of recorded videos.<br><br>-High number of live connections and visits. |
| 3  | Qualitative. Sample: 1 educator, 1 librarian and 20 students.<br><br>Technique: interviews recorded with the professor, the library administrator, and the students. | - Key elements for the success of the experience.<br><br>- Satisfaction of the participants.<br><br>- Reach of the broadcasts.                        | -Importance of technical supervision and planning.<br><br>-Benefits when the participation of students is promoted.<br><br>-Offering support to minimize connectivity problems.<br><br>-High satisfaction of the participants and high number of views.                                                                                                                                                                                                                                                                                                                                                                                  |
| 4  | Qualitative. Sample: librarian, professor, technical support personnel, students.<br><br>Technique: interviews recorded with different agents                        | - Acquisition of practical guidelines for planning.<br><br>- Aspects perceived as positive.<br><br>- Difficulties perceived.                          | -Need to chat using the students' digital media, importance of using open-source software, and not being intrusive.<br><br>-Satisfaction of all the agents, ability to connect, through the sessions, with other subjects of interest and resources available in the library.<br><br>-Difficulty for quality live productions.<br><br>-Possibility of disconnection of the students if the streamers don't catch their attention.                                                                                                                                                                                                        |

| N. | Methodology                                                                                                                                                                                                                      | Evidence evaluated                                                                                                                  | Results                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 5  | Mixed.<br>Sample: 50 students.<br>Technique: Barrado et al. survey for assessing the quality of teaching practice, plus classroom observations                                                                                   | -Interest in the subject<br>-Complement of theoretical content<br>-Improvement of teaching quality                                  | -90% of the students show an increase in the interest of the subject<br>-96% of the students satisfied with the experience<br>-98% of students confirm that the practical sessions proposed in the subject are a good complement to the theoretical contents<br>-94% of students consider that it improves the quality of their learning                                                                                                                                    |
| 6  | Quantitative.<br>Sample: 188 students.<br>Technique: Questionnaire from the universities course management system + two adhoc questionnaires                                                                                     | -Participation<br>-Elements affecting the quality of the experience<br>-Problems and issues                                         | -44% of students attend to all the streams<br>Improvement in course satisfaction (M = 4.66 / 5)<br>-1/3 of all positive feedback related to Twitch synchronous streams.<br>-Interaction in Twitch chat particularly valuable<br>-Some difficulties accessing the experience<br>-Some technical issues<br>-Still missing personal contact                                                                                                                                    |
| 7  | Qualitative. Sample: 45 students.<br>Technique: questionnaire (15 item scale).                                                                                                                                                   | - Perceived efficacy of communication.<br>- Perceived efficacy of learning.<br>- Use of YouTube Live as streaming tool.             | -The efficacy of communication of YouTube Live was considered good, in particular the levels of educator-student interaction.<br>-The efficacy of learning was considered good, in particular the possibility of being able to access the video multiple times.<br>-Depends on the digital literacy of the educator and the involvement of the university in the management of the project.                                                                                 |
| 8  | Quantitative. Sample: 128 students.<br>Technique: Validated scales. Pretest-posttest with three groups that used FB Live: One using self-regulated learning, another using experience-based learning (EBL), and a control group. | -Motivation of the students towards the experience (MSLQ)<br>- Satisfaction with the activity<br>- Learning results.                | -Motivation: Does not improve with self-regulated learning or EBL.<br>-Satisfaction: Does not improve with self-regulated learning, although it does with EBL.<br>-Learning: Does not improve with self-regulated learning or EBL, although it DOES with the combined use of both approaches.                                                                                                                                                                               |
| 9  | Mixed.<br>Sample: 52 students.<br>Technique: Case study and questionnaire with open-ended and close-ended questions.                                                                                                             | - Advantages perceived<br>- Challenges perceived<br>- Other emerging matters.                                                       | -Advantages: improvement of comprehension, increase in the desire to participate in the next sessions when perceiving feedback as personal and valuable, preference with respect to the written feedback. Increase in the relevant interactions and access to the teacher, impact on the tasks to be delivered.<br>-Perceived challenges: none beyond matters related with connections by some students.<br>-Other matters: general improvement in communication processes. |
| 10 | Qualitative. Sample: 10 active players and streamers<br>Technique: Interviews and treatment of data starting from the Grounded Theory.                                                                                           | -Perception of the education competences acquired when amplifying a learning process (in this case the use of game) through Twitch. | -General improvement of communication skills, development of ability to be "a better professor" and development of formal language or "composure" given the responsibility felt when broadcasting.                                                                                                                                                                                                                                                                          |

In all the studies we found some type of evaluation focusing on variables and/or phenomena through heterogeneous quantitative (n = 3), qualitative (n = 4), and mixed (n = 2) measurement strategies.

Although in many cases the information was based on the perception of the participants themselves, matters such as the positive effects of streaming in education processes, or the satisfaction with the experiences developed were mentioned. In some cases, these perceptions were extended to other variables such as learning.

On the other hand, some elements were pointed out for improving future experiences; these included elements such as the importance of connectivity, participants' literacy, or the fact that streaming tools and codes were not originally thought for formal educational experiences.

#### **4. Discussion and Conclusion**

The theoretical framework guiding this research—the gap between social/educational phenomena emerging in popular culture and the impact of these experiences in educational academic literature—extends beyond the realm of eLearning and targets educational innovation as a whole. Specifically, this study's emphasis on streaming—a technology and cultural phenomenon with clear influence on technology-mediated educational processes—and the fact that contemporary university education is inherently hybrid underscore the relevance of this work.

The article presents a systematic literature review on the use of streaming, understood as a contemporary cultural phenomenon, at university. The study allowed us to find different educational projects, and to analyze the academic efforts for documenting and evaluating these experiences, classifying the types of scientific productions, what elements of this culture did contain, and what evidence of its impact was measured.

While answering the first research question, the progressive application of filters reduced the number of experiences identified. This process was valuable in itself, as it showed that only a few educational experiences using streaming (as we understood it today) have been documented. This provides evidence on a great divide between today's contemporary media and culture manifestations and the main resources used in education. In fact, this seems to occur with other cultural forms in our present time, such as comic books (Serants, 2018), videogames (Oceja & González Fernández, 2018), or transmedia narratives (Dudacek, 2015). This is particularly evident in university education, as secondary education seems to be more permeable to these languages (Pozo-Sánchez, *et al.*, 2012). In a time where media (Spilker, Ask & Hansem, 2029) and society in general (Thorburn, 2014) actively participate from these forms of communication, the rigidity of university is reluctant to these practices which dilute the barrier between formal, non-formal, and informal education.

To provide an answer to the second research question, the scientific production including these practices was analyzed, confirming that some projects led to articles in indexed journals, but also conference proceedings. This could be related with the peculiarities of academic conferences, which sometimes provide the opportunity for educators to establish links with academia. In fact, some authors have reflected on the problem of overestimating academic journals as compared to other formats, which, along with congresses and conferences, could foster and represent better successful educational practices (De Vries & Pieters, 2007). The traditional gap between classrooms and scientific literature has led many authors to talk about the “failure of educational research” (Miller, 1999) or the “need [to] confront the sterility of past labors and take radical steps to conduct inquiry in more productive ways” (Reeves, 2000, p. 10).

Despite Scopus and WoS including conference proceedings within their records, it is important to point out, as a limitation of the study, that these databases have traditionally focused on journal articles, ignoring other types of literature such as book chapters and technical reports.

When answering the third question, we confirmed that only some experiences imported the culture of streaming in a comprehensive and significant manner (Pirker, Steinmaurer, & Karakas, 2021; Wang, Pradhan & Cousins, 2021). In most cases, the only identifiable elements were the platforms utilized. In this regard we think that if educators use a platform such as Twitch for being “fashionable” while maintaining a classical instructional and lecture-based approaches, the value of their proposals will be limited. Thus, we consider interesting the combination of today's streaming communication codes with the use of other active methodologies (Pirker, Steinmaurer, & Karakas, 2021; Steinmaurer & Gütl, 2023).

It was also confirmed that when educational experiences utilized open-source tools (software such as OBS) and open and flexible languages (such as the communication codes used in Twitch), although not originally created for teaching, the class was “amplified”, favoring the creation of communities, and serving to reach more people. If we consider the guidelines given in the European context on informal education and life-long learning since the 1990s (European Commission, 1999, 2001, 2006) these experiences would be aligned with the views expressed in these documents, with respect to the promotion of education at all levels and encompassing every area, from the formal to the non-formal.

According to Salinas (2013), the idea of openness is associated with quality personal learning environments. Thus, experiences using streaming culture could contribute towards the improvement of all agents learning

ecosystems (educators, students, personnel, etc.). Nothing would be more interesting, for example, than starting with streaming experiences in formal courses, with students posteriorly subscribing to those channels to broaden their knowledge in an informal way. These experiences would serve to “blend the borders between the face-to-face and the virtual, and the school-related and the non-school related”, as mentioned by Engel & Coll (2022, p. 225). In fact, what is the sense of a meaningful conversation maintained within a class not being “amplified” to other classmates, and by extension, to any person interested on the subject?

Thus, the projects documented not focusing on theoretical content transmission were particularly interesting. For example, research lines opened by Wu, Chao & Tsai (2021) linked to artistic and physical training, should be explored in-depth. Also those projects combining streaming with other emerging technologies (Calatayud, Mireia & Monsalve, 2023; Steinmaurer & Gütl, 2023) are particularly meaningful. Similarly, even though it was not strictly formal education, the work by Wang, Pradhan & Cousins (2021) was interesting, as it explores the development of competences generated in informal context such as playing videogames. Also, the work by Hertzog (2018) which proposes professionals from different sectors acting as streamers, should be followed.

As for the fourth research question, it was confirmed that the evaluation strategies were very different providing, most of them, positive results (Liu, Hung & Liang, 2023; Aniroh, Latifah, & Abdul Ghoftar Ariyanto, 2018; Wu, Chao & Tsai, 2021; Calatayud, Mireia & Monsalve, 2023; Steinmaurer & Gütl, 2023) and/or recommendations and good practices for better practices (ChanLin, 2020; Wu, Chao & Tsai, 2021). Nevertheless, many evaluations were superficial, and were based on subjective variables such as the individuals’ perception of the experience. It is important for these assessments to be complemented with others of a quantitative nature, centered on key variables such as learning achieved, PLEs improvement or the development of specific competences.

In addition, the evaluations pointed out some elements to be considered in future experiences, such as the importance of participants’ media literacy (in particular educators), the key role of live interactions, or the need to plan the projects meticulously.

The issues identified in this work (few projects, many of which are either not implemented or rigorously documented), negatively impacts areas such as engagement—a particularly pressing problem in higher education (Quaye et al., 2019)—or the missed opportunity to leverage the capital of informal learning within formal university processes (Law, 2015).

The conclusions of this work are provisional, as the research on streaming culture, its platforms, and its communication codes in higher education, is still very limited. In fact, although our limited final sample could be seen as a limitation of the study, perhaps it is also its main virtue. More specifically, the detection of deficiencies in the projects documented should allow us to formulate new research and intervention proposals.

The scarcity of projects around the educational possibilities of streaming aligns with the theoretical positions guiding this work: the existing gap between educational practice (and its permeability to popular culture) and the academic world. However, this leads us to formulate new questions that need to be addressed in the future: Are there few documented projects because academia is unable to detect them? Do university educators fail to create projects because they cannot find scientific evidence to support them?

We believe this is a multifactorial issue that is difficult to resolve. We know of educators who use Twitch to amplify classroom discussions and would never consider submitting an article to a journal. At the same time, the peculiarities of academic journals (publication timelines, trends, etc.) limit their ability to respond to these types of phenomena. Moreover, the eLearning field continues to be heavily influenced by instructional approaches that are reluctant to recognize the importance of formal and informal learning.

In today’s liquid context (Bauman, 2009), where all educational experiences are hybrid (as they are mediated by digital technologies), eLearning studies must move beyond the burden of instructional design frameworks to embrace forms of communication that break down the barriers between formal and informal learning.

In this sense, it is possible to plan some future lines of work: (1) identifying and analyzing other projects conducted by educators that have not been documented and/or that appear in grey literature; (2) improving the strategies for assessing learning; (3) designing new projects through dynamic approaches such as action-research, design-based research, and co-creation.

We believe that there was a need to explore the presence of these new forms of educommunication in the university. The article presents a complete, realistic, international, and in-depth analysis of the current situation

focusing on the existing scientific production aiming, at the same time, to serve as a basis for new transforming experiences.

## Acknowledgements

This work was carried thanks to the educational innovation projects awarded by the University of Cantabria (PID and GREDA) and also through Project PID2023-148476OA-I00 funded by MICIU/AEI/10.13039/501100011033 /10.13039/501100011033 and by FEDER, UE.

**Ethics statement:** This study does not involve human participants and is based entirely on publicly available literature.

**AI statement:** This paper was entirely written and developed by the authors without the use of artificial intelligence tools for drafting, editing, or data analysis.

## References

- Aniroh, K., Latifah, H. & Abdul Ghoffar Ariyanto, A., 2018. The effectiveness of youtube live streaming as digital learning media in tourism and guiding subject. *Arab World English Journal (AWEJ)* Special Issue on CALL, 4(4), pp.193-201. <https://doi.org/10.24093/awej/call4.15>
- Bauman, Z. (2013). *Liquid modernity*. John Wiley & Sons.
- Burroughs, B. & Rugg, A., 2014. Extending the broadcast: Streaming culture and the problems of digital geographies. *Journal of Broadcasting & Electronic Media*, 58(3), pp. 365-38. <https://doi.org/10.1080/08838151.2014.935854>
- Calatayud, L., Mireia, J. & Monsalve, L., 2023. ICT in the university environment: Twitch and Instagram as learning tools. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 18(1), pp.70-75. <https://doi.org/10.1109/RITA.2023.3250557>
- Castañeda, L., Salinas, J. & Adell, J., 2020. Hacia una visión contemporánea de la Tecnología Educativa. *Digital Education Review*, (37), pp. 240-268. Available at: <https://raco.cat/index.php/DER/article/view/375035>
- Catá, A. S., 2019. Convergence of rhetoric, labour, and play in the construction of inactive discourses on Twitch. *Digital Culture & Society*, 5(2), pp. 133-148. <https://doi.org/10.14361/dcs-2019-0209>
- ChanLin, L.J., 2020. Engaging university students in an ESL live broadcast. *The Electronic Library*, 38(1), pp.28-43. <https://doi.org/10.1108/EL-08-2019-0186>
- De Vries, B. & Pieters, J., 2007. Knowledge sharing at conferences. *Educational Research and Evaluation*, 13(3), pp. 237-247. <https://doi.org/10.1080/13803610701626168>
- Dudacek, O., 2015. Transmedia storytelling in education. *Procedia-Social and Behavioral Sciences*, 197, pp. 694-696. <http://dx.doi.org/10.1016/j.sbspro.2015.07.062>
- Engel, A. & Coll, C., 2022. Entornos híbridos de enseñanza y aprendizaje para promover la personalización del aprendizaje. *RIED. Revista Iberoamericana de Educación a Distancia*, 25(1), pp.225-242. <https://doi.org/10.5944/ried.25.1.31489>
- European Commission, 1999. *On the implementation, results and overall assessment of the European Year of Lifelong Learning*. Available at: <http://aei.pitt.edu/4937/1/4937.pdf>
- European Commission, 2001. *Making a European Area of Lifelong Learning a Reality*. Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2001:0678:FIN:EN:PDF>
- European Commission, 2006. *Recommendation on key competences for lifelong learning*. Available at: <https://eur-lex.europa.eu/LexUriServ/LexUri>
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A. & Pappas, G., 2008. Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB journal*, 22(2), pp. 338-342. Available at: <https://doi.org/10.1096/fj.07-9492LSF>
- Flores-Saviaga, C., Hammer, J., Flores, J. P., Seering, J., Reeves, S. & Savage, S., 2019. Audience and streamer participation at scale on twitch. In *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, pp. 277-278. <https://doi.org/10.1145/3342220.3344926>
- Garrigós, F.J., Narangajavana, Y. & Estellés, S., 2022. Twitch y su uso en educación. In *Proceedings INNODOCT/21. International Conference on Innovation, Documentation and Education*. Editorial Universitat Politècnica de València, pp. 699-709. <https://doi.org/10.4995/INN2021.2021.13949>
- Hertzog, P.E., 2018. The use of live streaming in design-based modules for open distance learning. *World Transactions on Engineering and Technology Education*, 16(3), pp.224-225.
- Jesson, J., Matheson, L. & Lacey, F. M., 2011. *Doing your literature review: Traditional and systematic techniques*. Sage.
- Kulkarni, A. V., Aziz, B., Shams, I. & Busse, J. W., 2009. Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *Jama*, 302(10), pp. 1092-1096. <https://doi.org/10.1001/jama.2009.1307>
- Law, P. (2015). Digital badging at The Open University: recognition for informal learning. *Open Learning: The Journal of Open, Distance and e-Learning*, 30(3), pp. 221-234. <https://doi.org/10.1080/02680513.2015.1104500>
- Li, K., Rollins, J. & Yan, E., 2018. Web of Science use in published research and review papers 1997-2017: A selective, dynamic, cross-domain, content-based analysis. *Scientometrics*, 115(1), pp. 1-20. <https://doi.org/10.1007/s11192-017-2622-5>

- Liu, I.F., Hung, H.C. & Liang, C.T., 2023. A study of programming learning perceptions and effectiveness under a blended learning model with live streaming: Comparisons between full-time and working students. *Interactive Learning Environments*, pp.1-15. <https://doi.org/10.1080/10494820.2023.2198586>
- Merriam-Webster, 2022. Dictionary/Encyclopedia entry. In *Merriam-Webster*, Merriam-Webster, Incorporated.
- Miller, D. W., 1999. The black hole of education research. *Chronicle of Higher Education*, 45(48), pp. 17-18.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, 151(4), pp. 264-269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>
- Oceja, J. & González Fernández, N., 2018. Videojuegos y aprendizaje. ¿Por qué la gamificación y los juegos educativos no son suficientes? In A. Torres-Toukoumidis & L. M. Romero-Rodríguez, eds. *Gamificación en Iberoamérica. Experiencias desde la Comunicación y la Educación*. Quito: Abya-Yala. Available at: <<http://dspace.ups.edu.ec/handle/123456789/17051>>
- Pirker, J., Steinmaurer, A. & Karakas, A., 2021. Beyond Gaming: The Potential of Twitch for Online Learning and Teaching. In *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1*, pp. 74-80. <https://doi.org/10.1145/3430665.3456324>
- Pozo-Sánchez, S., López-Belmonte, J., Fuentes-Cabrera, A. & López-Núñez, J. A., 2021. Twitch as a techno-pedagogical resource to complement the flipped learning methodology in a time of academic uncertainty. *Sustainability*, 13(9), p. 4901. <https://doi.org/10.3390/su13094901>
- Purssell, E. & McCrae, N., 2020. *How to perform a systematic literature review: a guide for healthcare researchers, practitioners and students*. Springer Nature.
- Quaye, S. J., Harper, S. R., & Pendakur, S. L. (Eds.). (2019). *Student Engagement in Higher Education: Theoretical Perspectives and Practical Approaches for Diverse Populations*. Routledge.
- Reeves, T. C., 2000. Enhancing the worth of instructional technology research through “design experiments” and other development research strategies. *International Perspectives on Instructional Technology Research for the 21st Century*, 27, pp. 1-15.
- Salinas, J., 2013. Enseñanza flexible y aprendizaje abierto, fundamentos clave de los PLEs. In *Entornos Personales de Aprendizaje: Claves para el ecosistema educativo en red*. Editorial Marfil, p. 53. Available at: <<http://hdl.handle.net/10201/30410>>
- Serantes, L. C., 2018. The possibilities of comics reading: Uncovering a complex and situated reading experience in Canada. In *Young People Reading*. Routledge, pp. 19-38.
- Sjöblom, M., Törhönen, M., Hamari, J. & Macey, J., 2019. The ingredients of Twitch streaming: Affordances of game streams. *Computers in Human Behavior*, 92, pp. 20-28. <https://doi.org/10.1016/j.chb.2018.10.01>
- Spilker, H. S., Ask, K. & Hansen, M., 2020. The new practices and infrastructures of participation: How the popularity of Twitch.tv challenges old and new ideas about television viewing. *Information, Communication & Society*, 23(4), pp. 605-620. <https://doi.org/10.1080/1369118X.2018.1529193>
- Steinbeck, H., Teusner, R. & Meinel, C., 2021. Teaching the masses on Twitch: An initial exploration of educational live-streaming. *Proceedings of the Eighth ACM Conference on Learning@ Scale*, June, pp. 275-278.
- Steinmaurer, A. & Gütl, C., 2023. Implementation and Experiences of a Flipped Lecture Hall - A Fully Online Introductory Programming Course. In: Auer, M. E., Pachatz, W. & Rüttmann, T., eds. *Learning in the Age of Digital and Green Transition*. ICL 2022. *Lecture Notes in Networks and Systems*, vol 633. Springer. [https://doi.org/10.1007/978-3-031-26876-2\\_79](https://doi.org/10.1007/978-3-031-26876-2_79)
- Thorburn, E. D., 2014. Assemblages: Live Streaming Dissent in the ‘Quebec Spring’. In *Social Media, Politics and the State*. Routledge, pp. 149-167.
- Toffler, A., 1980. *The third wave*. William Morrow.
- Twitch statistics & charts, 2018. *Budgetbytes*. Available at: <https://www.budgetbytes.com/tuscan-white-bean-pasta/> [Accessed 22 Nov. 2023].
- Wang, S., Pradhan, S. & Cousins, K., 2021. Toward a Game-Based Dialogical Pedagogy: Insights from Massively Multiplayer Online Role-Playing Games. In *European Conference on Games Based Learning*. Available at: <https://doi.org/10.34190/GBL.21.010>
- Woodcock, J. & Johnson, M. R., 2019. The affective labor and performance of live streaming on Twitch.tv. *Television & New Media*, 20(8), pp. 813-823. Available at: <https://doi.org/10.1177/1527476419851077>
- Wohn, D.Y. and Freeman, G., 2020. Live streaming, playing, and money spending behaviors in esports. *Games and Culture*, 15(1), pp.73-88. <https://doi.org/10.1177/1555412019859184>.
- Wu, C.C., Chao, H.W. & Tsai, C.W., 2021. The Effects of Facebook Live-Stream Teaching on Improving Students' Dance Skills: Impacts on Performance, Learning Motivation, and Physical Activity Class Satisfaction. *International Journal of Mobile and Blended Learning (IJMBL)*, 13(4), pp.45-62. <https://doi.org/10.4018/IJMBL.2021100103>
- Wulf, T., Schneider, F.M. & Beckert, S., 2020. Watching players: An exploration of media enjoyment on Twitch. *Games and Culture*, 15(3), pp.328-346. <https://doi.org/10.1177/1555412018788161>

# An Analysis of Factors Impacting Users' Choice of Freemium or Premium Services in a Mobile-Assisted Language Learning App

Farah Dita Ashilah, Nurul Hanina Efendi, Yelda Faizah Havara, Putu Wuri Handayani, Nabila Clydea Harahap

Faculty of Computer Science, Universitas Indonesia, Indonesia

[farah.dita@ui.ac.id](mailto:farah.dita@ui.ac.id)

[nurul.hanina@ui.ac.id](mailto:nurul.hanina@ui.ac.id)

[yelda.faizah@ui.ac.id](mailto:yelda.faizah@ui.ac.id)

[putu.wuri@cs.ui.ac.id](mailto:putu.wuri@cs.ui.ac.id) (corresponding author)

[nabila.clydea@ui.ac.id](mailto:nabila.clydea@ui.ac.id)

<https://doi.org/10.34190/ejel.23.1.3894>

An open access article under [CC Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)

**Abstract:** Since the COVID-19 pandemic, the popularity of foreign language learning on mobile-assisted language learning (MALL) applications has increased. The MALL also implements a freemium (free access) and premium business model where customers must purchase the subscription package provided. This research aims to find out what factors influence the migration of users from the free version to the premium MALL version. This study uses a quantitative and qualitative approach. Quantitative data obtained through questionnaires filled out by 1,232 respondents and qualitative data obtained through interviews with thirty interviewees. The results of this study show that dissatisfaction in the push factor was found to be factors that significantly influenced the user's desire to switch to the premium version of the MALL, while perceived intrusiveness and perceived inefficiency did not have a significant positive effect on the user's desire to switch to the premium version of the MALL. The pull factor, perceived enjoyment, perceived usefulness, and trialability showed a significant positive influence, while the perceived learner value aspect did not show a significant positive influence. Finally, all aspects of the mooring factor, namely switching cost, social influence, and inertia show a significant influence on user intention to switch to the premium version of the MALL. However, switching cost showed positive influence on users' switching intention, contrary to the predicted hypotheses. The results of this research are expected to be an input for MALL application developers to improve their services so that users switch to using the premium version of the MALL.

**Keywords:** Mobile-assisted language learning, Switching intention, Pull-push-mooring, Indonesia

## 1. Introduction

M-learning is now a global trend, especially for foreign language learning (Toto & Limone, 2019). With the increasing penetration of mobile technology into human daily life, m-learning is becoming integrated with foreign language learning. Mobile assisted language learning (MALL) is used to describe the use of mobile technology in foreign language learning (Arvanitis & Krystalli, 2021). Aliakbari and Mardani (2022) found that MALL is one of the foreign language learning approaches that provides the most benefits to learners and can improve foreign language skills. In Indonesia, learning a foreign language is a necessity to be able to enter the global community due to globalization and the opportunity to exchange knowledge with other countries (Sulyati, 2020). To date, Duolingo is the most prominent foreign language learning application globally (Ceci, 2023). Since the pandemic, the interest of Indonesian people in learning foreign languages through Duolingo has increased drastically (Anggraeni, 2023).

Duolingo implemented game-based elements in its app to engage individuals, such as leaderboards, level-system, badges, and rewards with the aim of assisting its users in improving their foreign language skills (Huynh, Zuo, and Iida, 2016). Through providing useful features for free, Duolingo hopes to attract users to switch to the premium version, where they can enjoy additional benefits and a more immersive learning experience. The premium version of Duolingo, also known as Super Duolingo, offers a paid service with no ads and no access restrictions, as well as providing additional features such as personalized lessons (Purnama, 2023). User switching behavior refers to the actions taken by users when they choose an alternative service over the previous one (Sun et al., 2017), such as switching from the free version to the premium version. Tsai (2022) explained the importance of recognizing the difference between free version users and premium versions because the factors that affect free version users to use premium are diverse.

The push-pull-mooring (PPM) model is often used by previous studies because it has been proven to accurately test users' switching intention to switch from one condition to another (Tsai, 2022; Lisana, 2022; Wang & Shin, 2022; Chen & Keng, 2019; Nayak et al., 2021). One of them is a study that explores the factors that affect students in switching to using m-learning as a way of learning which used to be in the form of face-to-face classes (Lisana, 2022). Magsayo (2022) also conducted a similar study with student research subjects in rural areas of the Philippines. Chen and Keng (2019) conducted a study in the context of switching intention from face-to-face English courses to online courses.

In addition to the educational context, this research will also focus on identifying the factors that affect the use of the premium version. One of the previous studies that resonated with this context was Tsai's (2022) research on switching user intentions to switch to the premium version of the over-the-top (OTT) platform, which is part of video on demand (VOD), in Taiwan using the PPM model framework. Among the previous studies that have been discussed, there has been no research that specifically discusses the transition of users of the free version to the premium version of the MALL application, especially Duolingo, with the framework of the PPM model. Thus, this study will adopt the framework of the PPM model. Also supported by Nayak et al. (2021) stated that the PPM model has the advantage of adapting to the context of the research to be carried out because it does not determine fixed variables in the aspects of push, pull, or mooring. Thus, the question of this research is what push, pull, and mooring factors affect the switching intention of MALL young customers to premium services? This research can provide guidance for MALL service providers in developing strategies to make users switch from free to premium services.

## 2. Literature Review

We used push-pull-mooring (PPM) as a main reference of our proposed model. Push-pull-mooring (PPM) is based on migration theory which is influenced by push and pull factors. The ability of the PPM model framework in analyzing factors that affect switching intention in the context of education has been proven by various previous studies (Chen & Keng, 2019; Nayak et al., 2021; Lin et al., 2021; Lisana, 2022). The selection of research variables was carried out by taking constructs from previous studies that have been proven to affect switching intentions. In addition, to enrich the context related to factors that can affect the use of an educational platform, this study also refers to research that discusses the adoption of an e-learning system. This study adopts several factors that are appropriate to the context of the research, namely the influence that resulted in users of the free version of Duolingo switching to the premium version of Duolingo.

The push factor is a life situation that causes dissatisfaction with the environment in which it is occupied, thus prompting an individual to leave (Dorigo & Tobler, 1983; Bansal, 2005). The pull factor is an attribute of another place that makes it appear attractive (Dorigo & Tobler, 1983). The mooring factor is defined as a situational factor that moderates the decision to migrate or not (Bansal, 2005; Tsai, 2022). The main research that will be referred to in this study discusses the factors that affect the continuity intention of MALL (Faozi & Handayani, 2023) and switching intention (Tsai, 2022; Ye et al., 2022; Hu et al., 2023; Lisana, 2022; Magsayo, 2022; Zhu, Peng and Yang, 2023). Based on the referenced studies, we chose variables that were proven to have an influence on switching intention and had a context relevant to this study. Figure 1 shows the research model used with 10 hypotheses and 11 variables.

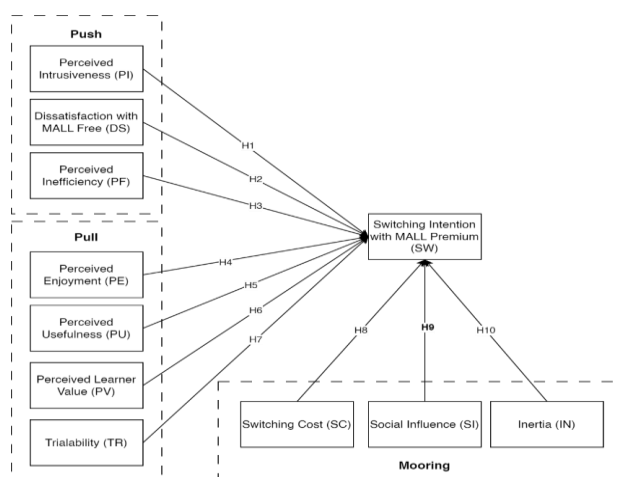


Figure 1: Proposed Research Model

Perceived intrusiveness is the level of how disturbed the user arises due to impaired individual cognitive processes (Li, Edwards, and Lee, 2002; Tsai, 2022). In the context of application use, disruption of cognitive processes can be caused by the appearance of advertisements (Li, Edwards, and Lee, 2002). Foulds, Azzopardi, and Halvey (2021) found users take much longer to complete tasks when there are ads and users also report more negative experiences such as frustration levels, distractions, and difficulty completing tasks. The distraction of these ads can cause irritation and trigger users to avoid the ads (Wei, Ko, and Pearce, 2021). Wei, Ko, and Pearce (2021) found that the presence of advertising disorders or perceived intrusiveness can increase the tendency of social media users to avoid advertising due to annoyance. Therefore, we define the following hypothesis.

*H1: Perceived intrusiveness (PI) has a positive influence on switching intention to MALL premium (SW)*

User satisfaction is often linked to the user's intention to use an application (Faozi & Handayani, 2023; Pozon-Lopez et al., 2020). Pozon-Lopez et al. (2020) found that user satisfaction is an important predictor of the intention to use MOOC platforms. Kuo (2020) found that user dissatisfaction causes consumers to switch to other mobile payment services. Dissatisfaction refers to the feeling of dissatisfaction of users with the product/service being used because they fail to meet the users' expectations or needs (Yoon and Lim, 2021). Sekarputri et al. (2024) found that users who feel uncomfortable using an app tend to stop using it and are encouraged to switch to another application. This is in line with Ye et al. (2022) who found that dissatisfaction is one of the factors that encourages e-commerce live streaming (ELS) users to leave the old platform and switch to another platform. Ye et al. (2022) and Sekarputri et al. (2024) prove that dissatisfaction can affect users' intention to switch to a product. Thus, we suggest the following hypothesis.

*H2: Dissatisfaction (DS) has a positive influence on switching intention to MALL premium (SW)*

Perceived inefficiency refers to the negative experience felt by individuals regarding the use of resources that have not been optimized (Phan and Ngu, 2021). An individual tends to consider the efficiency aspect when deciding to use a new technology and hopes to get an increase in efficiency in his/her activities (Murillo-Zegarra et al., 2020). Sarkar, Sarkar, and Sreejesh (2021) and Mazhar (2022) found that inefficiency can lead to dissatisfaction in customers and subsequently make customers switch to other services. Perceived inefficiency also refers to how far individuals have had negative experiences with inadequate content and wasted time when using a product or service (Kim, Choi and Choi, 2019). Zhu, Peng and Yang (2023) found that the intention to switch from lecturers who teach natural science courses has a high tendency to be influenced by perceived inefficiency felt in traditional classrooms. Then, we define the following hypothesis.

*H3: Perceived inefficiency (PF) has a positive influence on switching intention to MALL premium (SW)*

Generally, on platforms with a freemium business model, application developers deliberately reduce the sense of enjoyment of their users by limiting access or features to create demand for premium services (Hamari, Hanner, Koivisto, 2019). A sense of pleasure has a significant influence on the intention to purchase a product (Yulianto & Oroh, 2021) and is one of the main predictors of a platform's user intent to purchase premium services (Yulianto & Oroh, 2021). In a study conducted by Yulianto and Oroh (2021), it was found that a sense of pleasure has a significant influence on the intention to purchase premium services. In the context of education, perceived enjoyment can be defined as how learners enjoy using mobile learning to increase knowledge regardless of possible performance consequences (Kim, et al., 2020; Tsai, 2022). Previous studies have found that perceived enjoyment has a great influence on students' intention to adopt a new mobile learning system (Pramana, 2018; Lisana & Suciadi, 2021; Tsai, 2022). Tsai (2022) found that perceived enjoyment is one of the pull factors that can affect users' intention to switch using mobile learning systems. Previous studies have found a positive influence between perceived enjoyment and the intention to switch from free to premium services (Yulianto & Oroh, 2021) and on the use of mobile learning (Tsai, 2022). Thus, we propose the following hypothesis.

*H4: Perceived enjoyment (PE) has a positive influence on switching intention to MALL Premium (SW)*

Perceived usefulness is how the perception of users of an information system can improve the performance of their activities (Al-Rahmi et al., 2021). Perceived usefulness was found to have a great influence on adoption behavior in information system users (Ye et al., 2022). If users realize that there are useful uses for a new technology, then users will show positive behaviors that can affect their intention to adopt the technology (Zhu, Peng and Yang, 2023). As a result, the perceived usefulness of a platform can significantly affect a user's switching intention (Ye et al., 2022). Qashou (2020) also found that perceived usefulness has a significant positive influence on the intention to use mobile learning. Therefore, we suggest the following hypothesis.

*H5: Perceived usefulness (PE) has a positive influence on switching intention to MALL premium (SW)*

Perceived learner value refers to how users perceive the learning experience from the use of mobile learning (Magsayo, 2022). The adoption of mobile learning systems can be influenced by how the system can facilitate academic goals and satisfaction in the learning process (Magsayo, 2022). Perceived learner value describes a person's level of interest in learning and a sense of enthusiasm to get good results (Magsayo, 2022). These variables can help users to realize the value of adopting mobile learning in user learning (Magsayo, 2022). Ain, Kaur and Waheed (2016) argue that students' positive perception of learning through a learning management system (LMS) can affect students' intentions in spending more time and effort to be able to gain knowledge from the LMS. A learning platform will be considered to have more value if students are willing to spend time and effort to learn and how the system can facilitate quality learning (Ain, Kaur and Waheed, 2016; Magsayo, 2022). Then, we propose the following hypothesis.

*H6: Perceived learner value (PV) has a positive influence on switching intention to MALL premium (SW)*

Trialability is defined as a level at which a product innovation can be evaluated under limited conditions, such as a limited period or environment (Hsieh, 2021). Previous research has argued that a technological innovation will be easier to adopt if potential users can conduct trials first because that will reduce perceived risks (Raman et al., 2021; Wang, Douglas, and Hazen, 2021). This argument is supported by several studies on technology adoption that prove that trialability is an important aspect to include (Raman et al., 2021; Wang, Douglas, and Hazen, 2021; Hidayat-ur-Rehman et al., 2022). Hsieh (2021) proves that trialability is one of the crucial variables that has a positive impact on switching intention. Thus, we suggest the following hypothesis.

*H7: Trialability (TR) has a positive influence on switching intention to MALL premium (SW)*

The cost aspect is an important factor that needs to be tested on the switching intention variable (Wu, Vassileva, and Zhao, 2017; Monoarfa et al., 2023). Past study found that switching cost is a theory that is suitable to be used in proving user perception when using a new service (Yuen, Ng, and Wang, 2023). Wu, Vassileva, and Zhao (2017) researched the behavior of internet users towards the change of memory storage methods in the cloud, and found that low switching costs are interpreted when users are not too worried about losing effort in the service switching process. Switching cost is defined as the user's perception of sacrifices, both financial and non-financial, which are considered in carrying out new behaviors (Monoarfa et al., 2023). Then, we define the following hypothesis.

*H8: Switching cost (SC) has a negative influence on switching intention to MALL premium (SW)*

Social influence is defined as the perception of users over the opinion of important people (family, friends, or other relatives) that they should use a new technology, product, or service (Yee et al., 2022). Previous research has demonstrated the influence of social influence in predicting switching intentions. For example, a user in the transition to e-commerce live streaming (Ye et al., 2022) and the individual shift from cash payments to mobile payments (Hsieh, 2021). In the three studies, the social influence variable was placed as a mooring variable. In addition to predicting switching intention, social influence factors are also often used to predict other behavioral intentions, such as predicting the adoption rate of mobile payments in the younger generation (Wei, , Ko, and Pearce, 2021). Thus, we suggest the following hypothesis.

*H9: Social influence (SI) has a positive influence on switching intention to MALL premium (SW)*

When dealing with a new product or service, users may be influenced by inertia and choose not to move away from the original option (Sun et al., 2017). In making changes to the use of products, there are risks and costs that must be incurred by users (Monoarfa et al., 2023). In minimizing the uncertainty of risks and costs, users tend to choose more familiar products so that inertia is proven to affect the user's tendency to have switching intention (Hu et al., 2023). Inertia is defined as the user's preference to maintain the current behavior on the original option when faced with a better alternative judged from previous experience (Samuelson & Zeckhauser, 1988). In the context of MALL, inertia is defined as the tendency of users to choose to continue using the free version of the MALL. Finally, we propose the following hypothesis.

*H10: Inertia (IN) negatively influence switching intention to MALL premium (SW)*

### 3. Methodology

#### 3.1 Research Design

This study uses a mixed method approach by first distributing questionnaires and then followed by conducting interviews with respondents. Before distributing the questionnaire, we conducted a readability test on the respondents to confirm the accuracy of the use of diction and grammar, as well as the ease of understanding the questions. The readability test was conducted by interviewing 11 respondents face-to-face and online using video conferencing platforms, such as Google Meet and Zoom Meeting. Furthermore, a pilot study was conducted on 35 respondents to analyze Cronbach's Alpha (CA) scores. The CA result from the pilot study is 0.896 so that a wider distribution of questionnaires can be carried out. The questionnaire link is shared through social media platforms such as LINE, WhatsApp, Instagram, X, Telegram, Facebook which are widely used by Indonesians. The quantitative data collection period lasted about one month, from February 17, 2024, to March 16, 2024. The object of research in this study is one of the MALL platforms, namely Duolingo. The criteria for respondents involved in this study are respondents who have used the free version of Duolingo and know the offers provided by the premium version of Duolingo. All respondents have agreed to participate in filling out the questionnaire. Table 1 explains the summary of the demographics of the respondents involved in this study.

**Table 1: Summary of Demographic Data of Quantitative Respondents**

| Demographics |                                                | Number of Respondents | Percentage |
|--------------|------------------------------------------------|-----------------------|------------|
| Gender       | Man                                            | 406                   | 33.0%      |
|              | Woman                                          | 826                   | 67.0%      |
| Age          | < 17 years old                                 | 13                    | 1.1%       |
|              | 17 - 25 years                                  | 980                   | 79.5%      |
|              | 26 - 35 years old                              | 198                   | 16.1%      |
|              | 36 - 45 years old                              | 38                    | 3.1%       |
|              | > 45 years                                     | 3                     | 0.2%       |
| Education    | High school or below                           | 836                   | 67.9%      |
|              | Diploma                                        | 52                    | 4.2%       |
|              | Undergraduate                                  | 332                   | 26.9%      |
|              | Master student                                 | 12                    | 1.0%       |
| Domicile     | Greater Jakarta                                | 457                   | 37.1%      |
|              | Outside Greater Jakarta in Java Island         | 572                   | 46.4%      |
|              | Sumatra                                        | 129                   | 10.5%      |
|              | Kalimantan                                     | 33                    | 2.7%       |
|              | Sulawesi                                       | 27                    | 2.2%       |
|              | Bali, Nusa Tenggara Barat, Nusa Tenggara Timur | 12                    | 1.0%       |

| Demographics |                                      | Number of Respondents | Percentage |
|--------------|--------------------------------------|-----------------------|------------|
|              | Maluku and Papua                     | 2                     | 0.2%       |
| Occupation   | Unemployed                           | 44                    | 3.6%       |
|              | Students                             | 797                   | 64.7%      |
|              | Private employees                    | 184                   | 14.9%      |
|              | Employees of state-owned enterprises | 13                    | 1.1%       |
|              | Self employed                        | 71                    | 5.8%       |
|              | Civil servants                       | 9                     | 0.7%       |
|              | Housewives                           | 81                    | 6.6%       |
|              | Others                               | 33                    | 2.7%       |

### 3.2 Analysis Method

Data processing and analysis in this study used the covariance-based structural equation modeling (CB-SEM) method. The data processing process is carried out using several applications, including Google Sheet and AMOS 26. After quantitative data processing, we also collect data through interview sessions with respondents. This stage aims to confirm and identify supporting arguments for all hypotheses. The respondents involved in the interview were 30 respondents where most of the interviewees were aged 17-25 years, were undergraduate students, and resided in the Greater Jakarta area. Each interview session will be documented through a recording that has been approved by the respondent to be recorded. All data obtained from this study is anonymous. After conducting interviews with respondents, we used the content analysis method to analyze qualitative data.

### 3.3 Research Instruments

In this study, a questionnaire was used to collect quantitative data which was generally divided into demographic and measurement items totaling 44 statements. The answer to the statement uses a Likert scale with a value of 1 describing "Strongly Disagree" and a value of 5 describing "Strongly Agree". Appendix A explains the measurement items and Appendix B explains the interview questions used in this study.

## 4. Result

The first stage in the quantitative data analysis process is the specification of the research model by first making a path diagram. The loading factor value for each measurement item is above 0.7. The average variance extracted (AVE) value for each variable is above 0.5 and the CA and CR values are above 0.7 for each variable so that it passes the validity and reliability test according to Hair et al. (2014). Appendix C shows the AVE, CA and CR values for each variable. Before being able to conduct a hypothesis test, we conduct a Goodness of Fit (GoF) evaluation. In the structural model test, Goodness of Fit (GoF) was evaluated through Minimum Discrepancy Function by Degrees of Freedom divided (CMIN/df), Root Mean Square Error of Approximation (RMSEA), Normed Fit Index (NFI), Comparative Fit Index (CFI), Goodness of Fit Index (GFI), and Tucker-Lewis index (TLI) metrics. Appendix D shows the results of the GoF that are in accordance with the cut-off value.

The hypothesis test in this study will be carried out in a one-tailed with a significance level of five percent. The hypothesis is considered to have an influence if the p-value has a value of  $<0.05$  and has no influence if the value is  $\geq 0.05$  (Hair et al., 2014). In one-tailed research, the direction of influence needs to be examined to be able to conclude if a hypothesis is accepted (Kock, 2015). Table 2 explains the results of the hypothesis test.

Table 2: Hypothesis Testing Result

| Hypothesis |    |      |     | P     | Description |
|------------|----|------|-----|-------|-------------|
| H1         | SW | <--- | PI  | 0.181 | Rejected    |
| H2         | SW | <--- | DS  | 0.003 | Accepted    |
| H3         | SW | <--- | PF  | 0.053 | Rejected    |
| H4         | SW | <--- | PE  | 0.003 | Accepted    |
| H5         | SW | <--- | PU  | 0.047 | Accepted    |
| H6         | SW | <--- | PV  | 0.308 | Rejected    |
| H7         | SW | <--- | TR  | 0.003 | Accepted    |
| H8         | SW | <--- | SC  | 0.037 | Rejected    |
| H9         | SW | <--- | THE | 0.002 | Accepted    |
| H10        | SW | <--- | IN  | 0.002 | Accepted    |

## 5. Discussion

This study found that there was no effect of perceived intrusiveness on switching intention (SW). The results of this study are not in accordance with Tsai (2022), Foulds, Azzopardi, and Halvey (2021), and Wei, Ko, and Pearce (2021). Foulds, Azzopardi, and Halvey (2021) found that users reported negative responses due to the appearance of ads, such as feeling frustrated, distracted, and having difficulty completing their tasks. Tsai (2022) found that ad disruption is one of the main factors that encourage users to switch from a free OTT platform to a premium. Based on the interview data, it can be concluded that interviewees of the free version of MALL consider advertising not a significant distraction. This is because of the frequency of ads that can still be tolerated (*"There are not too many ads because they are only there after playing one unit of lessons."* - Interviewee 6). Some users feel that advertising can be an opportunity to take a break in the middle of learning to use the MALL (*"I consider it as a chance to take a break after a unit of lessons, I never mind the ads,"* - Interviewee 6).

This study also shows that dissatisfaction has a positive and significant influence on the intention of users to switch to premium services. This is in line with Kuo (2020) who stated that low user satisfaction or user dissatisfaction causes consumers to switch to other mobile payment services. Pozón-López et al. (2020) also stated that user satisfaction is an important factor that influences the intention to use MOOC platforms. This is contrary to Sarkar, Sarkar, and Sreejesh (2021) and Mazhar et al. (2022) who found that inefficiency can lead to dissatisfaction in customers and subsequently make customers switch to other services. Zhu, Peng and Yang (2023) found that the intention to switch lecturers who teach natural science courses has a high tendency to be positively influenced by the inefficiency of traditional classrooms. However, the results of this study are in line with Kim et al. (2019) who found that perceived inefficiency has no positive effect on individuals' intention to switch from traditional content services to content services with augmented or virtual reality technology.

This research shows that perceived enjoyment has a positive influence and significantly affects the user's intention to switch to premium services. Tsai (2022) also showed the same results, namely the significant positive influence of the perceived enjoyment variable that can encourage user intention to switch from traditional classes to the use of mobile learning. Yulianto and Oroh (2021) also showed a positive influence of perceived enjoyment on switching intention from a free service to a paid service. Feeling pleasure in using a product can encourage users to buy the product (Yulianto & Oroh, 2021). Of the total 30 interviewees, 77 percent of them said that with the loss of annoying restrictions by using the premium version, they can feel more comfortable and freer to use MALL. In addition, 10 interviewees also felt that the premium version of the MALL provides more variety of lessons and personalization that it makes users do not easily feel bored and can be more excited in using the MALL platform (*"I am more excited to learn focus better on learning because of the personalized lesson feature. Also, I can choose freely which material or topic I want to learn"* - Interviewee 16).

In addition, perceived usefulness has a positive influence on the user's intention to switch to premium services. The results of this hypothesis test are consistent with Ye et al. (2022). Ye et al. (2022) found that the perceived usefulness has a positive influence that can increase user intention to move from one ELS to another ELS platform. Based on the results of interviews, 63 percent of the interviewees agreed that the features in the premium version of the MALL are useful and can meet the needs of users (*"I feel that the features in the premium*

*version can fulfill my needs because I mainly want to be able to play without being limited by hearts," - Interviewee 2).*

Perceived learner value did not have a positive influence on the intention of users to switch from the free version of MALL to the premium version in this study. This contrasts with Magsayo (2022), Harja, Irawan and Ambarwati (2019) and Zacharis and Nikolopoulou (2022) who found that perceived learner value is a significant variable that can affect the intention of students to adopt a new mobile learning platform. Based on the results of the interview, MALL users feel that they do not get a different learning process in the premium version than the free version. This shows that there is no extra value to be gained from the learning process in the premium version of the MALL that can motivate users to commit to a subscription (*"The features in Duolingo premium are not too helpful for learning if compared to the free version. The useful features are the features that already exist by default, which is the gamification, for example, leaderboards and point systems."* - Interviewee 8).

Then, this study also found that there was a positive influence between trialability on switching intention to premium MALL. This is consistent with Hsieh (2021) who found that trialability is one of the crucial variables that positively affects the intention to switch from traditional health facility transactions to the use of technology. The experience of exploring or testing a product is an important opportunity (Raman et al., 2021; Wang, Douglas, and Hazen, 2021; Hidayat-ur-Rehman et al., 2022). The same thing was found in Raman et al. (2021), in which the existence of a trial period can affect the decision to adopt a new e-learning system by students. The results of the interview showed that 73 percent of the 30 interviewees were helped by the free trial and considered that the free trial was an important opportunity before committing to subscribe to the premium version of the MALL (*"I feel that free trial is very important so users don't feel cheated if it turns out that the premium version is not suitable for them when users have already paid,"* - Interviewee 16).

The study found that the sacrifices that users must make have a significant positive influence on their intention to subscribe to the premium version of the MALL. The results of this study are contrary to Hsieh (2021) which proves that switching costs have a negative effect on the intention to switch the behavior of patient transaction methods from cash to medical mobile payment. Anis and Noor (2021) also found that there was no moderation effect of switching costs that affected user satisfaction or intention to switch. Zakiy and Haryanto (2021) sees that even switching costs have a positive effect on customer loyalty. The higher the switching cost that is likely to be incurred by the customer, the higher the customer's loyalty not to turn away from the bank's services. Subscription prices are the main reason why users are reluctant to switch. The results of the interview showed that 57 percent of the interviewees felt that the costs that needed to be incurred to subscribe to the premium tended to be inexpensive (*"I think the price is still affordable. At that price, you can learn many languages, not just one language"* - Interviewee 1).

The study also found that users' perception of the opinions of important people around them had a positive influence on the intention to subscribe to the premium version of the MALL. The results of this study are consistent with Ye et al. (2022) which prove the influence of relatives in switching to other live streaming platforms. Hsieh (2021) showed that social influence has a positive impact on switching intentions, both between companies in the same offering and from offline to online transactions. Half of the interviewees agreed that opinions from friends, family, and reviews on social media influenced users to switch to the premium version (*"If I have friends who use the premium version of Duolingo, it can influence me to be interested in subscribing."* - Interviewee 25).

Finally, this study shows that inertia negatively affects the switching intention of users of the free version of Duolingo to the premium version. These results are consistent with Hsieh (2021) which proves that inertia makes patients less motivated to use medical mobile payment. An example of inertia action is that patients are used to relying on cash payment methods because they are more commonly used (Hsieh, 2021). Ye et al. (2022) in e-commerce live streaming (ELS), Hu et al. (2023) in the adoption of electric vehicles in China, and Dogra, Bakshi, and Gupta (2022) in e-health, all prove that inertia affects the intention to switch from one service to another. In addition, some interviewees feel that their needs have been met when using the free version of the MALL so there is no high urgency to switch to the premium version of the MALL (*"I feel that the free feature is enough to meet my casual use needs. I don't have a problem with the limitations of Duolingo,"* - Interviewee 6).

## 6. Conclusion

This study found that there is one push factor (dissatisfaction) and three pull factors (perceived enjoyment, perceived usefulness, and trialability) that positively affect users' intention to switch to the premium version of the MALL. Perceived intrusiveness as one of the push factors proposed in this research model was found to be

a factor that did not positively affect the user's desire to switch to the premium version of Duolingo from the free version. This means that even if users are required to watch ads after completing a lesson, it is not considered a significant distraction because the majority of users think that the frequency of ads and the length of ads are still tolerable and considered a normal thing to appear in the free version of an app. In addition, perceived inefficiency was also found to have no positive influence on user switching intentions. This can be caused by users who are already used to the limitations of Duolingo's features. On the other hand, perceived learner value is a pull factor that doesn't influence users to switch to Duolingo premium. This happens because users do not feel a significant difference between the learning process in the premium version and the free version. Then, it was also found that mooring factors, social influence and switching costs have a positive influence on users' intentions to switch users to premium MALL. Meanwhile, inertia as a mooring factor was found to have a negative influence on users' intentions to switch users to the premium version of the MALL. The limitations of this study are that the demographics of the respondents are dominated by female respondents (67 percent), aged 17-25 years (79.5 percent), and students (64.7 percent). Future research can conduct research to analyze user loyalty in using premium MALL.

Duolingo users feel that the premium version of Duolingo only provides a slight difference in features compared to the free version of Duolingo. Duolingo can focus on increasing the differentiation between free and premium services by increasing the benefits of premium rather than just adding limitations to its free services, such as ad restrictions and limited hearts. Duolingo can further develop the curriculum and optimize the features it already has. In further curriculum development, Duolingo can involve more comprehensive language learning materials, such as an in-depth introduction to letters in languages that do not use Latin characters and a more complete explanation of sentence structure (grammar). Duolingo can also provide alternative two-way learning with experts, such as a feedback feature where users can confirm the knowledge they have gained with experts. In optimizing the features that you already have, Duolingo can re-evaluate the use of technology in speaking and listening practice questions in personalized practice features.

In addition, Duolingo can differentiate between users of the free version of Duolingo and the premium version, such as creating a special badge for the premium version, rewards for loyal users of the premium version, or a customized avatar. Duolingo users love the variety of practice questions in the premium version of Duolingo because they don't get bored easily when using Duolingo. Therefore, Duolingo can add a variety of challenge types to its learning, such as by using movies or songs that are popular on the internet to be used in questions. Duolingo can also use different practice question formats, such as using videos. Duolingo users also argue that the surrounding environment, such as family or relatives, can influence a user's desire to subscribe to a premium subscription on Duolingo. In addition, users also show a positive response to features that support collaboration between users such as the friend's quest feature. Therefore, Duolingo can implement a feature that will support interaction between users of the premium version.

**Ethical Declaration:** This study is not required ethics approval.

**AI Statement:** We did not use artificial intelligence tools.

## Acknowledgements

We wish to convey our gratitude to the University of Indonesia for Grant Hibah PUTI Q1, with grant number NKB-136/UN2.RST/HKP.05.00/2024.

## References

- Ain, N., Kaur, K., and Waheed, M., 2016. The influence of learning value on learning management system use. *Information Development*, 32(5), pp. 1306–1321. <https://doi.org/10.1177/0266666915597546>
- Al-Rahmi, A.M., Al-Rahmi, W.M., Alturki, U., Aldraiweesh, A., Almutairy, S., and Al-Adwan, A.S., 2021. Exploring the Factors Affecting Mobile Learning for Sustainability in Higher Education. *Sustainability*, 13(14), pp. 7893-7915. <https://doi.org/10.3390/su13147893>
- Aliakbari, M., and Mardani, M., 2022. Mobile-Assisted Language Learning and Its Effects on Learners' Speaking Development. *Education Research International*, pp. 1–14. <https://doi.org/10.1155/2022/9043326>
- Anggraeni, L., 2023. Indonesia Jadi Pasar Terbesar Kedua Duolingo. [online] Medcom.id. Available at <<https://www.medcom.id/teknologi/news-teknologi/0kpPLZRk-indonesia-jadi-pasar-terbesar-kedua-duolingo>> [Accessed 2 February 2024].
- Anis, W.A.W. and Noor, N.A.M., 2021. The Moderating Effects of Switching Cost On The Relationship Between Satisfaction And Intention To Switch, And Alternative Attractiveness And Intention To Switch In Malaysian Prepaid Mobile Phone Services. *International Journal of Business and Economy*, 3(1), pp. 180-189.

- Arvanitis, P. and Krystalli, P., 2021. Mobile Assisted Language Learning (MALL): Trends from 2010 to 2020 Using Text Analysis Techniques. *European Journal of Education*, 4(1), pp. 13–22. <https://files.eric.ed.gov/fulltext/EJ1336725.pdf>
- Bansal, H.S., 2005. "Migrating" to New Service Providers: Toward a Unifying Framework of Consumers' Switching Behaviors. *Journal of the Academy of Marketing Science*, 33(1), pp. 96–115. <https://doi.org/10.1177/0092070304267928>
- Ceci, L., 2023. Top language learning apps by downloads 2021. [online] Statista. Available at <<https://www.statista.com/statistics/1239522/top-language-learning-apps-downloads/>> [Accessed 2 February 2024].
- Chang, H.H., Fu, C.S. and Jain, H.T., 2016. Modifying UTAUT and innovation diffusion theory to reveal online shopping behavior. *Information Development*, 32(5), pp. 1757–1773. <https://doi.org/10.1177/0266666915623317>
- Chen, Y.-H. and Keng, C.-J., 2019. Utilizing the Push-Pull-Mooring-Habit framework to explore users' intention to switch from offline to online real-person English learning platform. *Internet Research*, 29(1), pp. 167–193. <https://doi.org/10.1108/intr-09-2017-0343>
- Dogra, N., Bakshi, S. and Gupta, A., 2022. Exploring the switching intention of patients to e-health consultations platforms: blending inertia with push–pull–mooring framework. *Journal of Asia Business Studies*, 17(1), pp. 15–37. <https://doi.org/10.1108/jabs-02-2021-0066>
- Dorigo, G. and Tobler, W., 1983. Push-Pull Migration Laws. *Annals of the Association of American Geographers*, 73(1), pp. 1–17. <https://www.jstor.org/stable/2569342>
- Faozi, F.H. and Handayani, P.W., 2023. The Antecedents of Mobile-Assisted Language Learning Applications Continuance Intention. *Electronic Journal of E-Learning*, 21(4), pp. 299–313. <https://doi.org/10.34190/ejel.21.4.2744>
- Foulds, O., Azzopardi, L. and Halvey, M., 2021. Investigating the Influence of Ads on User Search Performance, Behaviour, and Experience during Information Seeking. In the 2021 Conference on Human Information Interaction and Retrieval. United States of America, New York, March 2021. New York: Association for Computing Machinery. <https://doi.org/10.1145/3406522.3446024>
- Hair, J.F., Sarstedt, M., Hopkins, L. and Kuppelwieser, V.G., 2014. Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *European Business Review*, 26(2), pp. 106–121. <https://doi.org/10.1108/eb-10-2013-0128>
- Hamari, J., Hanner, N. and Koivisto, J., 2019. "Why pay premium in freemium services?" A study on perceived value, continued use and purchase intentions in free-to-play games. *International Journal of Information Management*, 51, pp. 102040. <https://doi.org/10.1016/j.ijinfomgt.2019.102040>
- Harja, Y. D., Irawan, M. I. and Ambarwati, R., 2019. Measure The Significance of Learning Value and Trust Factors for Online Learning Technology Acceptance in Indonesia. *IPTEK the Journal for Technology and Science*, 31(2), pp. 188–200. <https://doi.org/10.12962/j20882033.v31i2.5583>
- Hidayat-ur-Rehman, I., Alzahrani, S., Rehman, M.Z. and Akhter, F., 2022. Determining the factors of m-wallets adoption. A twofold SEM-ANN approach. *PLOS ONE*, 17(1), pp. e0262954. <https://doi.org/10.1371/journal.pone.0262954>
- Hsieh, P.-J., 2021. Understanding medical consumers' intentions to switch from cash payment to medical mobile payment: A perspective of technology migration. *Technological Forecasting and Social Change*, 173, pp. 121074. <https://doi.org/10.1016/j.techfore.2021.121074>
- Hu, X., Wang, S., Zhou, R., Gao, L. and Zhu, Z., 2023. Determinants of consumers' intentions to switch to electric vehicles: a perspective of the push–pull–mooring framework. *Journal of Environmental Planning and Management*, 67(14), pp. 1–26. <https://doi.org/10.1080/09640568.2023.2232945>
- Huynh, D., Zuo, L. and Iida, H., 2016. Analyzing Gamification of "Duolingo" with Focus on Its Course Structure. *Game and Learning Alliance*, 10056, pp. 268–277. [https://doi.org/10.1007/978-3-319-50182-6\\_24](https://doi.org/10.1007/978-3-319-50182-6_24)
- Hussain, S., Seet, P.-S., Ryan, M., Iranmanesh, M., Cripps, H. and Salam, A., 2022. Determinants of switching intention in the electricity markets - An integrated structural model approach. *Journal of Retailing and Consumer Services*, 69, pp. 103094. <https://doi.org/10.1016/j.jretconser.2022.103094>
- Kim, S., Choi, M. J. and Choi, J.S., 2019. Empirical Study on the Factors Affecting Individuals' Switching Intention to Augmented/Virtual Reality Content Services Based on Push-Pull-Mooring Theory. *Information*, 11(1), pp. 25–46. <https://doi.org/10.3390/info11010025>
- Kock, N. (2015). One-Tailed or Two-Tailed P Values in PLS-SEM? *International Journal of E-Collaboration*, 11(2), pp. 1–7. <https://doi.org/10.4018/ijec.2015040101>
- Kuo, R.-Z., 2020. Why do people switch mobile payment service platforms? An empirical study in Taiwan. *Technology in Society*, 62, pp. 101312. <https://doi.org/10.1016/j.techsoc.2020.101312>
- Li, H., Edwards, S.M. and Lee, J.-H., 2002. Measuring the Intrusiveness of Advertisements: Scale Development and Validation. *Journal of Advertising*, 31(2), pp. 37–47. <https://doi.org/10.1080/00913367.2002.10673665>
- Magsayo, R. T., 2022. Mobile learning adoption continuance: role of locus of control on its determinants. *Interactive Technology and Smart Education*, 20(2), pp. 177–208. <https://doi.org/10.1108/itse-10-2021-0191>
- Monoarfa, T.A., Sumarwan, U., Suroso, A.I. and Wulandari, R., 2023. Switch or Stay? Applying a Push–Pull–Mooring Framework to Evaluate Behavior in E-Grocery Shopping. *Sustainability*, 15(7), pp. 6018. <https://doi.org/10.3390/su15076018>
- Nayak, B., Bhattacharyya, S.S., Goswami, S. and Thakre, S., 2021. Adoption of online education channel during the COVID-19 pandemic and associated economic lockdown: an empirical study from push–pull–mooring framework. *Journal of Computers in Education*, 9, pp. 1–23. <https://doi.org/10.1007/s40692-021-00193-w>

- Pozón-López, I., Higuera-Castillo, E., Muñoz-Leiva, F., & Liébana-Cabanillas, F. J. (2020). Perceived user satisfaction and intention to use massive open online courses (MOOCs). *Journal of Computing in Higher Education*, 33, pp. 85-120. <https://doi.org/10.1007/s12528-020-09257-9>
- Purnama, B.E., 2023. Duolingo Luncurkan Fitur Premium Super Duolingo di Seluruh Wilayah Asia Tenggara. [online] Mediaindonesia.com. Available at <<https://mediaindonesia.com/teknologi/578527/duolingo-luncurkan-fitur-premium-super-duolingo-di-seluruh-wilayah-asia-tenggara>> [Accessed 2 February 2024].
- Qashou, A., 2020. Influencing factors in M-learning adoption in higher education. *Education and Information Technologies*, 26, pp. 1755–1785. <https://doi.org/10.1007/s10639-020-10323-z>
- Samuelson, W. and Zeckhauser, R., 1988. Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1), pp. 7–59. <https://doi.org/10.1007/bf00055564>
- Sarkar, A., Sarkar, J. G. and Sreejesh, S., 2021. Managing customers’ undesirable responses towards hospitality service brands during service failure: The moderating role of other customer perception. *International Journal of Hospitality Management*, 94, pp. 102873. <https://doi.org/10.1016/j.ijhm.2021.102873>
- Sekarputri, J.A., Fitriani, W.R., Hidayanto, A.N. and Kurnia, S., 2024. The roles of privacy, security, and dissatisfaction in affecting switching intention on messenger applications. *Multimedia Tools and Applications*, 83, pp. 45625–45651. <https://doi.org/10.1007/s11042-023-17466-4>
- Sulyati, E., 2020. Learning Foreign Language: Between Globalization and Hegemony in Indonesia. *International Journal of Science and Society*, 1(1), pp. 81–90. <https://doi.org/10.54783/ijssoc.v1i1.50>
- Sun, Y., Liu, D., Chen, S., Wu, X., Shen, X.-L. and Zhang, X., 2017. Understanding users’ switching behavior of mobile instant messaging applications: An empirical study from the perspective of push-pull-mooring framework. *Computers in Human Behavior*, 75, pp. 727–738. <https://doi.org/10.1016/j.chb.2017.06.014>
- Tsai, L.L., 2022. A deeper understanding of switching intention and the perceptions of non-subscribers. *Information Technology & People*, 36(2), pp. 785-807. <https://doi.org/10.1108/itp-04-2021-0255>
- Toto, G.A. and Limone, P., 2019. Contemporary Trends in Studies on Mobile Learning of Foreign Languages: A Meta-Analysis. *International Journal of Engineering Education*, 1(2), pp. 85–90. <https://doi.org/10.14710/ijee.1.2.85-90>
- Wang, G. and Shin, C., 2022. Influencing Factors of Usage Intention of Metaverse Education Application Platform: Empirical Evidence Based on PPM and TAM Models. *Sustainability*, 14(24), pp. 17037. <https://doi.org/10.3390/su142417037>
- Wang, Y., Douglas, M. and Hazen, B., 2021. Diffusion of public bicycle systems: Investigating influences of users’ perceived risk and switching intention. *Transportation Research Part A: Policy and Practice*, 143, pp. 1–13. <https://doi.org/10.1016/j.tra.2020.11.002>
- Wei, X., Ko, I. and Pearce, A., 2021. Does Perceived Advertising Value Alleviate Advertising Avoidance in Mobile Social Media? Exploring Its Moderated Mediation Effects. *Sustainability*, 14(1), pp. 253. <https://doi.org/10.3390/su14010253>
- Wu, K., Vassileva, J. and Zhao, Y., 2017. Understanding users’ intention to switch personal cloud storage services: Evidence from the Chinese market. *Computers in Human Behavior*, 68, pp. 300–314. <https://doi.org/10.1016/j.chb.2016.11.039>
- Ye, D., Liu, F., Cho, D. and Jia, Z., 2022. Investigating switching intention of e-commerce live streaming users. *Heliyon*, 8(10), pp. e11145. <https://doi.org/10.1016/j.heliyon.2022.e11145>
- Yoon, C. and Lim, D., 2021. Customers’ Intentions to Switch to Internet-Only Banks: Perspective of the Push-Pull-Mooring Model. *Sustainability*, 13(14), pp. 8062. <https://doi.org/10.3390/su13148062>
- Yuen, K.F., Ng, W.H. and Wang, X., 2023. Switching intention in the online crowdsourced delivery environment: The influence of a platform’s technological characteristics and relational bonding strategies. *Technology in Society*, 72, pp. 102167. <https://doi.org/10.1016/j.techsoc.2022.102167>
- Yulianto, H.C. and Oroh, A.N.H., 2021. The Effects of Social Value, Value for Money, App Rating, and Enjoyment on the Intention to Purchase the Premium Service of the Spotify App. *KnE Social Sciences*, 5(5), pp. 266–281. <https://doi.org/10.18502/kss.v5i5.8815>
- Zacharis, G. and Nikolopoulou, K., 2022. Factors predicting University students’ behavioral intention to use eLearning platforms in the post-pandemic normal: an UTAUT2 approach with “Learning Value.” *Education and Information Technologies*, 27, pp. 12065–12082. <https://doi.org/10.1007/s10639-022-11116-2>
- Zakiy, M. and Haryanto, M., 2021. Linking of Service Quality and Switching Costs with Turnover Intentions: The Mediating Role of Customer Loyalty. In the 4th International Conference on Sustainable Innovation 2020-Accounting and Management (ICoSIAMS 2020), pp. 29–37. New York: Atlantis Press. <https://doi.org/10.2991/aer.k.210121.005>
- Zhu, Z., Peng, Z. and Yang, K., 2023. Utilizing the push-pull-mooring framework to explore university teachers’ intention to switch from traditional classrooms to smart classrooms in China. *Journal of Education and Training*, 65(3), pp. 470–491. <https://doi.org/10.1108/et-12-2021-0461>

## Appendix A: Measurement Items in the Questionnaire

| Code | Measurement Items                                                                                              | References  |
|------|----------------------------------------------------------------------------------------------------------------|-------------|
| PI1  | I feel that the ads in the free version of the Duolingo app were distracting me when working on the questions. | Tsai (2022) |

| Code | Measurement Items                                                                                                                                                                                                                                 | References                 |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|
| PI2  | I find that the ads in the free version of the Duolingo app are annoying. (e.g., the duration is too long).                                                                                                                                       |                            |
| PI3  | I feel that the ads in the free version of the Duolingo app are not to my liking.                                                                                                                                                                 |                            |
| PI4  | I am annoyed when I have to watch ads on the free version of Duolingo.                                                                                                                                                                            |                            |
| DS1  | I have trouble doing the lessons I wanted on the free version of Duolingo.                                                                                                                                                                        | Ye et al. (2022)           |
| DS2  | I feel that the features in the free version of Duolingo have not met my expectations.                                                                                                                                                            |                            |
| DS3  | I feel that the features in the free version of Duolingo do not meet my needs.                                                                                                                                                                    |                            |
| DS4  | I am not satisfied with my overall experience while using the free version of Duolingo's features.                                                                                                                                                | Sekarputri et al. (2024)   |
| PF1  | I feel like I wasted a lot of time learning a foreign language on the free version of Duolingo (e.g., ads take up my learning time).                                                                                                              | Kim, Choi and Choi (2019)  |
| PF2  | I feel that the process of learning a foreign language in the free version of Duolingo is inefficient (e.g., the limitation of hearts hinders the efficiency of my learning process).                                                             | Zhu, Peng and Yang (2023)  |
| PF3  | I find the learning process of the free version of Duolingo unproductive (e.g., the restriction on the features, such as Legendary level, made it difficult for me to be productive).                                                             | Kim, Choi and Choi (2023)  |
| PF4  | I don't get the results I want in learning a foreign language on the free version of Duolingo due to feature limitations.                                                                                                                         | Zhu, Peng and Yang (2023)  |
| PE1  | I believe the premium version of the Duolingo app has interesting features.                                                                                                                                                                       | Lisana (2022)              |
| PE2  | I'm very excited to use the features on the premium version of the Duolingo app.                                                                                                                                                                  |                            |
| PE3  | I believe the features in the premium version of Duolingo can make my learning process more enjoyable.                                                                                                                                            |                            |
| PE4  | I believe I will love the overall learning experience due to the features offered by the premium version of Duolingo.                                                                                                                             | Lisana and Suciadi (2021)  |
| PU1  | I feel that the features in the premium version of Duolingo can improve my learning performance. (e.g., the personalized lesson feature improves my performance)                                                                                  | Faozi and Handayani (2023) |
| PU2  | I feel that the features in the premium version of Duolingo can increase my learning productivity. (e.g., having no restrictions, such as on Legendary level and heart limitations, helps me be more productive)                                  |                            |
| PU3  | I feel that the features on the premium version of Duolingo can effectively help me learn a foreign language. (e.g., having an infinite heart makes my experience more effective because I don't have to wait for the system to refill my hearts) |                            |

| Code | Measurement Items                                                                                                           | References                      |
|------|-----------------------------------------------------------------------------------------------------------------------------|---------------------------------|
| PU4  | I feel that the features in the premium version of Duolingo will be useful to me.                                           |                                 |
| PV1  | I can be more proactive in learning a foreign language by using the premium version of Duolingo.                            | Magsayo (2022)                  |
| PV2  | I am able to increase my interest in learning by using the premium version of Duolingo                                      |                                 |
| PV3  | I am able to achieve my academic goals in learning languages by using the premium version of Duolingo.                      |                                 |
| PV4  | I am able to learn a foreign language more actively using the premium version of Duolingo.                                  |                                 |
| TR1  | I wanted to explore the premium version of Duolingo before I subscribed.                                                    | Hsieh (2021)                    |
| TR2  | I wanted to use the premium version of Duolingo features during the free trial period before I subscribed.                  |                                 |
| TR3  | I want to try the premium version of Duolingo in enough time to experience the benefits offered.                            |                                 |
| TR4  | I consider the opportunity for a trial important before deciding to buy the premium version of Duolingo.                    | Chang et al. (2016)             |
| SC1  | I find the procedure to start subscribing to the premium version of Duolingo is very complicated.                           | Hsieh et al. (2021)             |
| SC2  | I felt it took a lot of effort and time to switch to the premium version of Duolingo                                        | Ye et al. (2023)                |
| SC3  | I feel that I have to spend a lot of money to switch to the premium version of Duolingo.                                    |                                 |
| SC4  | I don't believe that I can get a better foreign language learning experience if I switch to the premium version of Duolingo | Yuen, Ng, and Wang (2023)       |
| SI1  | My relatives or friends invite me to subscribe to the premium version of Duolingo.                                          | Ye et al. (2022)                |
| SI2  | Many of my relatives or friends recommend the premium version of Duolingo to me.                                            |                                 |
| SI3  | Many of my relatives or friends are dissatisfied with the free version of Duolingo.                                         |                                 |
| SI4  | The opinion of my relatives or friends greatly influenced my decision to use the premium version of Duolingo.               | Hsieh et al. (2021)             |
| IN1  | I never thought of switching to the premium version of Duolingo.                                                            | Wang, Douglas, and Hazen (2021) |
| IN2  | I wouldn't switch to the premium version of Duolingo unless I'm very dissatisfied with the free version of Duolingo.        | Hu et al. (2023)                |

| Code | Measurement Items                                                                                        | References       |
|------|----------------------------------------------------------------------------------------------------------|------------------|
| IN3  | I would still choose the free version of Duolingo because it has become a habit.                         |                  |
| IN4  | I feel that the free version of Duolingo is enough to meet my needs.                                     |                  |
| SW1  | I would consider switching from the free version of Duolingo to the premium version of the Duolingo app. | Hu et al. (2022) |
| SW2  | I will most likely switch to the premium version of Duolingo.                                            |                  |
| SW3  | I believe I will be switching to the premium version of Duolingo.                                        |                  |
| SW4  | I would choose to use the premium version of Duolingo over the free version of Duolingo.                 |                  |

### Appendix B: Sample Interview Questions

1. How do the ads that appear affect your learning experience on the free version of the Duolingo app?
2. What do you think about the impact of these feature limitations on your learning experience?
3. How do you feel when using the premium version of Duolingo?
4. What are your expectations for the additional benefits you should be able to experience in using the premium version of Duolingo?
5. What do you think about the learning value you can get by using the premium version of Duolingo?
6. How do you think the value of learning can influence your desire to subscribe to the premium version of Duolingo?
7. How was your experience when using the premium version of Duolingo free trial?
8. How do you rate the Premium version of Duolingo's sign-up process? (Example: choosing a plan, choosing a payment method)
9. How does the presence of people around you who use the premium version of Duolingo affect your interest in using the premium version? How
10. What do you think about how your experience using the premium version of Duolingo will keep you subscribed to the premium version of Duolingo?

### Appendix C: AVE, CA and CR Values

| Variable | AVE   | CA    | CR    |
|----------|-------|-------|-------|
| DS       | 0.741 | 0.895 | 0.895 |
| IN       | 0.673 | 0.889 | 0.889 |
| PE       | 0.733 | 0.916 | 0.917 |
| PF       | 0.691 | 0.898 | 0.899 |
| PI       | 0.672 | 0.889 | 0.892 |
| PU       | 0.788 | 0.937 | 0.937 |
| PV       | 0.791 | 0.938 | 0.938 |
| SC       | 0.624 | 0.827 | 0.830 |
| THE      | 0.717 | 0.908 | 0.913 |
| SW       | 0.839 | 0.940 | 0.941 |
| TR       | 0.708 | 0.904 | 0.907 |

## Appendix D: Goodness of Fit Results

| GoF Criterion  | Cut-Off Value | Result | Information     |
|----------------|---------------|--------|-----------------|
| <b>CMIN/df</b> | $\leq 3$      | 2.136  | <i>Good Fit</i> |
| <b>RMSEA</b>   | $\leq 0.08$   | 0.034  | <i>Good Fit</i> |
| <b>NFI</b>     | $\geq 0.9$    | 0.959  | <i>Good Fit</i> |
| <b>CFI</b>     | $\geq 0.9$    | 0.978  | <i>Good Fit</i> |
| <b>GFI</b>     | $\geq 0.9$    | 0.932  | <i>Good Fit</i> |
| <b>TLI</b>     | $\geq 0.9$    | 0.974  | <i>Good Fit</i> |

# Beyond Face Recognition: A Multi-Layered Approach to Academic Integrity in Online Exams

Aivar Sakhipov<sup>1,2</sup>, Islam Omirzak<sup>1</sup> and Alexey Fedenko<sup>1</sup>

<sup>1</sup>Department of Computer Engineering, Astana IT University, Astana, Kazakhstan

<sup>2</sup>Department of Computer Science, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

[aivar.sakhipov@astanait.edu.kz](mailto:aivar.sakhipov@astanait.edu.kz)

[islamomirzak88@gmail.com](mailto:islamomirzak88@gmail.com) (corresponding author)

[supp.fed@gmail.com](mailto:supp.fed@gmail.com)

<https://doi.org/10.34190/ejel.23.1.3896>

An open access article under [CC Attribution 4.0](#)

**Abstract:** Ensuring academic integrity in online assessments is crucial for upholding fairness and credibility, especially with the widespread adoption of remote learning. This research addresses key vulnerabilities in preventing cheating and unauthorized collaboration, common in online assessments lacking direct supervision. To address these challenges, an intelligent proctoring system was developed and tested on BlockchainStudy.kz — an educational platform that offers online courses and issues blockchain-based certificates. This system establishes a controlled examination environment through facial recognition, user activity monitoring, and browser behavior tracking, effectively deterring dishonest practices. The study adopted a phased methodology, starting with pilot testing for feasibility, followed by large-scale deployment to assess scalability and effectiveness. The approach combines machine learning-based facial recognition for identity verification, user action logging, and browser monitoring to detect suspicious behaviors indicative of academic dishonesty. Findings demonstrated a marked decrease in cheating incidents, enhanced examination credibility, and improved perceptions of fairness among both students and instructors. By encouraging accountability, the system fostered a culture of honesty within the online education environment. Ethical concerns regarding privacy were addressed through robust safeguards in compliance with General Data Protection Regulation (GDPR), building student trust in the proctoring system. This research contributes to the field of e-learning by providing a scalable, effective solution for maintaining academic integrity in online assessments. It facilitates informed decision-making for educators, reduces dishonest behavior, and promotes a culture of integrity within digital education. Overall, this work enriches the body of e-learning knowledge by presenting a practical, adaptable strategy for overcoming the complex challenges of academic integrity in remote learning environments.

**Keywords:** Academic integrity, Intelligent proctoring, Machine learning, Blockchain, Online education, Cheating prevention

## 1. Introduction

The rapid expansion of online education has revolutionized the accessibility of learning, offering flexibility and inclusivity. However, it has also led to an alarming rise in academic dishonesty, posing significant challenges to assessment integrity (Dawson, 2021). Ensuring that online assessments are conducted fairly and honestly is increasingly difficult, given the remote nature of such exams and the lack of direct supervision (Bretag et al., 2019). The shift to online learning, accelerated by the COVID-19 pandemic, has exacerbated these challenges, leading to a rise in incidents of academic dishonesty, such as cheating and unauthorized collaboration (Lancaster and Cotarlan, 2021; Eaton, 2020).

Recent studies confirm a troubling surge in academic dishonesty in online assessments, particularly following the widespread shift to remote learning. A systematic review by Newton and Essex (2024) found that self-reported cheating rates in online exams increased from 29.9% before the pandemic to 54.7% during the pandemic, highlighting the growing prevalence of misconduct in remote education settings (Newton and Essex, 2024). Similarly, multiple universities in the United States, including the University of Missouri, North Carolina State University, Georgia Tech, and Boston University, reported substantial rises in academic dishonesty cases in online courses (Brandeis University, 2020). A study conducted in Pakistan revealed that 60% of students admitted to frequently cheating during online exams, with an additional 30% acknowledging they had cheated at least once (Malik et al., 2023). These findings illustrate the increasing sophistication of cheating strategies, such as the use of encrypted messaging apps and contract cheating services, reinforcing the urgent need for more effective and technologically advanced measures to uphold academic integrity in digital learning environments.

A viable solution to this challenge is implementing intelligent proctoring systems. These systems are designed to create a controlled examination environment by employing tools like facial recognition, behavior analysis,

and browser activity tracking (Nigam et al., 2021). Unlike traditional proctoring approaches, which often rely solely on browser lockdowns or restricted access to external resources, advanced proctoring integrates multi-faceted monitoring techniques, including face recognition, user activity logging, and automated detection of suspicious behavior, ensuring that examinations remain secure and credible.

Ensuring the integrity of online assessments is essential for maintaining the credibility of educational qualifications, as academic dishonesty undermines both individual learning outcomes and institutional reputation. (Bretag et al., 2018). Studies indicate that students are more likely to cheat when they perceive few consequences or a low risk of detection (Gamage et al., 2023). Thus, the deployment of such systems provides a deterrent against cheating, encouraging students to prepare adequately for exams and adhere to academic ethics (Dawson, 2021).

Despite the increasing adoption of online proctoring, concerns persist regarding its effectiveness in reducing cheating, ensuring fairness, and mitigating student stress. This study investigates the extent to which advanced automated systems can enhance academic integrity while maintaining a balanced and ethical approach.

To explore this, the study aims to answer the following research question: *How do intelligent proctoring systems impact cheating prevention, fairness, and student stress in online assessments?*

To answer this question, this paper examines the implementation and effectiveness of an advanced proctoring system on BlockchainStudy.kz, an educational platform that offers online courses and blockchain-based certificates. The study assesses the system's impact on reducing academic dishonesty, enhancing exam credibility, and addressing ethical concerns related to privacy and student stress levels. This research contributes to digital education security discourse by evaluating the effectiveness of multi-layered proctoring tools in preventing academic misconduct.

## 2. Literature Review

Maintaining academic integrity in online assessments presents significant challenges for educational institutions. Early remote proctoring systems primarily utilized basic measures such as browser lockdowns and time-restricted access to deter cheating (Tiong and Lee, 2021). However, as students developed methods to bypass these controls, the need for more advanced technologies became evident. Bilen and Matros (2021) observed that during the COVID-19 pandemic, students adapted quickly to circumvent basic proctoring mechanisms, which highlighted the urgency of implementing sophisticated monitoring technologies. The increasing sophistication of cheating tactics has driven educational institutions to explore newer and more effective methods for ensuring academic honesty. Simple methods of preventing cheating, such as limiting access to external resources or enforcing strict time constraints, were often insufficient against the evolving strategies used by tech-savvy students. This necessitated the integration of more advanced, multi-layered technologies to uphold the credibility of assessments.

The debate between traditional human invigilation and automated proctoring remains central to ensuring exam integrity. As shown in Table 1, these two approaches differ significantly in their monitoring methods, contextual judgment, cost, scalability, ethical considerations, and detection accuracy.

**Table 1: Comparison of Traditional Human Invigilation and Automated Proctoring**

| Feature                    | Human Invigilation                                    | Automated Proctoring                                      |
|----------------------------|-------------------------------------------------------|-----------------------------------------------------------|
| <b>Monitoring Method</b>   | In-person supervision by proctors                     | Software-based remote monitoring                          |
| <b>Contextual Judgment</b> | High; proctors can interpret nuanced behaviors        | Limited; relies on predefined algorithms                  |
| <b>Cost</b>                | High; requires personnel and physical space           | Lower; reduces need for physical resources                |
| <b>Scalability</b>         | Limited; constrained by available proctors and venues | High; can accommodate large numbers of examinees remotely |
| <b>Ethical Concerns</b>    | Lower; direct human oversight                         | Higher; concerns about data privacy and algorithmic bias  |
| <b>Detection Accuracy</b>  | Subjective; depends on proctor vigilance              | Objective; depends on algorithm effectiveness             |

Human invigilation has historically been the most effective method of ensuring academic integrity, as proctors can assess examinee behavior in real-time, intervene when necessary, and apply contextual judgment to

distinguish between unintentional actions and deliberate misconduct (Muzaffar et al., 2020). However, human-supervised exams require significant logistical and financial resources, making them impractical for large-scale online education. Automated proctoring systems, on the other hand, provide scalability and cost-efficiency but lack the nuanced decision-making abilities of human invigilators. These systems rely on rule-based flagging mechanisms that may incorrectly classify harmless behaviors, such as looking away from the screen or adjusting one's seating position, as suspicious activities (Balash et al., 2021).

The advent of machine learning (ML)-based proctoring tools has significantly enhanced the credibility of online assessments by introducing advanced detection capabilities. These systems employ techniques like facial recognition and behavioral analysis to monitor students in real-time, effectively detecting identity discrepancies and suspicious activities with greater accuracy. For instance, a study by Tiong and Lee (2021) introduced an e-cheating intelligence agent that utilizes ML to detect online cheating through IP and behavioral analysis. However, while these ML-based systems mark a clear advancement over basic rule-based methods, they are not without limitations. Concerns about algorithmic biases, data privacy, and potential false positives remain, suggesting that relying solely on ML may not fully address the complex dynamics of cheating. In response, some institutions have begun to explore complementary non-ML approaches—such as biometric authentication, continuous identity verification, and rule-based anomaly detection—to develop more robust and context-sensitive proctoring solutions.

Evaluations of intelligent proctoring systems across various educational contexts highlight both their promise and their challenges. Liu et al. (2024) developed a framework named CHEESE, which applies multiple instance learning to detect and localize cheating behaviors in online exams, achieving a frame-level Area Under the Curve (AUC) score of 87.58% on the Online Exam Proctoring dataset. This method effectively identifies suspicious activities based on data patterns but lacks real-time contextual awareness. In contrast, Moyo et al. (2023) proposed a video-based detector using OpenPose, which analyzes student posture and movement to identify deviations from normal exam conduct. Liu's model offers high scalability but risks misclassifying non-malicious behaviors, whereas Moyo's system improves contextual interpretation but faces computational constraints. A hybrid approach that integrates scalable anomaly detection with movement-based behavioral analysis could enhance both accuracy and fairness. These findings underscore the need for integrative models that combine automated detection with human oversight, ensuring adaptability across diverse educational environments.

Ethical considerations are crucial in proctoring system deployment. Coghlan, Miller and Paterson (2021) highlight biases in AI-based moderation, particularly in facial recognition, which may disproportionately affect underrepresented demographics, leading to false positives and undue stress. To ensure fairness, ongoing research focuses on refining algorithms for greater inclusivity and equity, preventing unintended disadvantages while upholding academic integrity.

Global studies further emphasize the diversity of challenges in implementing online proctoring systems. Raman et al. (2021), for instance, applied the diffusion of innovation theory to examine the adoption of Online Proctored Examinations (OPE) during the COVID-19 pandemic, revealing that factors such as relative advantage, compatibility, and ease of use positively influenced student acceptance. Using Aspect Level Sentiment Analysis, the researchers found that 55% of students held a positive attitude towards OPE, viewing it as advantageous and easy to use despite some challenges. This work adds an important perspective on the integration of proctoring technologies in higher education, especially during disruptive global events like the pandemic.

While existing literature demonstrates that both ML-based and automated proctoring systems can enhance exam integrity, significant gaps remain in integrating diverse monitoring methods, addressing ethical concerns such as bias in facial recognition and privacy issues, and adapting to varied regional contexts. Facial recognition struggles with demographic fairness (Coghlan, Miller and Paterson, 2021), rule-based anomaly detection often misclassifies normal behavior as suspicious (Balash et al., 2021), and many proctoring systems lack adaptability to different regulatory and technological environments (Raman et al., 2021). To address these issues, our study introduces a hybrid intelligent proctoring system on BlockchainStudy.kz that combines facial recognition, behavioral monitoring, biometric authentication, continuous identity verification, and rule-based anomaly detection—reinforced by human oversight. This multi-layered approach improves detection accuracy, enhances scalability by reducing reliance on manual supervision, and ensures fairness through bias-aware verification and strict privacy safeguards. Moreover, this study relies solely on anonymized, computer-generated data—without collecting personal user information—thus eliminating the need for additional ethical approvals.

In conclusion, leveraging advanced technologies like ML is crucial for upholding academic integrity in online assessments. Nonetheless, persistent challenges—including privacy concerns, algorithmic biases, and regional

variability—underscore the need for a more balanced approach. Future research should prioritize the development of hybrid proctoring models that integrate diverse technological methods with robust ethical safeguards. Furthermore, collaboration among technologists, educators, and policymakers is critical to establishing comprehensive standards and best practices, ensuring that proctoring systems not only deter academic dishonesty but also promote fairness and trust across varied educational contexts.

### 3. Materials and Methods

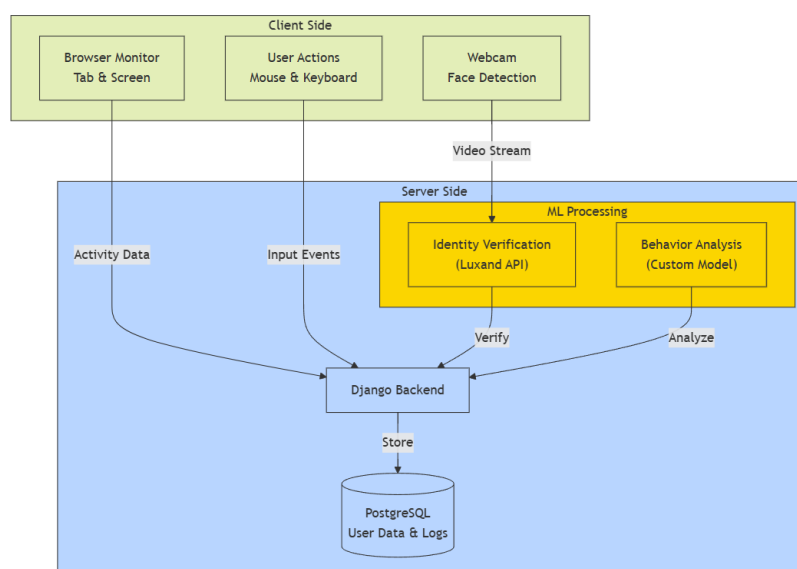
#### 3.1 System Design and Architecture

The advanced proctoring system incorporates multiple technologies to maintain exam integrity. Its architecture combines facial recognition, user activity logging, and behavioral monitoring to detect cheating or suspicious behavior. The system's core feature is identity verification, achieved through facial recognition, which compares a pre-uploaded profile photo with a live video stream captured during the exam. To balance accuracy and privacy, the system uses the Luxand Face Recognition API. Alternative solutions, such as OpenCV-based models or AWS Rekognition, were considered, but Luxand was selected due to its high accuracy, ease of integration, and compliance with GDPR regulations. Unlike OpenCV, which requires extensive training datasets and computational power, Luxand provides a pre-trained model that reduces the dependency on local biometric data storage, thereby lowering security risks.

To prevent cheating through tab switching or external browsing, the system enforces full-screen mode. JavaScript event listeners and the HTML5 Page Visibility API monitor for attempts to switch tabs or minimize the browser, flagging such actions for review. Mouse movements, keyboard inputs, and screen interactions are continuously captured and sent to the server via WebSocket for real-time processing.

For secure data handling, the system uses PostgreSQL as the database, chosen for its ACID compliance, advanced indexing capabilities, and encryption features. Alternatives such as MySQL and MongoDB were evaluated, but PostgreSQL's ability to handle high-volume, real-time event logs with strict integrity constraints made it the optimal choice. The database securely stores exam logs, including timestamps for every user action, such as key presses, mouse clicks, and screen state changes. These logs are indexed for quick retrieval and analyzed to detect anomalous behavior indicative of cheating.

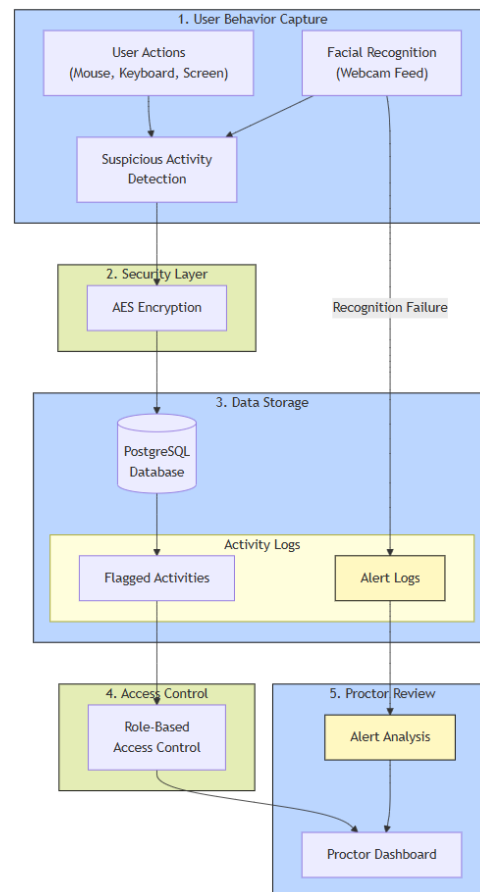
Figure 1 illustrates the multi-modal architecture of the system, where client-side monitoring captures user actions and webcam data, and server-side machine learning processes verify identity and analyze behavior. The backend was built using Django, which was selected due to its built-in security mechanisms, scalability, and support for role-based access control (RBAC). While Flask and Node.js were considered, Django's pre-configured authentication modules, secure session handling, and ORM-based database management provided a robust, secure, and efficient backend solution. The backend manages data flow between the client and the PostgreSQL database, ensuring secure storage of user data and exam logs. This integrated approach provides a robust proctoring solution that deters cheating and maintains academic integrity.



**Figure 1: Machine Learning-Enhanced Academic Integrity System: Multi-Modal Architecture**

It is important that the intelligent system has the ability to record user actions in real time to ensure the integrity of the exam. It records mouse movements, keyboard inputs, facial recognition data, and screen state changes, securely storing them in a PostgreSQL database with timestamps. Suspicious behaviors—such as frequent glances away, tab-switching, or unauthorized individuals in the webcam feed—are flagged for review and stored separately for detailed post-exam analysis.

To maintain security, the logged data is encrypted using Advanced Encryption Standard (AES) during both transmission and storage. Proctors can access timestamped logs and flagged events after the exam for review, allowing them to verify the legitimacy of student behavior. The system's architecture enables efficient indexing and retrieval of logs, allowing quick identification of suspicious activities. This logging process ensures secure monitoring, analysis, and storage of relevant data while maintaining the ability to review and audit user activity. Figure 2 illustrates the data flow from real-time monitoring to secure storage and post-exam review.



**Figure 2: Data Logging and Monitoring Architecture of the Intelligent Proctoring System**

### 3.2 Development Process and Prototyping

The development of the proctoring system followed an iterative, three-phase approach: prototyping, pilot testing, and large-scale deployment. This structured process was chosen to ensure technical feasibility, real-world applicability, and scalability, aligning with the study's goal of developing a reliable, privacy-compliant, and adaptable proctoring solution.

The prototyping phase was essential for assessing technical feasibility and refining system performance across various devices and environments. Early testing revealed challenges such as facial recognition inconsistencies due to low-resolution cameras and suboptimal lighting, leading to higher false-positive rates. To enhance accuracy, image pre-processing techniques, including histogram equalization and adaptive thresholding, were implemented, improving system robustness in diverse conditions.

The pilot testing phase evaluated the system's effectiveness in real-world conditions with a controlled group of students on BlockchainStudy.kz. This phase focused on detecting suspicious behaviors while ensuring a seamless

user experience. The insights gained allowed for adjustments, such as refining detection thresholds and reducing false positives, to enhance system stability and usability.

The large-scale deployment phase validated the system's scalability, ensuring it could maintain detection accuracy and compliance with ethical and privacy standards under high user loads. This phase demonstrated the system's adaptability to different technological infrastructures and institutional policies, confirming its practicality for widespread implementation in online education.

By structuring development in these phases, the study ensured continuous refinement, ethical integrity, and reliability, aligning with the broader objective of creating a proctoring system that is both effective in preventing academic dishonesty and respectful of user privacy.

### **3.3 Pilot Testing and Deployment**

The system underwent extensive pilot testing to assess its effectiveness in real-world online exam conditions. The testing took place on [blockchainstudy.kz](https://blockchainstudy.kz), an online educational platform offering courses and certification in blockchain technology, which was developed and launched by our team. The platform already has more than 800 active users and supported a wide range of courses and exam functionalities, making it an ideal environment for launching the intelligent proctoring system.

The initial pilot phase involved 66 students who were required to take an exam while their activities were monitored by the system. The main objective of the pilot was to evaluate the system's detection accuracy, false positive rates, and system stability. The system flagged suspicious behaviors such as tab-switching, unauthorized individuals appearing in the webcam feed, or prolonged distractions from the screen. In addition to monitoring behavior, the system's performance was assessed under varying internet speeds, device specifications, and camera qualities to ensure robust functionality across different environments. System load tests were conducted to measure performance under concurrent user sessions, verifying the scalability of real-time processing.

During the pilot phase, several important insights into the system's capabilities were gained. The proctoring system was successful in detecting irregular behaviors, such as students looking away from the screen for extended periods or the presence of additional individuals in the webcam feed. However, initial testing revealed a high rate of false positives, particularly for natural head movements and brief distractions. To mitigate this, anomaly detection techniques were implemented to refine classification models, significantly reducing incorrect flags.

After fine-tuning the detection parameters, the system was scaled up for broader use on the platform. A total of 770 students participated in the system's deployment across various courses on [blockchainstudy.kz](https://blockchainstudy.kz). These students accessed the exams using a range of operating systems, which tested the system's ability to handle a larger volume of concurrent users. The system demonstrated its capacity to maintain stable performance and efficiently monitor student activities in real time, with accurate tracking of behaviors and generation of event logs for post-exam analysis.

The successful scaling of the system within a single platform not only validated the system's robustness in handling large numbers of users but also confirmed its effectiveness in detecting and recording suspicious behavior. The feedback from this phase provided valuable insights for further refining the system before its potential expansion to other educational platforms or institutions.

### **3.4 Data Collection and Security**

The advanced proctoring system implements robust security measures to comply with data protection regulations, including GDPR. Sensitive data such as student identification, exam logs, and facial recognition data are encrypted using AES both during transmission and at rest. Secure communication protocols, like HTTPS (Hypertext Transfer Protocol Secure), are used to protect data during exams. Role-based access control (RBAC) restricts access to sensitive data based on user roles, ensuring that only authorized personnel can view it. Regular security audits are conducted to identify vulnerabilities and address them promptly.

Data collected during exams includes user activity logs, facial recognition data, and tab-switch events. This data is securely stored in an encrypted PostgreSQL database, with only suspicious actions (e.g., prolonged looking away or unauthorized tab-switching) being stored, reducing the exposure of personal information. Proctors review only flagged activities, and anonymization techniques are used to protect personally identifiable information during post-exam analysis.

### 3.5 Ethical Considerations

Ethical considerations were a fundamental aspect of the proctoring system's design, particularly regarding the use of facial recognition and behavioral monitoring. The system was developed in strict compliance with General Data Protection Regulation (GDPR) and other relevant data protection policies, ensuring that user privacy was preserved at all stages.

To participate in online exams, students voluntarily agreed to the platform's proctoring policies, which outlined the use of facial recognition for identity verification, user activity monitoring, and browser behavior tracking. No additional data collection procedures were introduced beyond what was necessary for maintaining academic integrity. All recorded data remained within the BlockchainStudy.kz platform's existing framework and was never used for experimental interventions, external research, or shared with third parties.

To safeguard personal information, no personally identifiable data were included in the research analysis. The study relied exclusively on fully anonymized and aggregated data, ensuring that individual students could not be identified at any stage. Only system-generated logs of user activity were analyzed, focusing on general statistical patterns such as flagged exam violations and system performance metrics.

Strict data retention policies were implemented, with all proctoring logs stored only for as long as necessary to verify exam integrity. Any flagged events were reviewed solely within the system's operational framework, and all data were automatically deleted after the designated retention period.

Because this study did not involve direct human subject research, external interventions, or the collection of identifiable user data, no formal ethics approval was required. The study adhered to best practices in ethical educational research, prioritizing student privacy, data security, and transparency in all aspects of system design and implementation.

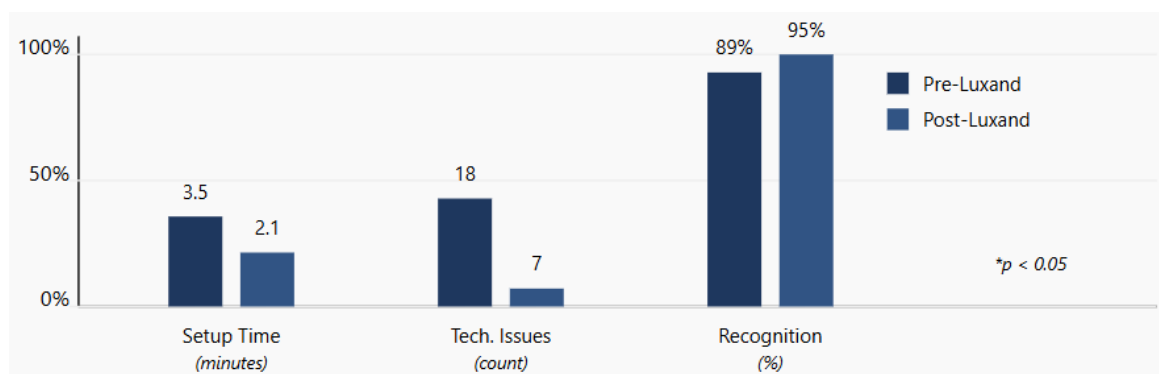
## 4. Results

### 4.1 Pilot Study Results

*Phase 1: Pilot Study.* The initial phase of testing involved a small group of 66 students, each from diverse backgrounds, to assess the system's functionality and gather feedback on its user experience. This pilot phase highlighted several performance challenges, particularly with the custom facial recognition model. Due to the model's high computational load, the system struggled with identity verification, causing delays and an extended setup time. These inefficiencies were particularly evident on lower-powered devices, leading to performance degradation.

To address these challenges, the development team integrated the Luxand Face Recognition API, which processed identity verification based on the user's uploaded profile photo. This switch significantly reduced the computational burden and resulted in a more stable and faster setup time. As a result, the system saw a reduction in technical issues and false positives, thereby improving user experience and enhancing the system's reliability.

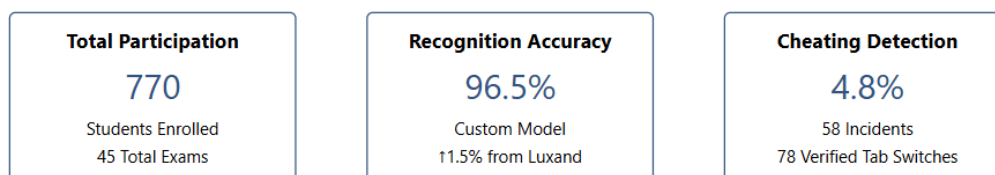
Figure 3 illustrates the comparative analysis of system performance metrics before and after the implementation of the Luxand API. Key improvements included a 35% reduction in computational requirements, which allowed the system to operate more smoothly and on a wider range of devices.



**Figure 3: Comparative Analysis of System Performance Metrics**

Despite these improvements, some areas for further optimization were identified. For instance, while Luxand's API helped streamline the recognition process, certain suspicious behaviors - such as repeated glancing away from the screen - were not fully captured by the existing model. This limitation prompted further adjustments and the introduction of a more sophisticated monitoring system for the next phase.

*Phase 2: Educational Platform Rollout.* Building on the success of the pilot study, the intelligent proctoring system was deployed on an online educational platform offering certification courses. This phase involved 770 students who voluntarily participated in the testing process. All participants received detailed instructions before the testing began and provided consent for participation. The primary objective of this phase was to assess the system's scalability and effectiveness in a larger, more diverse setting (Figure 4).



**Figure 4: Metrics of the Success of Implementation in Educational Institutions**

During this phase, while the Luxand API had improved efficiency, the existing model still struggled to detect certain suspicious behaviors, such as frequent glancing away or unauthorized interactions. To address this, the team developed a custom in-house model for enhanced behavioral monitoring. Deployed on a dedicated server, it enabled simultaneous processing of facial recognition and behavior analysis, ensuring stable performance even under high demand. This setup improved detection of cheating indicators, like tab switching or multiple individuals in the frame, without compromising system efficiency.

This phase also involved extensive collaboration with educators and technical teams to refine the system's ability to identify common cheating patterns and suspicious behaviors specific to online exams. As a result, the system's monitoring features were enhanced to detect additional indicators of suspicious activity, such as the presence of multiple people in the camera frame or inconsistencies in user behavior. The upgraded system could now provide a more detailed log of user actions, such as tab-switching or changes in screen focus, enabling instructors to conduct more thorough analyses of potential academic dishonesty.

To enhance security and ensure transparency, the system captures and stores webcam images and screen screenshots when suspicious activities or tab-switching occur. These logs serve as verifiable evidence, fostering trust between students and educators. Figures 5 and 6 illustrate the student and proctor interfaces, highlighting live monitoring, violation logging, and timestamped records of suspicious actions. Figure 5 shows the facial recognition and violation logging interface, while Figure 6 presents the post-test review screen for instructors to verify flagged events.

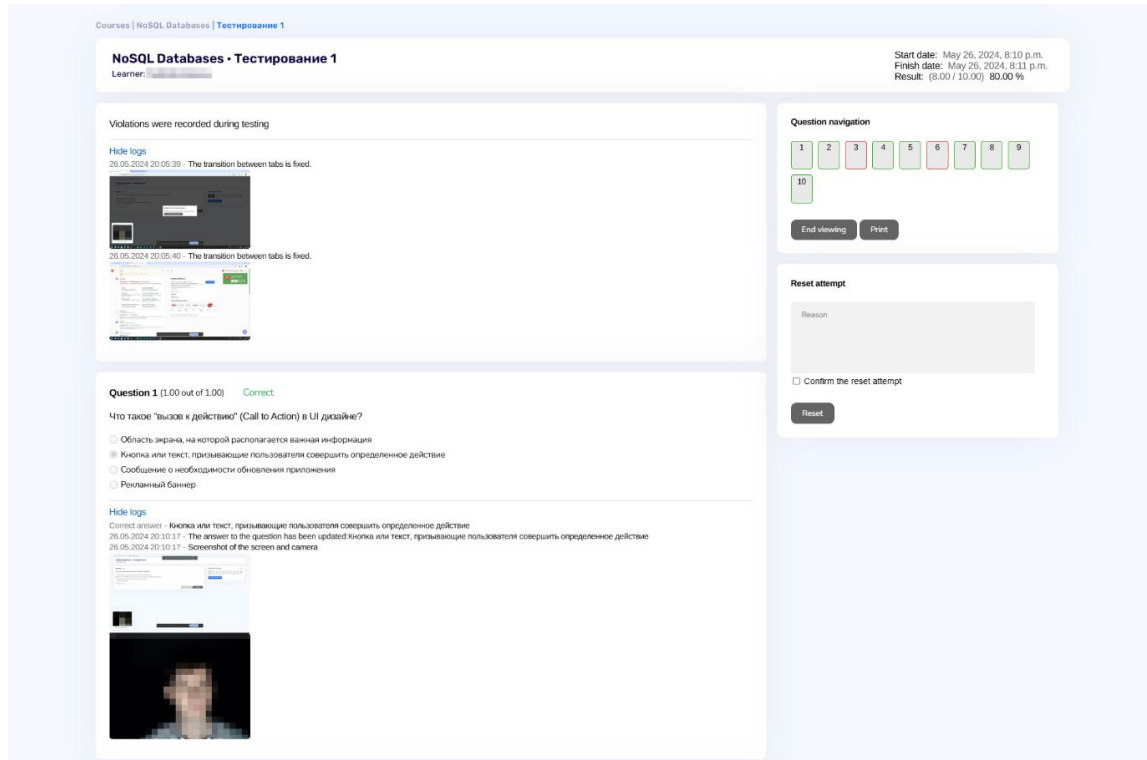


Figure 5: The Interface for Reviewing the Student's Results with Detailed Logs

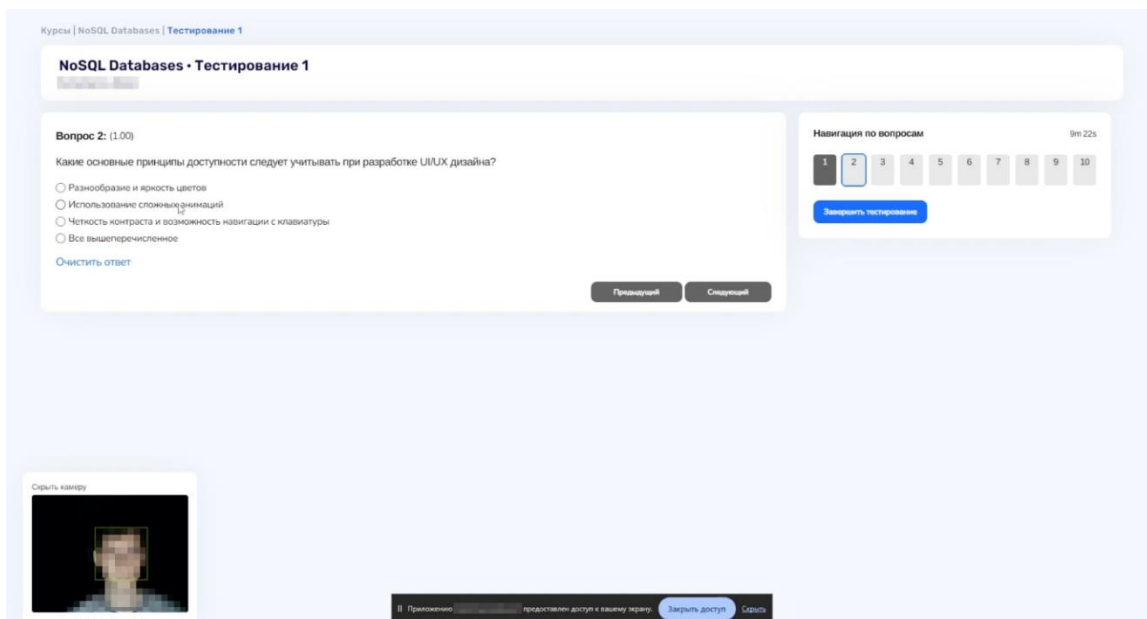
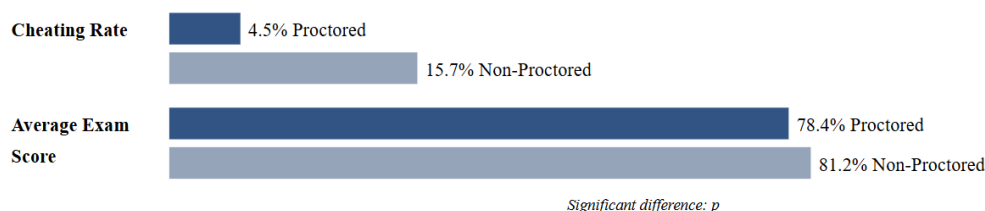


Figure 6: The Student's Interface During Testing with a Working Proctoring System

*Phase 3: Comparative Study.* Finally, a comparative study was conducted to assess the impact of the new in-house model on proctored versus non-proctored environments. This phase involved 770 students, split into two groups of 385 students each, allowing for a direct comparison between the effectiveness of the proctoring system in both settings. The custom model enabled deeper behavioral analysis, including the detection of suspicious actions, adding an additional layer of security to the proctoring system.

The primary goal of the comparative study was to evaluate not only the effectiveness of the system in reducing cheating but also its broader impact on academic integrity and student behavior. Weiner and Hurtz (2017) found that online proctoring can offer security and fairness levels comparable to traditional in-person exams, but emphasized the need for careful implementation to address student concerns and ensure consistent

performance outcomes. By including a non-proctored control group, the study allowed for clear insights into the differences between the two environments. The results indicated that the proctored group experienced a significantly lower rate of cheating and a slightly lower average exam score. Specifically, the cheating rate in the proctored group was 4.5%, compared to 15.7% in the non-proctored group. While the proctored group had a slightly lower average exam score (78.4%) compared to the non-proctored group (81.2%), this suggested that the system effectively deterred dishonest behavior, leading to a more genuine assessment of student knowledge (Figure 7).



**Figure 7: Proctored vs. Non-Proctored Performance Analysis**

The custom model's ability to provide detailed logs of student behavior proved invaluable in identifying common cheating strategies and continuously refining the system's detection capabilities. The system successfully detected instances where students attempted to communicate with others during the exam or use secondary devices, providing key insights into cheating behaviors. This data helped improve the model's ability to distinguish between legitimate behaviors and potential cheating attempts, making the system more robust in future applications. Moreover, the use of behavioral analysis allowed educators to gain a deeper understanding of student engagement during exams. By identifying suspicious behaviors, educators were better equipped to provide guidance on proper exam conduct and foster a culture of academic honesty. Thus, the system not only acted as a deterrent against cheating but also served as an educational tool to promote academic integrity.

The analysis of the system's performance further revealed the importance of continuous improvement. The system's ability to detect cheating was significantly enhanced by the custom model, reducing the incidence of cheating from 58 incidents (4.8% of exams monitored) in Phase 2 to 36 incidents (4.5% of exams) in Phase 3. The system also showed improved detection rates for behaviors such as tab-switching, multiple faces in the camera frame, and suspicious movement, demonstrating the value of ongoing refinement and optimization.

Additionally, the system's development process involved substantial iteration, with significant improvements in facial recognition accuracy, cheating incident detection, and overall system performance (Table 2). The phase also highlighted the importance of transparency and clear communication with students regarding the proctoring process. By ensuring students understood how their data would be used and how the proctoring system functioned, the development team successfully addressed privacy concerns and fostered trust. This approach helped in gaining student cooperation, ensuring the proctoring system was viewed not as intrusive surveillance but as a necessary tool for maintaining academic standards.

**Table 2: Performance Metrics of Intelligent Proctoring System Across Testing Phases**

| Metric/Feature                  | Pre-Luxand Phase 1 | Post-Luxand Phase 1   | Phase 2 (Luxand + Custom Model) | Phase 3 (Enhanced Custom Model) |
|---------------------------------|--------------------|-----------------------|---------------------------------|---------------------------------|
| Average Setup Time (minutes)    | 3.5                | 2.1                   | 2.1                             | 1.9                             |
| Facial Recognition Accuracy (%) | 89                 | 95                    | 96.5                            | 97                              |
| Number of Technical Issues      | 18                 | 7                     | 5                               | 3                               |
| Detected Cheating Incidents     | N/A                | N/A                   | 58 (4.8% of exams monitored)    | Reduced to 36 (4.5% of exams)   |
| Tab-Switch Detection Events     | N/A                | 150 flagged, 40 valid | 340 flagged, 78 confirmed       | 420 flagged, 92 confirmed       |
| Flagged Suspicious Behaviors    | N/A                | 30                    | 120                             | 200                             |

| Metric/Feature                           | Pre-Luxand Phase 1 | Post-Luxand Phase 1 | Phase 2 (Luxand + Custom Model) | Phase 3 (Enhanced Custom Model) |
|------------------------------------------|--------------------|---------------------|---------------------------------|---------------------------------|
| Devices Compatible                       | PCs only           | PCs, modern laptops | PCs, laptops, low-power devices | PCs, laptops, low-power devices |
| Privacy Compliance (Encryption)          | Basic encryption   | AES Encryption      | AES + GDPR Compliance           | Advanced controls with consent  |
| Exam Coverage (students)                 | 66                 | 66                  | 770                             | 770                             |
| Cheating Rate in Non-Proctored Exams (%) | 15.7               | 15.7                | 15.7                            | 15.7                            |
| Cheating Rate in Proctored Exams (%)     | Not Monitored      | 4.8                 | 4.8                             | 4.5                             |
| Data Storage Logs (GB/exam)              | 0.2                | 0.5                 | 0.8                             | 1.2                             |
| Multiple Face Detection Support          | No                 | Limited             | Supported                       | Enhanced Real-time Detection    |

The detailed monitoring capabilities of the proctoring system were further highlighted in the evaluation of various violation types. As shown in Table 3, different types of suspicious behavior, such as tab switching, multiple faces in view, and irregular keyboard patterns, were detected with high accuracy, and false positive rates were minimized through model optimization. These insights provided a clearer understanding of the common cheating tactics employed and allowed for ongoing improvements in the system's detection algorithms.

**Table 3: Violation Type Statistics in Online Proctoring System**

| Violation Type              | Frequency | Detection Rate | False Positive Rate |
|-----------------------------|-----------|----------------|---------------------|
| Tab Switching               | 320       | 92%            | 8%                  |
| Multiple Faces              | 65        | 95%            | 5%                  |
| Suspicious Movement         | 200       | 87%            | 13%                 |
| Irregular Keyboard Patterns | 145       | 93%            | 7%                  |
| Full Screen Exit            | 340       | 94%            | 6%                  |

#### 4.2 Impact on Academic Integrity

To assess the impact of the intelligent proctoring system, a qualitative survey was conducted among 18 mentors and 54 platform users who had directly experienced the system. The survey aimed to assess perceptions of the fairness, effectiveness, and overall user satisfaction with the proctored exams (Table 4). While 83% of mentors felt that proctored exams were significantly more secure and contributed to academic integrity, only 52% of students expressed satisfaction with the fairness of the system. Notably, 48% of students reported feeling more stress and anxiety during the proctored exams, contrasting with only 17% of mentors who shared similar concerns. This difference in opinion reflects the challenge of balancing the need for academic integrity with the student experience. While 94% of mentors viewed proctoring as essential for maintaining exam integrity, only 37% of students agreed with this view, and 63% preferred non-proctored exams. Despite these mixed perceptions, both groups acknowledged the usefulness of the proctoring system, with mentors valuing the enhanced security it provided, and students recognizing its role in deterring dishonest behavior.

Overall, Phase 3 of the study demonstrated that the system under development, with its custom model and enhanced behavioral analysis capabilities, significantly improved the detection of cheating and contributed to a more accurate assessment of student knowledge. The findings underscore the importance of continuous improvement, transparency, and clear communication with students, ensuring that proctoring systems are both effective in maintaining academic integrity and sensitive to the concerns and experiences of students.

**Table 4: Survey Results on Perceptions of Proctored Exams**

| Survey Question                               | Mentors (n = 18) | Students (n = 54) | Overall (n = 72) |
|-----------------------------------------------|------------------|-------------------|------------------|
| Satisfaction with fairness of proctored exams | 83 % (15)        | 52 % (28)         | 60 % (43)        |
| Perceived stress/anxiety due to proctoring    | 17 % (3)         | 48 % (26)         | 40 % (29)        |

| Survey Question                                | Mentors (n = 18) | Students (n = 54) | Overall (n = 72) |
|------------------------------------------------|------------------|-------------------|------------------|
| Preference for non-proctored exams             | 5.5 % (1)        | 63 % (34)         | 49 % (35)        |
| Importance of proctoring for exam integrity    | 94 % (17)        | 37 % (20)         | 51 % (37)        |
| Overall usefulness of proctoring for integrity | 94 % (17)        | 50 % (27)         | 61 % (44)        |

## 5. Discussion

The findings indicate that the intelligent proctoring system reduces cheating and positively influences student behavior. The presence of monitoring technologies serves as a deterrent, motivating students to focus more on preparation, knowing that dishonest behavior is likely to be detected. This aligns with previous research that emphasizes the deterrent effects of such systems on academic dishonesty, such as the work of Weiner and Hurtz (2017), who found that online proctoring can promote academic integrity in a similar manner to traditional in-person proctoring methods. However, several challenges persist - particularly concerning privacy, fairness, and ethical considerations - that must be addressed for these technologies to gain widespread acceptance.

### 5.1 Privacy Concerns and Data Collection

One of the primary concerns with proctoring systems is the significant amount of data collection required, such as facial recognition and detailed logs of user interactions. Mutumukwe et al. (2023) highlight that online proctoring systems may compromise contextual integrity by collecting sensitive and potentially excessive personal data. Students have expressed discomfort with the level of surveillance, fearing the misuse or potential breach of their personal data. This concern is particularly relevant in systems that use invasive techniques, like facial recognition, which can make students feel that their privacy is being invaded. To address these concerns, robust safeguards are essential, including encryption, strict data access controls, and compliance with privacy regulations. Transparent communication with students about how their data is stored, used, and protected is crucial for alleviating privacy concerns and fostering trust. As the study demonstrates, privacy compliance (AES encryption and GDPR adherence) was a key feature of the proctoring system, which likely played a role in gaining user confidence. To further enhance transparency, a potential improvement could be the introduction of a notification mechanism informing students whenever data collection is initiated. Additionally, implementing automated data deletion policies post-exam verification would help address concerns regarding unnecessary data retention while maintaining academic integrity.

### 5.2 Ethical Implications and Bias in Proctoring Technologies

Automated proctoring systems raise ethical concerns related to fairness, privacy, and student well-being. Burgess et al. (2022) highlight issues of algorithmic bias and over-surveillance, which can undermine trust and create a stressful exam environment. In this study, while mentors valued the system's security, 48% of students reported increased anxiety, suggesting that monitoring can unintentionally heighten exam pressure. A key concern is bias in facial recognition technologies, which may struggle to accurately identify individuals from diverse demographics (Burgess et al., 2022). This can lead to higher false-positive rates, causing undue scrutiny for certain groups. While system accuracy improved (from 89% to 97%), ongoing refinements—such as expanding training datasets—are needed to enhance fairness.

Another factor contributing to anxiety is the perceived lack of control over monitoring. Some students feared routine behaviors, like adjusting posture or briefly looking away, could be misinterpreted as suspicious. Mukherjee et al. (2024) suggest that allowing users to customize aspects of monitoring, such as blurring background elements, could reduce stress and build trust. Greater transparency and control over data collection may help mitigate these concerns while maintaining exam integrity.

### 5.3 Refinement of Behavioral Monitoring Capabilities

Beyond facial recognition, the advanced proctoring system must continuously improve its ability to detect and analyze a wide range of suspicious behaviors, such as unauthorized collaboration or the use of secondary devices. In Phase 2 and 3 of current study, we found that the system's detection capabilities, including the ability to flag tab-switching and detect multiple faces in the camera frame, significantly enhanced its monitoring accuracy. However, further research and development are needed to refine these systems. Ngo et al. (2024) propose a multi-modal approach combining behavioral data with environmental monitoring to detect abnormal activities during online exams. Such systems are better equipped to identify more subtle forms of cheating, like unauthorized collaboration or the use of hidden devices, while minimizing the risk of flagging legitimate student behaviors as misconduct.

Mukherjee et al. (2024) also suggest that the perception of fairness in online proctoring systems can be improved by implementing greater transparency and allowing students to control aspects of their monitoring, such as visual data obfuscation. This approach could help reduce students' feelings of being overly surveilled and improve the overall experience. Allowing students to have some control over their data - such as opting to blur background details or mask parts of their face during recognition - may help reduce the anxiety associated with being watched, fostering a more positive testing experience.

#### **5.4 Impact on Student Behavior and Performance**

The proctored environment has significantly reduced cheating, with a 4.5% cheating rate in the proctored group compared to 15.7% in the non-proctored group. This suggests that monitoring acts as a deterrent, encouraging students to prepare more thoroughly rather than relying on dishonest methods. This finding aligns with Weiner and Hurtz (2017), who emphasized that online proctoring upholds academic integrity similarly to traditional in-person exams.

However, the study also found that the average exam score was slightly lower in the proctored group (78.4%) than in the non-proctored group (81.2%), indicating that surveillance may induce stress, potentially affecting cognitive performance. Similar concerns have been reported in international studies, where students in proctored settings often experience higher anxiety levels, which can impact their test-taking ability. Universities globally are adopting hybrid proctoring models that combine automated monitoring with human oversight to mitigate such issues while maintaining fairness.

Beyond preventing cheating, proctoring also influences learning habits. Knowing they will be monitored, students are more likely to engage in deeper learning strategies rather than surface-level memorization. This suggests that proctoring may contribute to long-term improvements in student preparation and academic discipline. However, achieving a balance between security and student well-being remains essential. Measures such as providing structured breaks, reducing unnecessary monitoring intrusiveness, and increasing transparency in proctoring policies can help maintain academic integrity without negatively impacting student performance.

#### **5.5 Ethical Use and Transparency in Proctoring Systems**

The ethical deployment of enhanced proctoring systems requires balancing security with student rights. Moreno-Guerrero et al. (2020) emphasize that transparency and accessibility are critical for student acceptance of e-learning tools. Effective integration of proctoring into online education depends on clear communication about system functionality, data collection, and privacy safeguards to build trust. Ensuring fairness, inclusivity, and minimal invasiveness should remain a priority.

The findings of this study provide insights that can inform global e-learning practices. While surveillance concerns persist in many educational systems (Moreno-Guerrero et al., 2020), our results indicate that transparency in data protection and giving students some control over monitoring can alleviate anxiety and improve acceptance. Similar studies, such as Mukherjee et al. (2024), suggest that customizable monitoring features, such as background blurring, enhance student trust and perceived fairness.

This study demonstrates that privacy-conscious, adaptive proctoring solutions can serve as scalable models for other educational platforms. By integrating automated monitoring with human oversight, institutions can maintain exam integrity while addressing ethical concerns. Future research should explore how such systems can be refined for diverse educational settings, ensuring accessibility and compliance with different institutional and regulatory frameworks worldwide.

### **6. Limitations**

Despite its advancements, the intelligent proctoring system faces several challenges. Webcam dependency can lead to verification errors for students with low-quality cameras or unstable internet, potentially affecting fairness. Privacy concerns remain significant, as biometric data collection raises security and consent issues, despite robust encryption and compliance measures. Additionally, facial recognition biases, particularly for diverse demographics, necessitate ongoing improvements in dataset diversity and alternative verification methods like voice recognition. Lastly, the system's complexity and resource demands may pose adoption barriers for smaller institutions or those in developing regions, highlighting the need for a more lightweight and cost-effective version.

## 7. Conclusions and Future Work

The integration of advanced proctoring technologies has demonstrated significant potential in enhancing academic integrity within online learning environments. This study confirms that automated monitoring can effectively reduce cheating incidents, contributing to fairer and more credible assessments. However, ongoing challenges remain, particularly concerning privacy, inclusivity, and student well-being. Future improvements should focus on refining detection algorithms to minimize biases, enhancing system adaptability across different educational settings, and exploring alternative verification methods, such as voice recognition and multi-factor authentication. These refinements will ensure that proctoring remains both effective and ethically responsible.

Beyond preventing academic dishonesty, proctoring systems can be leveraged to improve learning behaviors. The inclusion of behavioral analytics—such as eye-tracking, gesture recognition, and posture analysis—could provide educators with insights into student engagement, cognitive load, and stress levels. This data could help refine assessment strategies to better support students during high-stakes exams. Additionally, ensuring compliance with evolving privacy regulations (e.g., GDPR) and introducing student-controlled monitoring options will be key to fostering trust and transparency.

For broader accessibility, future research should focus on developing lightweight, low-bandwidth-compatible versions of the system to support students in regions with limited technological infrastructure. Additionally, long-term studies should assess the impact of enhanced proctoring on academic performance, student perceptions, and test-taking behaviors to further optimize the system.

To maximize its global applicability, interdisciplinary collaboration between educators, technologists, and policymakers will be essential. By integrating ethical safeguards, promoting inclusivity, and addressing student concerns, smart proctoring systems can evolve into scalable, fair, and privacy-conscious solutions that uphold academic integrity while maintaining student trust in online education.

**Funding:** This research was funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP19177277).

**Ethics Statement:** This study was conducted using the BlockchainStudy.kz platform, where students voluntarily participated in online courses and assessments. The intelligent proctoring system recorded data related to facial recognition, user activity monitoring, and browser behavior tracking solely for the purpose of ensuring academic integrity. All recorded data remained within the platform's existing operational framework and were never used for experimental interventions, external research, or third-party sharing. No personally identifiable information was included in the research analysis. All data used for this study were fully anonymized and aggregated before any evaluation was conducted, ensuring that individual students could not be identified at any stage. The research relied exclusively on statistical summaries of system performance and general behavioral patterns, with no direct association to specific users. No new data collection procedures were introduced for research purposes. The proctoring system operated under standard platform policies, which all users agreed to before taking their exams. All participants provided consent for data processing as part of the BlockchainStudy.kz platform's user agreement, ensuring that data usage aligned with the policies users were informed of in advance. Because this study did not involve direct human subject research, external interventions, or personally identifiable data analysis, no formal ethics approval was required. The study fully complied with General Data Protection Regulation (GDPR) principles, ensuring that all data were handled securely, access was strictly controlled, and retention policies were followed. All proctoring logs were stored only as necessary for integrity verification and were automatically deleted after the designated retention period. This study adheres to best practices in ethical educational research and privacy protection. Since only system-generated, anonymized data were used in a retrospective analysis, and no personal data were collected outside the platform's normal operational scope, it does not fall under research requiring formal institutional ethics approval.

**AI Statement:** This research paper was written without the assistance of artificial intelligence (AI) tools. All content, including the conceptual framework, data analysis, writing, and editing, was developed entirely by the authors. No AI-generated text, automated editing, or machine-assisted research synthesis was used in the preparation of this paper. The findings, arguments, and conclusions presented in this study are the result of human-led research and critical analysis, ensuring academic integrity and adherence to ethical research standards.

## References

- Balash, D., Kim, D., Shaibekova, D., Fainchtein, R., Sherr, M. and Aviv, A., 2021. Examining the examiners: Students' privacy and security perceptions of online proctoring services. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2106.05917>
- Bilen, E. and Matros, A., 2021. Online Cheating Amid COVID-19. *Journal of Economic Behavior & Organization*, 182, pp. 196-211. <https://doi.org/10.1016/j.jebo.2020.12.004>
- Brandeis University, 2020. Academic dishonesty and COVID-19: A biological explanation. *Write Now*. Available at: <https://www.brandeis.edu/writing-program/write-now/2020-2021/arie-rotem/index.html> [Accessed 16 February 2025].
- Bretag, T., Mahmud, S., Wallace, M., Walker, R., James, C., Green, M. and East, J., 2018. Contract cheating: A survey of Australian university students. *Studies in Higher Education*, 43(11), 1921-1939. <https://doi.org/10.1080/03075079.2018.1462788>
- Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., van Haeringen, K., Saddiqui, S. and Rozenberg, P., 2019. Contract cheating and assessment design: Exploring the connection. *Assessment & Evaluation in Higher Education*, 44(5), 676-691. <https://doi.org/10.1080/02602938.2018.1527892>
- Burgess, B., Ginsberg, A., Felten, E. W. and Cohnsey, S., 2022. Watching the watchers: Bias and vulnerability in remote proctoring software. *arXiv preprint*. Available at: <https://doi.org/10.48550/arXiv.2205.03009> [Accessed 15 November 2024].
- Coghlan, S., Miller, T., and Paterson, J., 2021. Good Proctor or "Big Brother"? Ethics of Online Exam Supervision Technologies. *Philosophy & Technology*, 34, pp. 1581–1606. <https://doi.org/10.1007/s13347-021-00476-1>
- Dawson, P., 2021. Defending Assessment Security in a Digital World: Preventing E-Cheating and Supporting Academic Integrity in Higher Education. *Routledge*. <https://doi.org/10.4324/9780429324178>
- Eaton, S., 2020. Academic Integrity During COVID-19: Reflections from the University of Calgary. *International Studies in Educational Administration*, 48(1), pp. 80-85. <https://dx.doi.org/10.11575/PRISM/38013>
- Gamage, K., Dehideniya, S., Zhiheng, X. and Tang, X., 2023. Contract cheating in higher education: Impacts on academic standards and quality. *Journal of Applied Learning & Teaching*. 6(2). <https://doi.org/10.37074/jalt.2023.6.2.24>
- Nigam, A., Pasricha, R., Singh, T. and Churi, P., 2021. A systematic review on AI-based proctoring systems: Past, present and future. *Education and Information Technologies*, 26(5), pp.6421–6445. <https://doi.org/10.1007/s10639-021-10597-x>
- Lancaster, T. and Cotarlan, C., 2021. Contract cheating by STEM students through a file sharing website: A Covid-19 pandemic perspective. *International Journal for Educational Integrity*, 17(1), 1-16. <https://doi.org/10.1007/s40979-021-00070-0>
- Liu, Y., Ren, J., Xu, J., Bai, X., Kaur, R. and Xia, F., 2024. Multiple Instance Learning for Cheating Detection and Localization in Online Examinations. *arXiv preprint*. Available at: <https://arxiv.org/abs/2402.06107> [Accessed 15 November 2024].
- Malik, A.A., Hassan, M., Rizwan, M., Mushtaque, I., Lak, T.A. and Hussain, M., 2023. Impact of academic cheating and perceived online learning effectiveness on academic performance during the COVID-19 pandemic among Pakistani students. *Frontiers in Psychology*, 14, p.1124095. <https://doi.org/10.3389/fpsyg.2023.1124095>
- Moreno-Guerrero, A. J., Aznar-Díaz, I., Cáceres-Reche, M. P. and Alonso-García, S., 2020. E-Learning in the Teaching of Mathematics: An Educational Experience in Adult High School. *Mathematics*, 8(5), 840. <https://doi.org/10.3390/math8050840>
- Moyo, R., Ndebvu, S., Zimba, M. and Mbelwa, J., 2023. A Video-based Detector for Suspicious Activity in Examination with OpenPose. *arXiv preprint*. Available at: <https://arxiv.org/abs/2307.11413> [Accessed 15 November 2024].
- Mukherjee, S., Distler, V., Lenzini, G. and Cardoso-Leite, P., 2024. Balancing the perception of cheating detection, privacy and fairness: A mixed-methods study of visual data obfuscation in remote proctoring. *arXiv preprint*. Available at: <https://arxiv.org/abs/2406.15074> [Accessed 15 November 2024].
- Mutumukwe, C., Han, S., Viberg, O. and Cerratto-Pargman, T., 2023. Privacy as Contextual Integrity in Online Proctoring Systems in Higher Education: A Scoping Review. *arXiv preprint*. Available at: <https://doi.org/10.48550/arXiv.2310.18792>
- Muzaffar, A.W., Tahir, M., Anwar, M.W., Chaudry, Q., Mir, S.R. and Rasheed, Y., 2021. A systematic review of online exams solutions in e-learning: Techniques, tools and global adoption. *arXiv preprint*. Available at: <https://arxiv.org/abs/2010.07086> [Accessed 16 February 2025].
- Newton, P.M. and Essex, K., 2024. How common is cheating in online exams and did it increase during the COVID-19 pandemic? A systematic review. *Journal of Academic Ethics*, 22, pp. 323-343. <https://doi.org/10.1007/s10805-023-09485-5>
- Ngo, D. A., Nguyen, T. D., Dang, T. L. C., Le, H. H., Ho, T. B., Nguyen, V. T. K. and Nguyen, T. T. H., 2024. Examining monitoring system: Detecting abnormal behavior in online examinations. *arXiv preprint*. Available at: <https://arxiv.org/abs/2402.12179> [Accessed 15 November 2024].
- Raman, R., Bandlamudi, S., Gangadharan, V., Vachharajani, H. and Nedungadi, P., 2021. Adoption of online proctored examinations by university students during COVID-19: Innovation diffusion study. *Education and Information Technologies*, 26, pp.7339–7358. <https://doi.org/10.1007/s10639-021-10581-5>
- Tiong, L. C. O. and Lee, H. J., 2021. E-cheating Prevention Measures: Detection of Cheating at Online Examinations Using Deep Learning Approach - A Case Study. *arXiv preprint*. Available at: <https://arxiv.org/abs/2101.09841> [Accessed 15 November 2024].

# Evaluating ChatGPT's Reliability in Second Language Acquisition (SLA): Insights on Language Skills and Technology's Role

Albatool Ahmad Alhazmi and Muneera Muftah

Department of English, College of Languages and Translation, Najran University, Najran, Saudi Arabia

[Aaalhazmi@nu.edu.sa](mailto:Aaalhazmi@nu.edu.sa)

[munmef5@gmail.com](mailto:munmef5@gmail.com) (Corresponding Author)

<https://doi.org/10.34190/ejel.23.1.3762>

An open access article under [CC Attribution 4.0](#)

**Abstract:** Artificial Intelligence (AI) has become a transformative tool across various fields, including education, offering innovative opportunities for second language (L2) learning and teaching. The potential of AI-driven tools like ChatGPT in Second Language Acquisition (SLA) remains an area of emerging interest, with implications for enhancing pedagogical approaches and providing scalable, interactive support for learners and educators. This study evaluates the reliability and quality of ChatGPT-4.0's responses to 48 key questions related to SLA, encompassing themes such as technology's role and language skills (reading, writing, listening, and speaking). The questions were rated by five expert linguists using a Likert scale to assess accuracy, and relevance. Statistical analyses, including the Kruskal-Wallis H test ( $p < 0.05$ ) and post-hoc pairwise comparisons, revealed significant differences among evaluators ( $p < 0.001$ ), with most responses rated as good (26.25%) or very good (55.41%). Despite these positive outcomes, Fleiss's Kappa coefficient (0.006) highlighted low inter-rater agreement, signaling a need for standardized evaluation frameworks. While ChatGPT shows promise in addressing diverse SLA topics, its limitations—such as occasional inaccuracies—underscore the necessity for ongoing research to optimize its utility in educational contexts. This study contributes to SLA pedagogy by demonstrating the potential and boundaries of integrating AI into language learning, advocating for cautious and evidence-based adoption.

**Keywords:** Artificial intelligence (AI), ChatGPT, Language education, Language learning technologies, Reliability, Second language acquisition (SLA)

## 1. Introduction

Artificial intelligence (AI) language models have been transforming the landscape of education and scientific research. These novel models attract the attention of researchers, educators, and users across various fields due to their impressive and versatile capabilities, including content generation in multiple domains such as texts, images, and videos (Liu et al., 2023; Ray, 2023). Among these models, ChatGPT, based on the Generative Pre-trained Transformer (GPT) architecture, has gained significant recognition and acceptance in professional and educational sectors, including medicine, healthcare, finance, business, management, science, translation, and education (Baskara and Mukarto, 2023; Kalla, Smith and Carolina, 2023; Kung et al., 2023; Lund and Wang, 2023; Ray, 2023).

This revolutionary technology is designed to perform several natural language processing (NLP) tasks that mimic human capabilities, such as learning, reasoning, data analysis, decision-making, question-answering, and problem-solving (Kalla, Smith and Carolina, 2023; Kung et al., 2023; Liu et al., 2023; Muftah, 2024). As these models continue to evolve, their integration into daily life and education becomes increasingly inevitable, with notable implications for individualized learning and interaction. However, concerns regarding the accuracy and quality of responses generated by ChatGPT have prompted scholars to call for more rigorous evaluations of its potential benefits, limitations, and risks in academic contexts.

Recent studies have highlighted the need to explore the role of AI technologies in education. For instance, Huallpa et al. (2023) and Baskara and Mukarto (2023) emphasize the importance of investigating ChatGPT's impact on higher education language learning. Similarly, Ray (2023) underscores the necessity of examining ChatGPT's potential to advance deep and personalized learning and scientific research. While the technology offers significant promise, researchers such as Chaka (2023), Liu et al. (2023), and Sallam (2023) advocate for more research to assess the validity of its responses, identify its limitations, and establish strategies to mitigate risks associated with its use.

Despite these discussions, one critical gap in the literature is the lack of empirical studies assessing the reliability of ChatGPT's responses specifically within Second Language Acquisition (SLA) contexts. Current research has largely focused on AI's general application in education, with limited attention to its specific impact on SLA-

related areas such as reading, writing, listening, and speaking skills. Additionally, there is insufficient analysis of how well AI-generated content aligns with established SLA theories and pedagogical frameworks.

To bridge this gap, this study critically evaluates the accuracy and reliability of ChatGPT's responses to SLA-related inquiries, particularly in the context of language skills learning. While AI has demonstrated potential in language education, its effectiveness remains uncertain due to inconsistencies in generated responses, potential misinformation, and lack of alignment with SLA principles. Understanding these factors is essential for assessing whether ChatGPT can serve as a reliable tool for learners and educators.

Given these considerations, the present study seeks to evaluate the reliability of ChatGPT in providing accurate and high-quality responses to inquiries concerning key aspects of SLA. These aspects include the role of technology, reading, writing, listening, and speaking, with a focus on the model's capacity to contribute meaningfully to SLA-related issues. This evaluation addresses a notable gap in research, as limited studies have investigated the reliability and accuracy of AI models within the branches and sub-branches of linguistics.

To achieve this aim, the study addresses the following research questions:

- How reliable are the responses generated by ChatGPT in addressing key questions in SLA?
- To what extent does ChatGPT provide accurate and meaningful information on specific language skills (reading, writing, listening, and speaking) in SLA?
- What are the limitations and areas for improvement in ChatGPT's performance as evaluated by experts in SLA?

By addressing these questions, this study aims to contribute to the growing body of knowledge on the use of AI in SLA and provide insights into its potential applications and limitations in language education.

## **2. Literature Review**

### **2.1 Advancements in AI and ChatGPT Applications**

AI, particularly through language models like the Generative Pretrained Transformer (GPT) series, has revolutionized natural language processing (NLP). These models perform tasks such as contextual understanding, real-time interaction, and personalized support Brown et al. (2020) and Liu et al. (2023) discuss how these advancements enhance AI's adaptability in various fields, particularly education. ChatGPT, a generative chatbot, exemplifies these capabilities, offering human-like conversational responses and demonstrating human-level performance in various professional and academic domains. For instance, Firat (2023) highlights ChatGPT's impact on academic writing and research assistance, while Ray (2023) evaluates its performance in subject-specific tasks.

AI applications, including ChatGPT, are extensively employed across industries such as medicine, business, education, and scientific research (Baskara and Mukarto, 2023; Kalla, Smith and Carolina, 2023; Kung et al., 2023). For instance, ChatGPT aids medical diagnostics by providing interactive and timely responses for disease prevention and treatment planning (Branum and Schiavenato, 2023; Sarraju et al., 2023; Tanaka et al., 2023). Similarly, in education, AI tools like ChatGPT support diverse academic functions, including content creation, metadata generation, and personalized tutoring (Huallpa et al., 2023; Lund and Wang, 2023).

### **2.2 AI and Education: Opportunities and Challenges**

AI's integration into education has been transformative, particularly in enhancing individualized learning experience (Srinivasa, Kurni and Saritha, 2022). Lund and Wang (2023) note that AI enhances self-directed learning by providing real-time feedback, an observation echoed by Firat (2023), who highlights ChatGPT's potential in fostering learner autonomy, motivation, and self-directed learning by providing personalized guidance and prompt feedback. Similarly, Sharma and Yadav (2023) discuss AI's contribution to open and online education, enabling interactive and adaptive learning.

Despite these advantages, AI adoption raises concerns about accuracy, ethical implications, and privacy. For example, Baskara and Mukarto (2023) warn about biases in ChatGPT's responses, which can lead to culturally insensitive outputs. Furthermore, Huallpa et al. (2023) argue that AI tools must be adapted to diverse learners' skills to avoid inequities in learning outcomes. These challenges highlight the need for responsible integration of AI in education.

### 2.3 ChatGPT in Language Education

The application of ChatGPT in language education holds immense potential, supporting language learning through automated tutoring, writing assistance, language assessment, and personalized language practice in simulated real-life conversational contexts (Barrot, 2023; Brown et al., 2020; Huang, Hew and Fryer, 2022). For example, ChatGPT enhances English proficiency by enriching vocabulary, improving writing skills, fostering reading comprehension, and supporting translation activities (Baskara and Mukarto, 2023). It also aids in improving speaking skills through activities such as storytelling and narration.

Baskara and Mukarto (2023) emphasize the wealth of opportunities ChatGPT provides for language learning, highlighting its role in enhancing communicative proficiency and fostering personalized, collaborative language learning experiences. They illustrate how ChatGPT supports English teaching and learning both inside and outside classrooms, enhancing learners' abilities in writing, grammar, and vocabulary acquisition while developing their social and interactional skills. Similarly, Karataş et al. (2024) found that ChatGPT positively affects students' learning experiences, helping them improve their writing abilities, grammar, and vocabulary while fostering critical social skills.

In addition to its pedagogical applications, ChatGPT offers several technological affordances, including timeliness, ease of use, and personalization. Huang, Hew and Fryer (2022) outline five pedagogical roles of AI chatbots in language learning: as interlocutors, simulations, transmission tools, helplines, and recommendation agents. These roles contribute significantly to enhancing students' social skills through affective and coherent communication. Beyond language learning, AI applications, including ChatGPT, support various areas in applied linguistics, such as language translation, language acquisition, and computer-assisted learning (Chaka, 2023).

Despite these advantages, challenges such as cognitive overload and technological limitations must be addressed to optimize the use of AI technologies in language learning. Researchers, including Karataş et al. (2024), advocate for a balanced approach to integrating AI tools like ChatGPT into language curricula, ensuring their effective application while mitigating potential drawbacks.

### 2.4 Reliability and Challenges of AI Applications

The reliability of AI tools like ChatGPT has been a focus of extensive research across various disciplines, particularly in education and healthcare. While ChatGPT demonstrates considerable potential, it has also revealed limitations that highlight the need for cautious integration into human-centered fields. For instance, in the medical domain, studies have examined ChatGPT's ability to perform diagnostic and educational tasks. Kung et al. (2023) evaluated ChatGPT's performance in the United States Medical Licensing Examination (USMLE), finding that it scored near the passing threshold without expert training, demonstrating its potential role in medical education and clinical decision-making. Similarly, Tanaka et al. (2023) found that ChatGPT provided effective answers on orthodontic topics such as clear aligners, temporary anchorage devices, and digital imaging. However, deficiencies were noted in generating verifiable responses, as Branum and Schiavenato (2023) reported instances of fabricated and self-generated information, underscoring concerns about its trustworthiness.

In educational settings, ChatGPT raises similar reliability issues. While it offers numerous benefits, such as personalized learning experiences and enhanced accessibility (Barrot, 2023; Lund and Wang, 2023), its inability to consistently provide precise, culturally aware, or ethically sound responses presents challenges. Concerns about accuracy, privacy, and data security are common among educators and students alike (Hualpa et al., 2023). For example, Baskara and Mukarto (2023) highlight that ChatGPT may generate inappropriate or misleading content due to biases in its training data, which often lacks human nuance, cultural awareness, and higher-order thinking. Additionally, Hualpa et al. (2023) emphasize that adapting AI-generated materials to suit diverse learner backgrounds, skills, and interests remains a critical issue.

Ethical considerations further complicate the integration of ChatGPT and similar AI tools in education. Researchers like Lund and Wang (2023) discuss concerns such as transparency, accountability, and intellectual property, emphasizing the importance of addressing these issues to avoid inaccurate or inequitable outcomes. Additionally, academic integrity and the preservation of students' critical and creative thinking skills are significant factors in evaluating the role of AI in learning and teaching (Barrot, 2023; Hualpa et al., 2023). The use of AI in language learning illustrates both the promise and the pitfalls of these technologies. Huang, Hew and Fryer (2022) identify limitations such as cognitive load, technological constraints, and the novelty effect, while Karataş et al. (2024) stress the importance of adopting a balanced approach when integrating AI tools into language curricula. Researchers agree that while AI can enhance learners' experiences by improving skills such

as writing, grammar, and vocabulary acquisition, its limitations necessitate further research and methodological refinement to ensure effective and ethical application (Huang, Hew and Fryer, 2022; Baskara and Mukarto, 2023; Sallam, 2023).

In brief, while ChatGPT and similar AI tools demonstrate substantial potential in education and other fields, their limitations in reliability, ethical considerations, and adaptability demand a cautious and balanced approach. Further research is essential to address these challenges and to develop strategies that maximize their benefits while minimizing risks.

## **2.5 ChatGPT and SLA: Bridging AI and Language Acquisition**

Second Language Acquisition (SLA) is a multidisciplinary field dedicated to understanding the processes and factors involved in learning languages other than one's first language (Ellis, 2015). With its foundations in linguistics, psychology, cognitive science, sociology, and education, SLA research has provided invaluable insights into cognitive processes, the similarities and differences between first language (L1) and second language (L2) acquisition, and the development of effective language teaching methodologies (Muftah, 2023a, 2023b; VanPatten, Smith and Benati, 2020). These findings not only inform language education but also contribute to broader fields such as global communication and cognitive science.

In recent years, advancements in artificial intelligence (AI) have introduced novel tools, such as ChatGPT, with potential applications in SLA. ChatGPT offers opportunities to simulate conversational interactions, provide linguistic feedback, and adapt to learners' individual needs, thereby enhancing learning experiences. Despite these promising features, the reliability of ChatGPT's responses to SLA-specific queries remains underexplored. Current research lacks sufficient evaluation of its accuracy and applicability within the nuanced context of SLA.

This study seeks to bridge this gap by evaluating the reliability and potential of ChatGPT in addressing key aspects of SLA. Specifically, it examines the accuracy and quality of ChatGPT's responses through content analysis of self-generated and self-answered queries. By focusing on critical facets of SLA, this research aims to assess whether ChatGPT can effectively support learning processes, address the complex dynamics of L2 acquisition, and complement traditional teaching methodologies. Building on decades of SLA research, which has uncovered the cognitive, social, and pedagogical factors influencing language learning, this study emphasizes the importance of integrating AI into SLA responsibly. While ChatGPT offers innovative possibilities, understanding its limitations and potential applications is crucial to leveraging AI effectively in language education.

## **3. Materials and Methods**

The study aimed to assess the accuracy and relevance of ChatGPT's responses to key questions related to SLA. To achieve this, ChatGPT was used as an interactive tool to generate responses to 48 pre-designed questions covering various SLA topics, such as the role of technology, and language skills including reading, writing, listening, and speaking. These questions were crafted to evaluate ChatGPT's ability to provide accurate, detailed, and coherent answers. The chatbot responded autonomously to each query without additional contextual prompts, ensuring that its responses were generated based solely on the questions asked.

### **3.1 Selection of Questions**

For the evaluation process, ChatGPT-4 was prompted with 48 questions covering six key areas in SLA: (1) key questions in SLA, (2) the role of technology in SLA, (3) reading skills, (4) writing skills, (5) listening skills, and (6) speaking skills. These questions were selected based on their frequency in academic inquiries and their significance in SLA research, drawing from textbooks, scholarly articles, and online discussions. A timeline-based approach was used to ensure a systematic selection process. Over three months, a comprehensive review of SLA literature and academic discourse was conducted to identify the most frequently discussed topics. Eight representative questions were formulated for each category, ensuring broad coverage of SLA concepts. While specific websites were not used as the sole source, the design was guided by theoretical constructs, research studies, and expert consultation. The final selection was refined and validated by experts before being used to prompt ChatGPT-4. The AI-generated responses to these questions were collected and documented in Table 1.

### **3.2 Evaluation Process**

To assess the quality of the responses, five expert researchers in SLA, each possessing over ten years of academic experience in Applied Linguistics and SLA studies, independently rated the answers on a five-point Likert scale. The scale ranged from 1 (very poor) to 5 (very good), with higher scores indicating superior information quality. The Likert scale used in the evaluation assessed ChatGPT's responses across two dimensions: accuracy and

relevance, each rated on a five-point scale. Accuracy measured the correctness and alignment of the responses with established SLA theories, where 1 indicated significant errors and 5 represented error-free and precise information. Relevance assessed the focus and appropriateness of the responses to the questions posed, ranging from 1 for off-topic or tangential answers to 5 for highly relevant and contextually appropriate responses. These criteria ensured a systematic evaluation process aligned with the study's objectives and theoretical underpinnings, offering a structured approach to assess ChatGPT's performance comprehensively. The evaluation considered the amalgamation of the best available linguistic evidence and academic expertise.

### 3.3 Addressing Inter-Rater Reliability

Before commencing the scoring process, a calibration session was held to establish a shared understanding of the scoring system among the evaluators, reducing subjectivity and enhancing consistency. The study employed a crowd score strategy, as the evaluation of ChatGPT's responses lacked an established "ground truth" and was inherently subjective. Emphasis was placed on the median scores assigned by the evaluators for each answer. To further address inter-rater reliability concerns, Fleiss's kappa coefficient (Fleiss, 1971) was computed to determine the level of agreement among evaluators, with results interpreted based on Landis and Koch's (1977) guidelines, categorizing agreement levels as slight, fair, moderate, substantial, or almost perfect. Despite some variability in ratings, this method ensured a systematic and evidence-based approach to evaluating ChatGPT's responses.

### 3.4 Data Analysis Procedure

The scores provided by the evaluators were recorded in Microsoft Excel and analyzed using the Statistical Package for Social Sciences v. 25 (SPSS) software. Each question's median, interquartile range (IQR), and full score range were calculated. Evaluators were assigned random identifiers, and Fleiss's kappa was used to assess the consistency of scores among them (Fleiss, 1971). Moreover, the reliability of the questionnaire, which included the questions, was measured using Cronbach's alpha. The choice of statistical tests was based on the nature of the collected data. The Kruskal–Wallis H test was used to identify any differences in scores among the evaluators due to its suitability for non-parametric data and independent samples (Corder and Foreman, 2011; Kruskal and Wallis, 1952). All statistical analyses were conducted with a significance level of  $p < 0.05$ . Additionally, when performing post-hoc pairwise comparisons, the Bonferroni correction was applied to manage multiple testing and reduce the likelihood of Type I errors. This structured methodology ensures that the study maintains high levels of validity and reliability, providing a robust assessment of ChatGPT's performance in answering SLA-related questions.

## 4. Results

Table 1 presents the questions and answers generated by ChatGPT, alongside the median, interquartile range, and overall range of scores provided by evaluators.

**Table 1: The queries and answers generated by ChatGPT, along with a summary of evaluators scores**

| Topics               | N | Questions generated by ChatGPT 4.0                                         | Answers generated by ChatGPT 4.0                                                                                                                                                                                                                                                                                                                                                                                                                              | Median | IQR | Range   |
|----------------------|---|----------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----|---------|
| Key questions in SLA | 1 | How did the field of second language acquisition (SLA) research originate? | The field of second language acquisition (SLA) research originated as scholars began to systematically investigate the processes and factors involved in individuals acquiring a second language. This development gained momentum in the mid-20th century, with linguists and psychologists exploring various aspects of language learning and its underlying mechanisms.                                                                                    | 4.0    | 2.0 | 3.0-5.0 |
|                      | 2 | Describe the stages and progression involved in the development of SLA?    | The development of SLA involves distinct stages and progression, beginning with the initial exposure to second language (L2). Initial exposure leads to the establishment of rudimentary linguistic competence, with subsequent stages marked by syntactic and morphological refinement, expanding lexicon, and pragmatic development, followed by stages of comprehension, production, and eventual mastery. These stages are shaped by factors such as age, | 4.0    | 2.0 | 3.0-5.0 |

| Topics | N | Questions generated by ChatGPT 4.0                                                  | Answers generated by ChatGPT 4.0                                                                                                                                                                                                                                                                                                                                                                                                                        | Median | IQR | Range   |
|--------|---|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----|---------|
|        |   |                                                                                     | exposure, and individual differences, contributing to the complexity of the SLA process.                                                                                                                                                                                                                                                                                                                                                                |        |     |         |
|        | 3 | What functions do input and output serve in the process of SLA?                     | Input and output play crucial roles in the process of SLA. Input, as articulated by Stephen Krashen's Input Hypothesis, in the form of exposure to L2, provides the linguistic material for learners, while output posited by Merrill Swain's Output Hypothesis, or the production of language, allows learners to practice and internalize linguistic structures. Both input and output contribute to the gradual development of language proficiency. | 5.0    | 1.0 | 4.0-5.0 |
|        | 4 | What characterizes the initial state in SLA?                                        | The initial state in SLA is characterized by a learner's starting point in acquiring a second language. Factors such as the learner's prior linguistic knowledge, cognitive abilities, and exposure to L2 contribute to shaping this initial state.                                                                                                                                                                                                     | 3.0    | 1.5 | 3.0-4.5 |
|        | 5 | Is it possible for second language (L2) learners to attain native-like proficiency? | While achieving native-like proficiency is challenging, research suggests that second language (L2) learners can attain a high level of proficiency. However, various factors, including age of acquisition, intensity of exposure, and individual differences, influence the extent to which learners can approach native-like competence.                                                                                                             | 5.0    | 1.0 | 4.0-5.0 |
|        | 6 | How does instructional intervention impact SLA?                                     | Instructional intervention can significantly impact SLA. Effective teaching methods, feedback, and exposure to meaningful language input can enhance the learning process and accelerate language acquisition. However, the effectiveness of instructional methodologies is contingent upon the alignment of pedagogical strategies with learner variables, including motivation, cognitive styles, and prior linguistic experience.                    | 4.0    | 1.0 | 3.5-4.5 |
|        | 7 | What significance does explicit learning have in SLA?                               | Explicit learning, involving conscious awareness of linguistic rules and structures, is recognized as a supplementary mechanism in SLA. While implicit learning remains primary, explicit instruction can expedite the acquisition of intricate grammatical features and foster metalinguistic awareness.                                                                                                                                               | 4.0    | 1.5 | 3.5-5.0 |
|        | 8 | What are individual differences, and how do they influence SLA?                     | Individual differences, encompassing variables such as cognitive aptitude, motivation, and learning strategies, exert discernible effects on the trajectory of SLA. These factors interplay with instructional contexts, shaping the pace and extent of linguistic development among L2 learners. Understanding and accommodating these individual differences are imperative for tailored pedagogical approaches                                       | 5.0    | 1.0 | 4.0-5.0 |
|        |   |                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                         |        |     |         |
|        | 1 | How has technology impacted SLA in recent years?                                    | Technology has revolutionized second language acquisition by providing interactive and immersive learning experiences. Online resources, language learning apps, and virtual communication tools offer learners access to authentic language use and diverse cultural content.                                                                                                                                                                          | 5.0    | 1.0 | 4.0-5.0 |

| Topics                        | N | Questions generated by ChatGPT 4.0                                                                                  | Answers generated by ChatGPT 4.0                                                                                                                                                                                                                                                                                                   | Median | IQR | Range   |
|-------------------------------|---|---------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----|---------|
| The role of technology in SLA | 2 | What are the advantages of incorporating language learning apps in SLA?                                             | Language learning apps provide flexibility, allowing learners to practice at their own pace. They often incorporate gamification elements, making the learning process engaging. Apps also offer instant feedback and personalized learning experiences.                                                                           | 5.0    | 1.0 | 4.0-5.0 |
|                               | 3 | How can virtual language exchanges enhance SLA?                                                                     | Virtual language exchanges enable learners to practice with native speakers or peers from around the world through online platforms. This fosters real-life communication, cultural exchange, and the application of language skills in authentic contexts                                                                         | 4.0    | 1.0 | 4.0-5.0 |
|                               | 4 | In what ways can multimedia content contribute to SLA?                                                              | Multimedia content, such as videos, podcasts, and interactive simulations, offers authentic language input with contextual cues. This aids in the development of listening and comprehension skills while exposing learners to diverse accents and language registers                                                              | 5.0    | 1.5 | 3.5-5.0 |
|                               | 5 | What role do online language communities play in supporting SLA?                                                    | Online language communities provide a platform for learners to connect, share experiences, and practice the target language. These communities facilitate collaborative learning, cultural exchange, and the development of a supportive language-learning network.                                                                | 5.0    | 0.5 | 4.5-5.0 |
|                               | 6 | How can virtual reality (VR) technology benefit second language learners?                                           | VR technology creates immersive language learning environments, allowing learners to virtually visit and interact in target language settings. This enhances cultural understanding, situational language use, and overall language proficiency through realistic simulations.                                                     | 5.0    | 1.0 | 4.0-5.0 |
|                               | 7 | What considerations should educators keep in mind when integrating technology into language classrooms?             | Educators should consider the accessibility of technology, learner preferences, and the alignment of digital tools with pedagogical objectives. Additionally, ensuring a balance between technology and traditional teaching methods is essential for an effective language learning environment.                                  | 5.0    | 1.5 | 3.5-5.0 |
|                               | 8 | Are there any potential challenges associated with technology use in SLA?                                           | Challenges may include issues related to digital access, unequal technology resources, and potential distractions. Additionally, maintaining a balance between technology and face-to-face interaction, and ensuring the quality of online content, are considerations for educators and learners alike.                           | 4.0    | 1.0 | 4.0-5.0 |
|                               |   |                                                                                                                     |                                                                                                                                                                                                                                                                                                                                    |        |     |         |
| Reading in SLA                | 1 | How does the process of reading contribute to SLA, and what specific skills are involved in L2 reading development? | The process of reading contributes significantly to SLA by exposing learners to authentic language use, expanding vocabulary, and enhancing overall language proficiency. Specific skills involved in L2 reading development include decoding, comprehension, inference, and the ability to analyze and critically evaluate texts. | 4.0    | 1.0 | 3.5-4.5 |
|                               | 2 | What role do phonics and decoding skills play in the early stages of L2 reading acquisition?                        | Phonics and decoding skills play a crucial role in the early stages of L2 reading acquisition. These skills help learners recognize and decipher unfamiliar words, facilitating the development of a strong                                                                                                                        | 5.0    | 1.5 | 3.5-5.0 |

| Topics         | N | Questions generated by ChatGPT 4.0                                                                                                                                          | Answers generated by ChatGPT 4.0                                                                                                                                                                                                                                                                                                                  | Median | IQR | Range   |
|----------------|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----|---------|
|                |   |                                                                                                                                                                             | foundation for understanding written language.                                                                                                                                                                                                                                                                                                    |        |     |         |
|                | 3 | How can educators tailor reading instruction to address the diverse needs and proficiency levels of L2 learners?                                                            | Educators can tailor reading instruction for L2 learners by differentiating content based on proficiency levels, providing culturally relevant texts, offering additional support for vocabulary development, and incorporating varied instructional approaches, such as collaborative reading activities and language-rich discussions.          | 4.0    | 2.0 | 3.0-5.0 |
|                | 4 | What impact does extensive reading, such as independent reading of various texts, have on vocabulary acquisition and overall language proficiency in L2 learners?           | Extensive reading, characterized by independent reading of a wide range of texts, has a positive impact on vocabulary acquisition and overall language proficiency in L2 learners. Exposure to diverse materials helps learners encounter new words in context and reinforces language patterns.                                                  | 4.0    | 1.0 | 4.0-5.0 |
|                | 5 | How do different reading comprehension strategies, such as summarization and inferencing, influence the development of higher-order thinking skills in SLA?                 | Different reading comprehension strategies, such as summarization and inferencing, contribute to the development of higher-order thinking skills in SLA. These strategies encourage learners to engage deeply with the text, fostering critical analysis and interpretation.                                                                      | 4.0    | 2.5 | 2.0-4.5 |
|                | 6 | What challenges do L2 learners commonly encounter in the process of reading, and how can educators mitigate these challenges to promote successful language acquisition?    | L2 learners commonly face challenges in reading, including vocabulary gaps, cultural unfamiliarity, and difficulty with complex syntax. Educators can mitigate these challenges by providing explicit vocabulary instruction, incorporating cultural awareness activities, and scaffolding reading tasks to support comprehension.                | 5.0    | 1.0 | 4.0-5.0 |
|                | 7 | How does the integration of authentic and culturally relevant texts enhance the effectiveness of reading instruction in SLA?                                                | The integration of authentic and culturally relevant texts enhances the effectiveness of reading instruction in SLA by making the content more engaging and meaningful for learners. Exposure to real-world language use and cultural contexts aids in comprehension and language internalization.                                                | 5.0    | 0.5 | 4.5-5.0 |
|                | 8 | In what ways can technology, such as digital platforms and online resources, be leveraged to support and supplement L2 reading instruction for diverse learner populations? | Technology, such as digital platforms and online resources, can be leveraged to support and supplement L2 reading instruction for diverse learner populations. Interactive e-books, language learning apps, and online reading materials offer opportunities for independent practice, immediate feedback, and personalized learning experiences. | 5.0    | 1.5 | 3.5-5.0 |
|                |   |                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                   |        |     |         |
| Writing in SLA | 1 | How does the process of writing contribute to SLA, and what specific skills are involved in the development of writing proficiency in L2?                                   | The process of writing significantly contributes to SLA by providing learners with opportunities to practice and apply language skills. Specific skills involved in the development of writing proficiency in L2 include grammar usage, vocabulary selection, sentence structure, and organizational coherence.                                   | 4.0    | 2.0 | 3.0-5.0 |
|                | 2 | What role does grammar and syntax play in the early                                                                                                                         | In the early stages of L2 writing acquisition, grammar and syntax play a foundational role in establishing basic communication                                                                                                                                                                                                                    | 4.0    | 2.0 | 3.0-5.0 |

| Topics | N | Questions generated by ChatGPT 4.0                                                                                                                                                            | Answers generated by ChatGPT 4.0                                                                                                                                                                                                                                                                                                                                     | Median | IQR | Range   |
|--------|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----|---------|
|        |   | stages of L2 writing acquisition, and how do these skills evolve over time?                                                                                                                   | skills. These skills evolve over time as learners progress, moving from simple sentence constructions to more complex structures, enabling them to express more nuanced ideas and arguments                                                                                                                                                                          |        |     |         |
|        | 3 | How can educators effectively teach and assess writing skills to address the diverse needs and proficiency levels of L2 learners?                                                             | Educators can effectively teach and assess writing skills in L2 learners by implementing differentiated instruction that considers diverse proficiency levels. Strategies may include providing clear writing prompts, offering targeted feedback, and incorporating peer review processes to address individual needs.                                              | 5.0    | 0.5 | 4.5-5.0 |
|        | 4 | What impact does extensive writing practice, including essay composition and creative writing, have on vocabulary acquisition and overall language proficiency in L2 learners?                | Extensive writing practice, including essay composition and creative writing, has a positive impact on vocabulary acquisition and overall language proficiency in L2 learners. Engaging in varied writing tasks allows learners to experiment with language, expand their lexicon, and apply grammar rules in context.                                               | 5.0    | 1.5 | 3.5-5.0 |
|        | 5 | How do different writing strategies, such as pre-writing planning and revision techniques, influence the development of higher-order thinking skills in SLA?                                  | Different writing strategies, such as pre-writing planning and revision techniques, play a crucial role in the development of higher-order thinking skills in SLA. These strategies encourage learners to organize their thoughts, analyze content critically, and refine their written expression                                                                   | 5.0    | 1.0 | 4.0-5.0 |
|        | 6 | What challenges do L2 learners commonly encounter in the process of writing, and how can educators provide targeted support to address these challenges?                                      | L2 learners commonly encounter challenges in writing, including difficulty with grammatical accuracy, vocabulary limitations, and struggles with text coherence. Educators can provide targeted support through explicit instruction, individualized feedback, and scaffolded writing tasks                                                                          | 4.0    | 1.0 | 4.0-5.0 |
|        | 7 | How does the incorporation of authentic and culturally relevant writing tasks enhance the effectiveness of writing instruction in SLA?                                                        | The incorporation of authentic and culturally relevant writing tasks enhances the effectiveness of writing instruction in SLA by making the content more relatable and engaging. Authentic tasks encourage learners to connect language use with real-world contexts, promoting meaningful expression                                                                | 4.0    | 1.5 | 3.5-5.0 |
|        | 8 | In what ways can technology, such as word processing tools and collaborative online platforms, be leveraged to support and supplement L2 writing instruction for diverse learner populations? | Technology, such as word processing tools and collaborative online platforms, can be leveraged to support and supplement L2 writing instruction for diverse learner populations. These tools provide opportunities for collaborative writing, immediate feedback, and the integration of multimedia elements to enhance language learning experiences                | 3.0    | 3.0 | 2.0-5.0 |
|        |   |                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                      |        |     |         |
|        | 1 | How does the development of listening skills contribute to overall SLA, and what specific aspects of listening proficiency are critical for language learners?                                | The development of listening skills significantly contributes to overall SLA by enabling learners to comprehend spoken language, recognize linguistic patterns, and enhance communicative competence. Critical aspects of listening proficiency for language learners include accurate comprehension of spoken content, recognition of diverse accents and dialects, | 5.0    | 1.0 | 4.0-5.0 |

| Topics           | N | Questions generated by ChatGPT 4.0                                                                                                                                                | Answers generated by ChatGPT 4.0                                                                                                                                                                                                                                                                                                                                                                                                      | Median | IQR | Range   |
|------------------|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----|---------|
| Listening in SLA |   |                                                                                                                                                                                   | and the ability to infer meaning from context.                                                                                                                                                                                                                                                                                                                                                                                        |        |     |         |
|                  | 2 | What role does exposure to authentic spoken language play in the early stages of L2 listening acquisition, and how does it impact learners at different proficiency levels?       | Exposure to authentic spoken language plays a crucial role in the early stages of L2 listening acquisition. It exposes learners to natural language use, intonation, and pronunciation, fostering a deeper understanding of spoken communication. The impact of exposure varies at different proficiency levels, influencing beginners in basic comprehension and more advanced learners in nuanced understanding and interpretation. | 5.0    | 1.5 | 3.5-5.0 |
|                  | 3 | How can educators design listening activities that cater to the diverse needs and proficiency levels of L2 learners?                                                              | Educators can design listening activities for L2 learners that cater to diverse needs and proficiency levels by incorporating a range of difficulty levels, diverse accents, and varied topics. Tailoring activities to individual learning styles and providing opportunities for repetition and reinforcement help address the unique challenges learners may face.                                                                 | 5.0    | 1.0 | 4.0-5.0 |
|                  | 4 | What influence does active engagement with various audio materials, such as podcasts or language learning apps, have on vocabulary acquisition and comprehension in L2 listening? | Active engagement with various audio materials, such as podcasts or language learning apps, positively influences vocabulary acquisition and comprehension in L2 listening. Exposure to real-world language use in context enhances vocabulary retention and allows learners to apply their knowledge in practical situations.                                                                                                        | 5.0    | 1.5 | 3.5-5.0 |
|                  | 5 | How do different listening comprehension strategies, such as prediction and summarization, contribute to the development of higher-order thinking skills in SLA?                  | Different listening comprehension strategies, such as prediction and summarization, contribute to the development of higher-order thinking skills in SLA. These strategies encourage learners to engage critically with spoken content, enhancing their ability to analyze, synthesize, and evaluate information.                                                                                                                     | 5.0    | 1.5 | 3.5-5.0 |
|                  | 6 | What common challenges do L2 learners encounter in the process of listening, and how can instructional approaches be adapted to address these challenges effectively?             | Common challenges encountered by L2 learners in the process of listening include difficulty with rapid speech, unfamiliar accents, and processing complex linguistic structures. Instructional approaches can be adapted by providing pre-listening activities, offering explicit strategies for comprehension, and gradually increasing the complexity of listening tasks.                                                           | 5.0    | 0.5 | 4.5-5.0 |
|                  | 7 | How does the integration of culturally relevant and authentic listening materials enhance the effectiveness of instruction in SLA?                                                | The integration of culturally relevant and authentic listening materials enhances the effectiveness of instruction in SLA by making the content more engaging and meaningful. Learners connect language use with real-world contexts, fostering a deeper appreciation of cultural nuances and communicative purposes.                                                                                                                 | 4.0    | 1.5 | 3.5-5.0 |
|                  | 8 | In what ways can technology, such as digital platforms and online resources, be utilized to support and augment L2 listening instruction for diverse learner populations?         | Technology, such as digital platforms and online resources, can be utilized to support and augment L2 listening instruction for diverse learner populations. Interactive tools, language learning apps, and online platforms offer opportunities for personalized learning experiences, allowing learners to practice listening skills in varied contexts and at their own pace.                                                      | 5.0    | 1.5 | 3.5-5.0 |

| Topics          | N | Questions generated by ChatGPT 4.0                                                                                                                                            | Answers generated by ChatGPT 4.0                                                                                                                                                                                                                                                                                                                                                                           | Median | IQR | Range   |
|-----------------|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----|---------|
|                 |   |                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                            |        |     |         |
| Speaking in SLA | 1 | How does the development of speaking skills contribute to overall SLA, and what specific aspects of speaking proficiency are critical for language learners                   | The development of speaking skills significantly contributes to overall SLA by providing learners with the ability to express themselves verbally, engage in effective communication, and participate actively in various social contexts. Critical aspects of speaking proficiency for language learners include pronunciation, fluency, vocabulary usage, and the ability to convey complex ideas.       | 5.0    | 1.0 | 4.0-5.0 |
|                 | 2 | What role does pronunciation play in the early stages of L2 speaking acquisition, and how does it impact learners at different proficiency levels?                            | Pronunciation plays a crucial role in the early stages of L2 speaking acquisition as it influences the clarity of communication. Learners at different proficiency levels may experience challenges ranging from basic sound production to mastering intonation and rhythm, impacting their overall communicative competence.                                                                              | 5.0    | 2.0 | 3.0-5.0 |
|                 | 3 | How can educators design speaking activities that cater to the diverse needs and proficiency levels of L2 learners?                                                           | Educators can design speaking activities for L2 learners that cater to diverse needs and proficiency levels by incorporating varied topics, real-life scenarios, and interactive tasks. Differentiating instruction based on learners' strengths and weaknesses, providing opportunities for peer collaboration, and offering a supportive environment contribute to effective speaking skill development. | 5.0    | 1.5 | 3.5-5.0 |
|                 | 4 | What influence does regular practice and interaction, such as participating in conversations or language exchange, have on vocabulary acquisition and fluency in L2 speaking? | Regular practice and interaction, such as participating in conversations or language exchange, have a positive influence on vocabulary acquisition and fluency in second language speaking. Engaging in authentic communication fosters vocabulary retention, allows learners to experiment with language use, and enhances overall speaking proficiency.                                                  | 5.0    | 1.5 | 3.5-5.0 |
|                 | 5 | How do different speaking strategies, such as role-playing and storytelling, contribute to the development of higher-order thinking skills in SLA?                            | Different speaking strategies, such as role-playing and storytelling, contribute to the development of higher-order thinking skills in SLA. These activities encourage learners to think critically, organize their thoughts coherently, and express complex ideas, promoting cognitive growth.                                                                                                            | 5.0    | 1.5 | 3.5-5.0 |
|                 | 6 | What common challenges do L2 learners encounter in the process of speaking, and how can instructional approaches be adapted to address these challenges effectively?          | Common challenges faced by L2 learners in the process of speaking include anxiety, hesitation, and difficulty with spontaneous expression. Instructional approaches can be adapted by incorporating supportive activities, providing opportunities for low-pressure practice, and gradually increasing the complexity of speaking tasks.                                                                   | 4.0    | 1.0 | 4.0-5.0 |
|                 | 7 | How does the integration of culturally relevant and authentic speaking tasks enhance the effectiveness of instruction in SLA?                                                 | The integration of culturally relevant and authentic speaking tasks enhances the effectiveness of instruction in second language acquisition by making the content more engaging and meaningful. Learners connect language use with real-world contexts, fostering a deeper appreciation of cultural nuances and communicative purposes.                                                                   | 5.0    | 1.5 | 3.5-5.0 |

| Topics | N | Questions generated by ChatGPT 4.0                                                                                                                                                             | Answers generated by ChatGPT 4.0                                                                                                                                                                                                                                                                                                                                                         | Median | IQR | Range   |
|--------|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----|---------|
|        | 8 | In what ways can technology, such as language learning apps and collaborative online platforms, be utilized to support and supplement L2 speaking instruction for diverse learner populations? | Technology, such as language learning apps and collaborative online platforms, can be utilized to support and supplement L2 speaking instruction for diverse learner populations. Interactive tools, video conferencing, and language exchange platforms offer opportunities for personalized practice, immediate feedback, and engagement with a global community of language learners. | 5.0    | 1.0 | 4.0-5.0 |

Generally, the evaluators rated ChatGPT as providing valuable information on the assessed topics: key questions in SLA=  $4.200 \pm 0.823$ , the role of technology in SLA=  $4.525 \pm 0.640$ , reading in SLA=  $4.250 \pm 0.927$ , writing in SLA=  $4.225 \pm 0.974$ , listening in SLA=  $4.475 \pm 0.751$  and speaking in SLA=  $4.425 \pm 0.813$ . The total median score for all main topics was 4.7, showing no statistical difference among them ( $p > 0.05$ ) as indicated in Table 2.

Table 2 displays descriptive statistics for various aspects of SLA, including key questions, the role of technology, reading, writing, listening, and speaking. The data for the writing in SLA category showed greater variability in scores.

**Table 2: Descriptive statistics regarding SLA related issues**

| Topics                        | Mean  | Standard deviation | Median | P value |
|-------------------------------|-------|--------------------|--------|---------|
| Key questions in SLA          | 4.200 | 0.823              | 4.0    | 0.210   |
| The role of technology in SLA | 4.525 | 0.640              | 5.0    |         |
| Reading in SLA                | 4.250 | 0.927              | 4.5    |         |
| Writing in SLA                | 4.225 | 0.974              | 5.0    |         |
| Listening in SLA              | 4.475 | 0.751              | 5.0    |         |
| Speaking in SLA               | 4.425 | 0.813              | 5.0    |         |

Statistical difference for  $p < 0.05$ .

Table 3 displays the distribution of scores assigned by evaluators, ranging from "very poor" to "very good." The overall findings reveal that the highest percentage of scores in the entire dataset fell within the "Very Good" category (55.41%), followed by "Good" (26.25.1%) and "Acceptable" (16.67%). Conversely, the categories of "Poor" and "Very Poor" had the lowest percentages: 1.25% and 0.42%, respectively.

**Table3: Total count (n) and proportion (%) of median scores assigned by each evaluator**

| Evaluators |     | (1) Very Poor | (2) Poor | (3) Acceptable | (4) Good | (5) Very Good |
|------------|-----|---------------|----------|----------------|----------|---------------|
| 1          | n   | 0             | 0        | 6              | 8        | 34            |
|            | (%) | 0.0%          | 0.00%    | 12.50%         | 16.67%   | 70.83%        |
| 2          | n   | 0             | 1        | 8              | 20       | 19            |
|            | (%) | 0.0%          | 2.08%    | 16.67%         | 41.67%   | 39.58%        |
| 3          | n   | 0             | 0        | 0              | 9        | 39            |
|            | (%) | 0.0%          | 0.0%     | 0.0%           | 18.75%   | 81.25%        |
| 4          | n   | 0             | 0        | 1              | 10       | 37            |
|            | (%) | 0.0%          | 0.0%     | 2.08%          | 20.83%   | 77.09%        |
| 5          | n   | 1             | 2        | 25             | 16       | 4             |
|            | (%) | 2.08%         | 4.18%    | 52.08%         | 33.33%   | 8.33%         |
| Total      | n   | 1             | 3        | 40             | 63       | 133           |
|            | (%) | 0.42%         | 1.25%    | 16.67%         | 26.25%   | 55.41%        |

These findings indicate that most evaluators assigned scores within the higher quality categories (Good and Very Good), suggesting a generally positive evaluation of the topics. The low percentages in the "Poor" and "Very Poor" categories suggest that the topics were perceived to be reasonable and satisfactory in quality by the evaluators. These lower ratings primarily stemmed from ChatGPT's oversimplifications, lack of depth in complex SLA concepts, and occasional inaccuracies in theoretical explanations. For example, responses regarding the characterization of the "initial state in SLA" demonstrated a tendency toward broad generalizations, failing to differentiate adequately between competing theoretical perspectives, such as Universal Grammar-based models and emergentist approaches. Additionally, ChatGPT's responses on leveraging technology in writing instruction were noted to lack specificity, often providing generic recommendations rather than engaging with established research on the pedagogical effectiveness of AI-driven writing assistants. Another area of concern was the response to questions about fossilization in SLA, where ChatGPT provided a surface-level definition but omitted key discussions on how various linguistic, cognitive, and sociocultural factors interact to contribute to fossilization processes. These findings highlight specific gaps in ChatGPT's capabilities and underscore the need for its enhancement in handling nuanced and research-intensive queries. The qualitative observations reinforce the statistical results, indicating that while ChatGPT performs well in providing general explanations, it struggles with depth, precision, and the integration of competing theoretical viewpoints.

In terms of the scores given by evaluators, there was a minimal level of agreement observed (Landis and Koch, 1977), with a combined Fleiss's Kappa coefficient of 0.006. This discrepancy reflects diverse interpretations of the evaluation criteria and underscores the subjective nature of assessing AI-generated content. The Kruskal–Wallis test revealed a notable inconsistency in scores among the evaluators ( $p < 0.001$ ), which has been attributed to differences in individual expertise, perspectives, and familiarity with SLA-specific issues. A comprehensive pairwise comparison of the evaluators' scores is provided in Table 4.

**Table 4: Pairwise comparison of evaluators scores.**

| Evaluators | Bonferroni pairwise comparison |         |
|------------|--------------------------------|---------|
|            | t                              | p-value |
| 1vs.2      | 2.817                          | 0.028   |
| 1vs.3      | -2.223                         | 0.428   |
| 1vs.4      | -1.359                         | 0.725   |
| 1vs.5      | -1.248                         | 0.824   |
| 2vs.3      | -4.916                         | 0.000   |
| 2vs.4      | -4.391                         | 0.000   |
| 2vs.5      | 2.158                          | 0.591   |
| 3vs.4      | 0.658                          | 0.990   |
| 3vs.5      | -1.723                         | 0.482   |
| 4vs.5      | -1.969                         | 0.539   |

Statistical difference for  $p < 0.05$ .

## 5. Discussion

*Research Question 1: How reliable are the responses generated by ChatGPT in addressing key questions in SLA?*

This article investigated how ChatGPT, a sophisticated AI, is currently being used, with a focus on how accurately and efficiently it responds to questions about SLA, the impact of technology, and language skills including reading, writing, listening, and speaking in SLA. In this research, the chatbot quickly provided answers to all questions within seconds. These responses were later evaluated by experienced linguists. The present study represents a groundbreaking cross-sectional evaluation of ChatGPT's precision in addressing contemporary linguistic questions.

The data reflects positive evaluation of responses presented by ChatGPT from the experts in SLA participated in this study. The assessment of ChatGPT's comprehension and interaction indicates a vastly accurate and high-quality information provided by this NLP model regarding various domains in SLA. This finding aligns with several studies that have shown that AI exhibited impressive performance, providing accurate responses across various

topics as evaluated by experts from different academic disciplines, albeit with significant constraints (Firat, 2023; Huallpa et al., 2023; Lund and Wang, 2023). While AI is not a recent innovation, ChatGPT has gained widespread popularity and acceptance (Kalla, Smith and Carolina, 2023; Liu et al., 2023; Ray, 2023). As early mentioned in this paper, various forms of AI have been extensively utilized across disciplines such as management, science, biological and medical diagnostics as well as education (Baskara and Mukarto, 2023; Kalla, Smith and Carolina, 2023; Kalla, Smith and Carolina, 2023; Lund and Wang, 2023; Ray, 2023). These AIs can be trained using diverse datasets and applied to a multitude of tasks within different domains (Lund and Wang, 2023; Makaremi, Lacaule and Mohammad-Djafari, 2019; Ray, 2023; Sharma and Yadav, 2023).

However, the reliability of using these human-like applications has been the concern of several studies across various disciplines including language learning fields. One particular study, (Chaka, 2023), concluded that AI chatbots are currently not considered trustworthy or reliable sources of information within the realm of applied English language studies (AELS). Hence, there is a call for serious work for adopting strategies that ensure effective use of such models in language education (for instances Al-khresheh, 2024; Barrot, 2023; Karataş et al., 2024). Various researchers (e.g., Baskara and Mukarto, 2023; Huallpa et al., 2023; Sallam, 2023) also assert the need to conduct more research and develop methods that equip learners with the tools they need for efficient use of these models and enable useful implementing for such new technologies across different educational contexts.

*Research Question 2: To what extent does ChatGPT provide accurate and meaningful information on specific language skills (reading, writing, listening, and speaking) in SLA?*

It is noteworthy that previous studies did not address the realm of SLA. In this study, the vast majority of responses were deemed highly satisfactory by the five evaluators. The responses that received the lowest scores were “Q: Key questions in SLA: What characterizes the initial state in SLA? / A: The initial state in SLA is characterized by a learner's starting point in acquiring a second language. Factors such as the learner's prior linguistic knowledge, cognitive abilities, and exposure to L2 contribute to shaping this initial state.” (median: 3.0) and “Q: Writing in SLA: In what ways can technology, such as word processing tools and collaborative online platforms, be leveraged to support and supplement L2 writing instruction for diverse learner populations? / A: Technology, such as word processing tools and collaborative online platforms, can be leveraged to support and supplement L2 writing instruction for diverse learner populations. These tools provide opportunities for collaborative writing, immediate feedback, and the integration of multimedia elements to enhance language learning experiences.” (median: 3.0).

As can be noted from the data, responses from ChatGPT are straightforward, and a linguist with poor training and lacking linguistic expertise may mistakenly perceive all answers as accurate, potentially leading to misinformation. Conversely, linguists with proper training can leverage their academic background to expand or adapt existing theories to accommodate new functions and contexts. However, when dealing solely with written text from ChatGPT, the situation becomes more complex and without experts to discern truth from falsehood, we lack the intermediary needed to navigate intricate subjects effectively (Barrot, 2023; Huallpa et al., 2023; Lund and Wang, 2023; Tanaka et.al., 2023). While large linguistic algorithms excel in knowledge-based assessments, they often struggle with language instruction, learning topics, and literature (Barrot, 2023; Baskara and Mukarto, 2023). To enhance their performance, these AI models require training on high-quality datasets (Huallpa et al., 2023; Liu et al., 2023; Sallam, 2023; Ray, 2023). Nonetheless, their current training on possibly biased datasets could account for the inaccuracies observed when responding to specific research-related queries.

*Research Question 3: What are the limitations and areas for improvement in ChatGPT's performance as evaluated by experts in SLA?*

Hence, considering that AI is not a novel concept, what has propelled this recent advancement into the mainstream? The intriguing idea of a textbot capable of proficiently generating content on a wide range of topics may be attributed to human curiosity and desire to maximize knowledge (Lund & Wang, 2023; Karataş et al., 2024; Ray, 2023; Sun, 2023). Despite ChatGPT providing robust answers across the six examined subjects, it's crucial to note that this AI learns from an extensive text dataset derived from books, articles, and web pages. Linguists and L2 learners seek more precise responses, but ChatGPT incorporates both accurate scientific information and potentially misleading content found in advertisements, social media, and websites (Barrot, 2023; Baskara and Mukarto, 2023; Huallpa et al., 2023).

Still, ChatGPT, an AI chatbot, effectively summarizes information, generates sophisticated text, and presents itself convincingly (Baskara and Mukarto, 2023; Liu et al., 2023; Lund and Wang, 2023). In this study, its linguistic responses demonstrated a general accuracy: among 240 answers, most were categorized as either very good (55.41%) or good (26.25%). Yet, it requires improvement in accuracy regarding issues related to SLA. The present findings were broad, realistic, and comprehensive, showcasing a solid understanding of the subject matter, but without delving into specific details. However, because ChatGPT functions differently from a search engine, the accuracy of its responses is contingent upon the accuracy of the source information (Barrot, 2023; Liu et al., 2023; Tanaka et al., 2023). While ChatGPT generally delivers accurate responses regarding SLA, it encounters limitations such as: (1) its inability to critically evaluate or analyze findings from previous literature, (2) reliance on a knowledge base limited to data up to 2020 with no updates, (3) occasional misinterpretation of various theories, (4) incapacity to discern between reputable and predatory journal sources, and (5) concerns about scientific precision, potential biases, and the distribution of misinformation to users.

At this point, AI should be viewed as a supplementary tool rather than a replacement for linguists. Researchers like Sharma and Yadav, (2023) and Karataş et al. (2024) emphasize the extent to which AI applications contribute to education that is complementing rather than replacing the facilitator role of humans. The utilization of ChatGPT can be seen as unavoidable, yet linguists must actively pursue theoretical and scientific validation. Although there is considerable interest in utilizing ChatGPT, it lacks extensive training on language acquisition data and specific responses. It is probable that second language learners will resort to ChatGPT for guidance in interpreting data and grasping technological and various language learning strategies. While ChatGPT is perceived as a potentially advantageous aid in language learning environments, apprehensions persist regarding its precision, dependability, and educational implications (Al-khresheh, 2024; Barrot, 2023; Baskara and Mukarto, 2023; Huang, Hew and Fryer, 2022; Karataş et al., 2024).

In sum, ChatGPT has demonstrated impressive proficiency in the evaluated language acquisition areas. However, variations in individual evaluations underscore the inherent complexity and subjectivity of any assessment process. It is imperative for L2 learners and linguists to acknowledge the limitations and ethical considerations associated with ChatGPT, and to consistently validate information using reliable sources. Prior to integrating these AI models into the fields of L2 learning and teaching, concerted efforts are necessary to improve their reliability and to enhance teachers' and learners' awareness about ethical and technical issues that must be considered when implementing such NLP models into education (Al-khresheh, 2024; Chaka, 2023; Huang, Hew and Fryer, 2022; Karataş et al., 2024; Sharma and Yadav, 2023).

## **6. Conclusion and Implications**

This study explored ChatGPT's accuracy, relevance, and reliability in addressing foundational and contemporary topics in SLA, demonstrating its potential as an interactive tool in language education. While ChatGPT produced coherent and contextually relevant responses, limitations such as inconsistencies in evaluator ratings and the absence of expert validation during question design highlighted areas for improvement. These findings underscore the importance of robust research methodologies when evaluating AI tools in educational contexts. The study's implications suggest that AI can serve as a valuable supplement in language learning, providing support for complex SLA concepts. However, educators and researchers must approach its application cautiously, ensuring it complements rather than replaces traditional methods. Future research should prioritize expert validation and inter-rater reliability analyses to enhance methodological rigor and better harness AI's pedagogical potential in SLA. Future research could also explore practical applications of AI tools, such as ChatGPT, in classroom settings, curriculum design, and teacher training programs. Additionally, providing detailed guidelines on how AI-generated content can complement traditional SLA methods would offer educators concrete strategies for integration. These steps aim to bridge the gap between theoretical insights and practical applications, making the study's contributions more impactful for the field.

## **7. Limitations and Recommendations**

Despite its contributions, this study has several limitations that should be addressed in future research. One key limitation is the lack of expert validation for the questions used in the study. While the questions were designed to cover foundational aspects of Second Language Acquisition (SLA) and reflect contemporary trends in SLA discourse, expert input could have ensured greater accuracy and relevance, thereby strengthening the study's methodological rigor. Future research should incorporate expert validation in the question design phase to enhance the precision and reliability of assessment tools.

Another limitation is the absence of an inter-rater reliability analysis prior to the evaluation process. Given that the evaluators had different levels of experience with AI-based tools, the consistency of their ratings could have been improved with such an analysis. Although Fleiss's Kappa coefficient revealed minimal agreement among evaluators, this inconsistency highlights a concern that future studies could address. To mitigate evaluator variability, future research should implement pre-evaluation calibration sessions to align evaluators' expectations and criteria, ensuring a more standardized rating process. Additionally, future studies could explore evaluators' familiarity with AI-based tools like ChatGPT, as understanding their prior exposure to such technologies could refine study design and improve the interpretation of rating discrepancies.

Beyond methodological considerations, the study highlights the potential of AI tools like ChatGPT in SLA research and language education. Future research could explore practical applications of AI in classroom settings, curriculum design, and teacher training programs. For instance, AI-powered chatbots could provide real-time feedback to students on writing tasks, while adaptive learning models could tailor language instruction to individual learners' proficiency levels. Additionally, integrating AI-assisted assessment tools into language programs could help educators streamline feedback processes and improve learning outcomes. These avenues for future research will not only expand the current understanding of AI's role in language learning but also provide practical insights into its pedagogical integration.

**Competing interests:** The authors have no conflict of interest to declare.

**Authors Contributions:** Albatool Alhazmi was responsible for the Introduction, Literature Review, Resources, and Writing—Reviewing and Editing. Muneera Muftah contributed to Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, and Writing—Original Draft Preparation and Editing.

**Ethics declarations: Ethics statement:** The authors obtained ethical approval (Approval No. 202402-076-018543-041930) from Najran University's Research Ethics Committee, ensuring adherence to ethical standards, including informed consent, maintaining confidentiality, and complying with regulations.

**Ethics approval and consent to participate:** Confidentiality and the participants' identities were protected by collecting data anonymously. All participants were assured that their responses would only be used for scientific research purposes. All the participants have voluntarily taken part in the study.

**AI statement:** This research did not utilize any AI-generated content in the writing, analysis, or interpretation of the findings. The study was conducted using original human-authored research methods, including manual data collection, statistical analysis, and literature synthesis. While ChatGPT was the subject of investigation in this study, it was not employed in drafting, editing, or formulating the manuscript. All conclusions and discussions presented in this paper are based on empirical findings and human interpretation

## References

- Al-Khresheh, M., 2024. Bridging technology and pedagogy from a global lens: Teachers' perspectives on integrating ChatGPT in English language teaching. *Computers and Education: Artificial Intelligence*, 6(1), pp.1-12. <https://doi.org/10.1016/j.caeai.2024.100218>
- Barrot, J.S., 2023. Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57(1), pp.1-6. <https://doi.org/10.1016/j.asw.2023.100745>
- Baskara, R. and Mukarto, M., 2023. Exploring the implications of ChatGPT for language learning in higher education. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 7(2), pp.343-358. <http://dx.doi.org/10.21093/ijeltal.v7i2.1387>
- Branum, C. and Schiavenato, M., 2023. Can ChatGPT accurately answer a PICOT question? Assessing AI response to a clinical question. *Nurse Educator*, 48(5), pp.231-233. <https://doi.org/10.1097/NNE.0000000000001436>
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. and Amodei, D., 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, pp.1877-1901.
- Chaka, C., 2023. Generative AI chatbots-ChatGPT versus YouChat versus Chatsonic: Use cases of selected areas of applied English language studies. *International Journal of Learning, Teaching and Educational Research*, 22(6), pp.1-19. <https://doi.org/10.26803/ijlter.22.6.1>
- Corder, G.W. and Foreman, D.I., 2011. *Nonparametric statistics for non-statisticians*. Hoboken: John Wiley & Sons, pp.99-105. <http://dx.doi.org/10.1002/9781118165881>
- Ellis, R., 2015. *Understanding second language acquisition* (2nd ed.). UK: Oxford University Press.
- Firat, M., 2023. How ChatGPT can transform autodidactic experiences and open education?. [online] Available at: <https://doi.org/10.31219/osf.io/9ge8m>

- Fleiss, J.L., 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), pp.378-382. <https://doi.org/10.1037/h0031619>
- Huallpa, J., Arocutipa, J., Panduro, W., Huete, L., Limo, F., Herrera, E., Callacna, R., Flores, V., Romero, M., Quispe, I. and Hernández, F., 2023. Exploring the ethical considerations of using ChatGPT in university education. *Periodicals of Engineering and Natural Sciences*, 11(4), pp.105-115. <http://dx.doi.org/10.21533/pen.v11i4.3770>
- Huang, W., Hew, K. and Fryer, L., 2022. Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38(1), pp.237-257. <https://doi.org/10.1590/2177-6709.28.5.e2323183.oar>
- Jung, S.K. and Kim, T.W., 2016. New approach for the diagnosis of extractions with neural network machine learning. *American Journal of Orthodontics and Dentofacial Orthopedics*, 149(1), pp.127-133. <https://doi.org/10.1016/j.ajodo.2015.07.030>
- Kalla, D., Smith, N. and Carolina, N., 2023. Study and analysis of ChatGPT and its impact on different fields of study. *International Journal of Innovative Science and Research Technology*, 8(3), pp.827-833.
- Karataş, F., Abedi, F., Ozek Gunyel, F., Karadeniz, D. and Kuzgun, Y., 2024. Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, pp.1-24. <https://doi.org/10.1007/s10639-024-12574-6>
- Kruskal, W.H. and Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), pp.583-621. <https://doi.org/10.1080/01621459.1952.10483441>
- Kung, T., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G. and Maningo, J., 2023. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS Digital Health*, 2(2), pp.1-12. <https://doi.org/10.1371/journal.pdig.0000198>
- Landis, J.R. and Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1), pp.159-174. <https://doi.org/10.2307/2529310>
- Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Lie, X., Qianga, N., Shenf, D., Liu, T. and Ge, B., 2023. Summary of ChatGPT-related research and perspective towards the future of large language models. *Meta-Radiology*, 1(2), pp.1-14. <https://doi.org/10.1016/j.metrad.2023.100017>
- Lund, B. and Wang, T., 2023. Chatting about ChatGPT: How may AI and GPT impact academia and libraries?. *Library Hi Tech News*, 40(3), pp.26-29. <https://doi.org/10.1108/LHTN-01-2023-0009>
- Makaremi, M., Lacaule, C. & Mohammad-Djafari, A., 2019. Deep learning and artificial intelligence for the determination of the cervical vertebra maturation degree from lateral radiography. *Entropy*, 21(12), p.1222. Available at: <https://doi.org/10.3390/e21121222>.
- Muftah, M., 2023a. Error analysis in second language acquisition (SLA): Types and frequencies of grammatical errors of simple present and past tense in the elicited written production task of Arab EFL undergraduate learners. *Colombian Applied Linguistics Journal*, 25(1), pp.42-56. Available at: <https://doi.org/10.14483/22487085.19202>.
- Muftah, M., 2023b. Data-driven learning (DDL) activities: Do they truly promote EFL students' writing skills development? *Education and Information Technologies*, 28(10), pp.13179-13205. Available at: <https://doi.org/10.1007/s10639-023-11620-z>.
- Muftah, M., 2024. Impact of social media on learning English language during the COVID-19 pandemic. *PSU Research Review*, 8(1), pp.211-226. Available at: <https://doi.org/10.1108/PRR-10-2021-0060>.
- OpenAI, 2024. GPT-4. Available at: <https://openai.com/research/gpt-4>.
- Ray, 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, pp.121-154. Available at: <https://doi.org/10.1016/j.iotcps.2023.04.003>.
- Sallam, M., 2023. The utility of ChatGPT as an example of large language models in healthcare education, research and practice: Systematic review on the future perspectives and potential limitations. *MedRxiv*, pp.1-34. Available at: <https://doi.org/10.1101/2023.02.19.23286155>.
- Sarraj, A., Bruemmer, D., Van Iterson, E., Cho, L., Rodriguez, F. & Laffin, L., 2023. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. *Jama*, 329(10), pp.842-844. Available at: <https://doi.org/10.1001/jama.2023.1044>.
- Sharma, S. & Yadav, R., 2023. ChatGPT – A technological remedy or challenge for education system. *Global Journal of Enterprise Information System*, 14(4), pp.46-51. Available at: <https://www.gjeis.com/index.php/GJEIS/article/view/698>.
- Srinivasa, K., Kurni, M. & Saritha, K., 2022. Harnessing the power of AI to education. In: *Learning, Teaching, and Assessment Methods for Contemporary Learners*, pp.311-342. Singapore: Springer. Available at: [https://doi.org/10.1007/978-981-19-6734-4\\_13](https://doi.org/10.1007/978-981-19-6734-4_13).
- Sun, C., 2023. Curiosity driven learning in artificial intelligence and its applications. Doctoral thesis, Nanyang Technological University, Singapore. Available at: <https://hdl.handle.net/10356/172831>.
- Tanaka, O., Gasparello, G., Hartmann, G., Casagrande, F. & Pithon, M., 2023. Assessing the reliability of ChatGPT: A content analysis of self-generated and self-answered questions on clear aligners, TADs and digital imaging. *Dental Press Journal of Orthodontics*, 28, pp.1-22. Available at: <https://doi.org/10.1590/2177-6709.28.5.e2323183.oar>.
- VanPatten, B., Smith, M. and Benati, A.G., 2020. *Key questions in second language acquisition: An introduction*. UK: Cambridge University Press.

# Blended Learning and Math Achievement: A Meta-Analytic Review Highlighting the Effectiveness and Heterogeneity

Amelia Defrianti Putri, Dadang Juandi, Turmudi and Suparman

Department of Mathematics Education, Universitas Pendidikan Indonesia, Indonesia

[ameliadefrianti@upi.edu](mailto:ameliadefrianti@upi.edu)

[dadang.juandi@upi.edu](mailto:dadang.juandi@upi.edu) (corresponding author)

[turmudi@upi.edu](mailto:turmudi@upi.edu)

[arman95@upi.edu](mailto:arman95@upi.edu)

<https://doi.org/10.34190/ejel.23.1.3781>

An open access article under [CC Attribution 4.0](#)

**Abstract:** The role of technology in mathematics education is growing more significant as contemporary learning evolves, particularly with the advent of blended learning approaches that merge traditional in-person instruction with digital and online tools. A significant amount of research has examined the connection between blended learning and mathematics education. Nevertheless, the effect of blended learning on math achievement has shown inconsistent results, indicating a need for a more comprehensive analysis. This research seeks to address this gap by analyzing the varying effects of blended learning on math achievement across different educational systems and learning environments. To achieve this objective, a meta-analytic review was conducted, synthesizing 34 empirical studies published between 2014 and 2023. These studies involved a total of 2,996 students and provided 34 units of effect size for analysis. Various statistical techniques were employed, including sensitivity analysis, publication bias assessment, Z-test, and Cochrane's Q test, all conducted using Comprehensive Meta-Analysis software. The findings from this meta-analysis reveal that the implementation of blended learning in mathematics has a significantly positive and robust effect on students' math achievement ( $g = 1.090$ ;  $p < 0.05$ ). This suggests that blended learning, when implemented effectively, can enhance students' understanding and performance in mathematics. Furthermore, the current meta-analytic review has examined that two moderating factors, such as educational level and digital platform, significantly differentiated students' math achievement in the environment of blended learning. Meanwhile, there has not been adequate evidence to conclude that class capacity and geographical location significantly differentiated students' math achievement in the blended learning environment. These findings highlight that while some contextual factors, such as educational level and the choice of digital platforms, play a crucial role in influencing math achievement, others, like class size and geographical location, may not be as impactful in blended learning contexts. This study offers an in-depth insight into the ways blended learning can effectively optimize math achievement across different educational settings. The positive outcomes associated with blended learning suggest that educators and policymakers should consider incorporating digital tools and resources into the mathematics curriculum more strategically.

**Keywords:** Blended learning, Math achievement, Meta-Analysis, Effect size, Systematic review

## 1. Introduction

Math achievement plays a critical role in students' academic success, serving as a key indicator of their ability to develop logical reasoning, problem-solving skills, and critical thinking (Nida, Usodo and Saputro, 2020). It encompasses a range of competencies, including mastery of mathematical concepts, problem-solving, mathematical representation, critical thinking, mathematical communication, and mathematical literacy (Ayob et al., 2023; Kilpatrick, 2001; Lin, Tseng and Chiang, 2016). These skills are essential for students to navigate various real-world contexts, apply mathematical reasoning to complex situations, and make informed decisions (Azid et al., 2022). Although these competencies are crucial, international assessments like the Programme for International Student Assessment (PISA) continue to reveal significant disparities in students' math achievement (Putri et al., 2024). The 2022 PISA findings, for example, showed a 15-point decline in average mathematics performance among OECD countries compared to 2018, raising concerns about the current state of math education worldwide (OECD, 2023). Various studies suggest that this decline has been attributed to multiple factors, including the disruptions caused by the coronavirus disease 2019 (COVID-19), insufficient teacher training in modern pedagogical methods, and unequal access to learning resources, particularly in underprivileged areas (İDİL, GÜLEN and DÖNMEZ, 2024; Nguyen et al., 2021). These underlying issues underscore the need for targeted interventions to address these challenges and improve students' mathematical abilities (Tong, Uyen and Ngan, 2022).

Traditional instructional methods in mathematics often fall short of fostering the full range of skills required for comprehensive math achievement (Alsalhi et al., 2021). These methods, while effective in developing procedural

knowledge, are often limited in promoting deeper conceptual understanding, critical thinking, and active engagement with mathematical concepts. As a result, students may become disengaged, leading to suboptimal learning outcomes and limited preparedness for solving complex, real-world problems (Aldalalah, Shatat and Ababneh, 2019). These limitations emphasize the need for innovative approaches that not only build foundational skills but also encourage higher-order thinking and personalized learning experiences. Blended learning has emerged as a promising alternative to overcome these challenges (Hrastinski, 2019; Lin, Tseng and Chiang, 2016). Blended learning is an educational approach that merges traditional classroom instruction with online learning components, aiming to create a more adaptable, interactive, and customized learning experience (Almasi and Zhu, 2018; Bonk and Graham, 2012; Graham, 2006; Machumu, Zhu and Sesabo, 2016). By combining in-person teaching with online resources, interactive activities, and personalized learning pathways, blended learning caters to individual student needs and learning styles (Garrison and Kanuka, 2004). This pedagogical approach allows students to not only access diverse resources but also engage in collaborative problem-solving and receive real-time feedback, enabling them to build both procedural and conceptual mathematical competencies effectively (Fardian, Suryadi and Prabawanto, 2025).

In mathematics education, this flexible instructional model provides opportunities for differentiated instruction, real-time feedback, and collaborative learning, thereby fostering deeper conceptual understanding and problem-solving skills among students (Cao, 2023). The rationale for investigating the effect of blended learning on math achievement lies in its potential to address longstanding challenges and limitations associated with traditional instructional methods (Pokorny, 2019). Research indicates that conventional approaches to teaching mathematics often fail to effectively engage students, resulting in disinterest, frustration, and poor academic performance (Aldalalah, Shatat and Ababneh, 2019). Blended learning, emphasizing active learning, student-centered pedagogy, and multimedia integration, offers a promising alternative to reinvigorate the learning process and improve outcomes in mathematics education (Bernard et al., 2014; Garrison and Kanuka, 2004). Despite its growing popularity, significant debate persists about the effectiveness of blended learning in improving students' math achievement (Indrapangastuti, Surjono and Yanto, 2021). While numerous studies have explored its impact, the results have often been mixed, with some research highlighting significant benefits while others report minimal or no effect (Kiviniemi, 2014). These inconsistent findings suggest the presence of heterogeneity in the effectiveness of blended learning, which may be influenced by various factors such as instructional design, implementation strategies, student characteristics, and contextual variables (Cao, 2023; Kintu, Zhu and Kagambe, 2017).

Several prior studies employing systematic reviews and meta-analyses have examined blended learning's impact on mathematics skills. For example, Mawardi, Budiningsih and Sugiman (2023) reviewed 26 primary studies and synthesized 37 effect sizes, finding an effect size of 1.01, which is considered large. This suggests that blended learning significantly enhances students' mathematical abilities compared to traditional methods. In contrast, Setiawan, Muhtadi and Hukom (2022) analyzed 36 effect sizes, with a random-effects model revealing an effect size of 1.269, categorized as very high. This indicates that, on average, blended learning effectively improves mathematical skills among Indonesian students. Similarly, Vo, Zhu and Diep (2017) conducted a meta-analysis on blended learning at the course level in higher education, reporting a significant positive effect ( $g = 0.385$ ,  $p < .001$ ). Although smaller than the findings of Mawardi and Setiawan, this result further confirms blended learning's potential in improving student performance across diverse disciplines. However, these studies have limitations, as they primarily focus on specific regions and datasets that may not incorporate the latest developments in educational technology. Additionally, there is often a lack of comprehensive analysis of moderator variables that could affect blended learning's effectiveness across different educational settings.

This study offers several significant advantages over previous meta-analyses on blended learning by addressing key limitations and broadening the scope of investigation. One of its main strengths lies in utilizing primary data from the most recent studies up to 2023, ensuring a more up-to-date understanding of blended learning's impact on students' mathematical achievement. For instance, earlier meta-analyses, such as Setiawan, Muhtadi and Hukom (2022), only included data up to 2021, which overlooked recent advancements in educational technology. For example, the integration of advanced platforms such as GeoGebra for interactive mathematical modeling, Google Classroom for streamlined communication and resource sharing, and gamified tools like Kahoot! and Quizizz for improving student engagement are now widely adopted. These tools have transformed how blended learning is implemented, making it more personalized and interactive.

Similarly, Mawardi, Budiningsih and Sugiman (2023) synthesized findings from studies focusing predominantly on Indonesian students and platforms like LMS and social media, limiting the generalizability of their conclusions to global contexts. Their work primarily centered on a narrow range of tools, leaving gaps in understanding the

broader technological innovations in blended learning. In contrast, this study bridges these gaps by incorporating a wider range of studies and platforms, offering a more global and comprehensive framework for understanding the effectiveness of blended learning in mathematics education. Unlike Setiawan, Muhtadi and Hukom (2022), which concentrated on a single geographical region, this study examines geographical location as a moderator, recognizing how blended learning's impact varies across cultural and contextual settings. This aligns with findings by Cao (2023), which highlights significant variations in blended learning outcomes across countries. In addition, this study also investigates critical moderators, such as class size, educational levels, and technological tools, to provide a nuanced perspective compared to previous research.

This research further extends prior analyses by exploring a diverse array of technological platforms. While Samritin et al. (2023), focused on a limited set of tools, this study incorporates a broader spectrum, including GeoGebra, Edmodo, Google Classroom, LMS, Microsoft Mathematics, Moodle, PowerPoint, Schoology, Video, and WhatsApp. By examining these platforms, this research provides detailed insights into how specific tools influence the effectiveness of blended learning in varied educational contexts, addressing the gaps left by Mawardi, Budiningsih and Sugiman (2023) and others. Additionally, this study builds on prior meta-analyses, such as Vo, Zhu and Diep (2017), which demonstrated the variability of blended learning's impact across disciplines and educational levels. However, unlike Vo, Zhu and Diep (2017), this research specifically addresses mathematics education, integrating findings from studies like Cao (2023) and Samritin et al. (2023) to validate and expand its conclusions. By synthesizing these perspectives, this study offers a holistic understanding of blended learning's effectiveness while addressing gaps in previous research.

The main goal of this meta-analysis is to provide a comprehensive evaluation of the impact of blended learning on students' mathematics performance, emphasizing various aspects such as comprehension of mathematical concepts, problem-solving abilities, representation, critical thinking, communication, and mathematical literacy. By integrating a wide array of empirical studies, this research seeks to uncover both the potential advantages and challenges associated with the adoption of blended learning in mathematics education. Through a detailed examination of moderator variables and the inclusion of the latest data, this meta-analysis aspires to provide meaningful insights for educators and policymakers on optimizing blended learning strategies to improve mathematics achievement in diverse educational settings. This study specifically aims to explore these key research inquiries:

*RQ1: What is the general effect of blended learning interventions on students' math achievement, and how significantly does the integration of blended learning enhance students' math achievement?*

*RQ2: How do blended learning interventions impact students' math achievement when considering factors like class size, grade level, digital platform, and geographical location?*

## 2. Literature Review

### 2.1 Math Achievement

Mathematics achievement refers to the extent to which students attain proficiency in mathematical skills and knowledge, as evidenced by their performance on assessments and their ability to apply mathematical concepts in various contexts (Azid et al., 2022; Lin, Tseng and Chiang, 2016). According to Kilpatrick (2001) mathematics achievement encompasses not only the ability to solve mathematical problems but also the understanding of fundamental concepts and the ability to communicate and represent mathematical ideas effectively. This perspective is endorsed by the National Council of Teachers of Mathematics (NCTM), which defines mathematics achievement as a complex construct encompassing students' conceptual understanding, procedural fluency, and the ability to apply mathematics in real-world contexts (Leinwand, 2014).

In the context of this study, mathematics achievement is understood through several key components that collectively define students' proficiency in mathematical skills. Student achievement is typically measured through performance on mathematical tasks, including standardized assessments, classroom tests, and problem-solving exercises, which reflect both procedural fluency and conceptual understanding (Fazal and Bryant, 2019). Understanding of mathematical concepts involves the ability to grasp fundamental principles, recognize patterns, and establish connections between different mathematical ideas, enabling students to apply their knowledge flexibly in various contexts (Yaghmour, 2016). Mathematical problem-solving refers to the capacity to analyze, strategize, and systematically resolve complex mathematical situations, which is crucial for higher-order thinking and real-world application (Pertiwi et al., 2019). Mathematical representation plays a vital role in expressing mathematical ideas through symbols, graphs, tables, and models, facilitating a deeper comprehension of abstract concepts and their applications (Khairiyah, Mulyono and Fauzi, 2021). Critical

thinking and mathematical communication are also essential, as they allow students to articulate their reasoning, justify solutions, and engage in discussions that enhance their understanding of mathematical concepts (Nida, Usodo and Saputro, 2020; Setiyani, 2019). Mathematical literacy reflects the ability to apply mathematical knowledge in everyday situations, such as financial decision-making, data interpretation, and problem-solving in professional settings, highlighting the importance of mathematics beyond academic contexts (Kilpatrick, 2001).

Furthermore, theoretical frameworks provide a deeper understanding of how mathematical achievement develops. Vygotsky's social constructivism highlights the role of collaboration and guided learning, where social interactions play a crucial role in developing problem-solving skills and conceptual understanding (Gredler, 2011). Cognitive load theory explains the importance of reducing extraneous cognitive load during the learning process, allowing students to focus on intrinsic mathematical tasks (Paas, Van Gog and Sweller, 2010). These theories offer insights into the mechanisms that underpin mathematical achievement across various educational settings.

## **2.2 Blended Learning**

Blended learning is an educational approach that merges traditional classroom instruction with online learning components, aiming to create a more adaptable, interactive, and customized learning experience (Alsalhi et al., 2021; Graham, 2006; Setiawan, Muhtadi and Hukom, 2022). This study defines blended learning as any instructional model that combines face-to-face and online components, particularly in mathematics education. Staker and Horn (2012) categorize blended learning into four models: Rotation Model, Flex Model, Self-Blend Model, and Enriched Virtual Model. The Rotation Model involves structured shifts between online and face-to-face learning and consists of four subtypes: Station Rotation, where students move between different learning stations, including an online component; Lab Rotation, where students alternate between a computer lab and classroom instruction; Flipped Classroom, where students study online at home before engaging in problem-solving activities in class; and Individual Rotation, which personalizes schedules based on student needs. The Flex Model relies primarily on online learning with optional face-to-face support, while the Self-Blend Model allows students to supplement their coursework with online resources. The Enriched Virtual Model prioritizes online instruction with occasional in-person sessions.

Each model varies in effectiveness based on the educational context. Flipped classrooms enhance self-regulated learning and critical thinking, making them ideal for secondary and higher education (Means et al., 2013). In contrast, Rotation Models provide structured guidance, making them more suitable for primary education. The Flex Model supports self-paced learners, while the Self-Blend and Enriched Virtual Models cater to independent students in remote or hybrid settings. By aligning blended learning approaches with specific learner needs, this review highlights their impact on mathematics education. The theoretical underpinnings of blended learning include multimedia learning theory and social constructivism. Multimedia learning theory explains how visual and auditory integration enhances comprehension, particularly in online components of blended learning (Mayer, 2009). In addition, social constructivism emphasizes collaboration and peer interaction, which are fostered through blended learning environments that combine online and in-person activities (Gredler, 2011).

## **2.3 Moderating Factors**

The disparity in students' math achievement within technology-assisted learning environments suggests the presence of several moderating factors. These factors indirectly contribute to the variations in students' math achievement, resulting in different levels of achievement. While some students achieve high levels of math achievement, others perform at lower levels, with many falling in the middle. Thus, it is essential to explore and evaluate the impact of these factors on students' math achievement. According to various studies, such as those by Helsa et al. (2023) and Tawaldi et al. (2023), there are generally two types of moderating factors: substantial and extrinsic. Helsa et al. (2023) highlight that substantial factors are those that have a direct connection with either the independent or dependent variables, such as class size, level of education, digital tools, and geographical region. Conversely, extrinsic factors pertain to aspects that do not have a direct link to the independent or dependent variables, including the year of publication, type of document, source, and database. Figure 1 illustrates the categorization of these moderating factors into substantial and extrinsic, along with examples for each category. This study concentrates specifically on substantial factors such as class size, educational level, digital platforms, and geographical region because these elements significantly influence variations in students' math achievement.

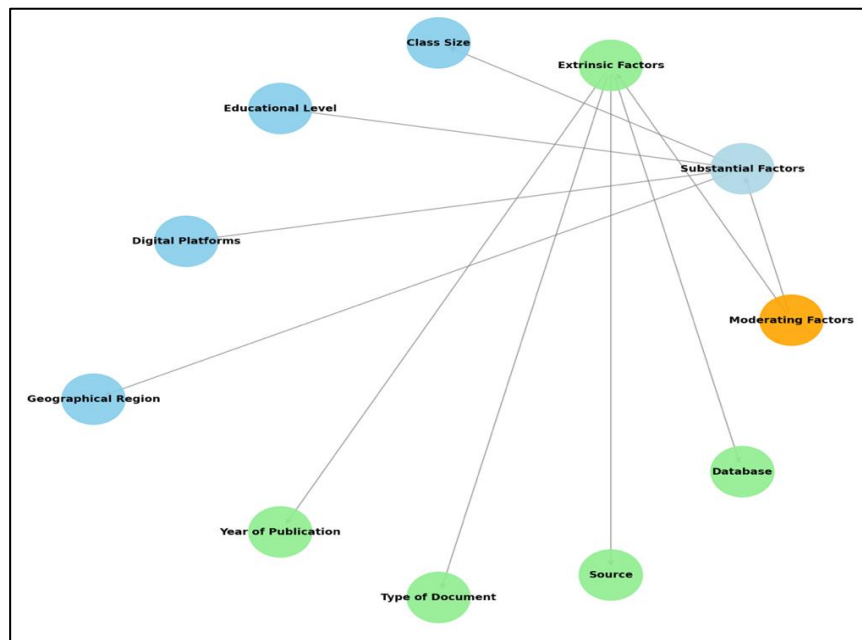


Figure 1: Circular network analysis: moderating factors

### 3. Method

#### 3.1 Research Design

This study employed a meta-analysis approach, specifically using a random effects model to account for various factors, such as class size, educational level, platform, and geographic region (Borenstein et al., 2021). Numerous academic sources have outlined seven distinct stages for conducting a meta-analysis (Putri, Juandi and Turmudi, 2024; Suparman and Juandi, 2022). Figure 2 presents these methods.

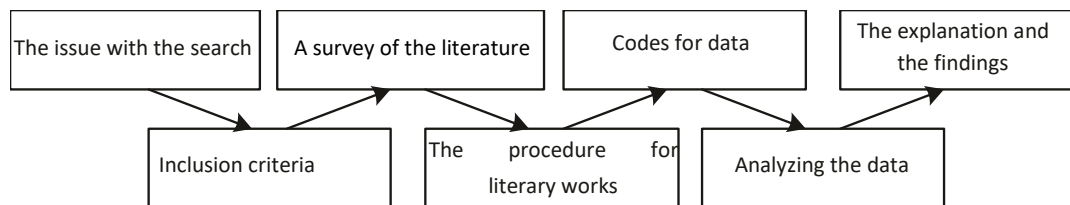


Figure 2: Diagram of the phases in the meta-analysis process

#### 3.2 Inclusion Criteria

To ensure the rigor and relevance of this meta-analytic review, specific inclusion criteria were established to define the scope of the investigative study. These criteria were designed to ensure that the selected studies align with the research objectives and provide sufficient data for analysis. The inclusion criteria are summarized in Table 1 below:

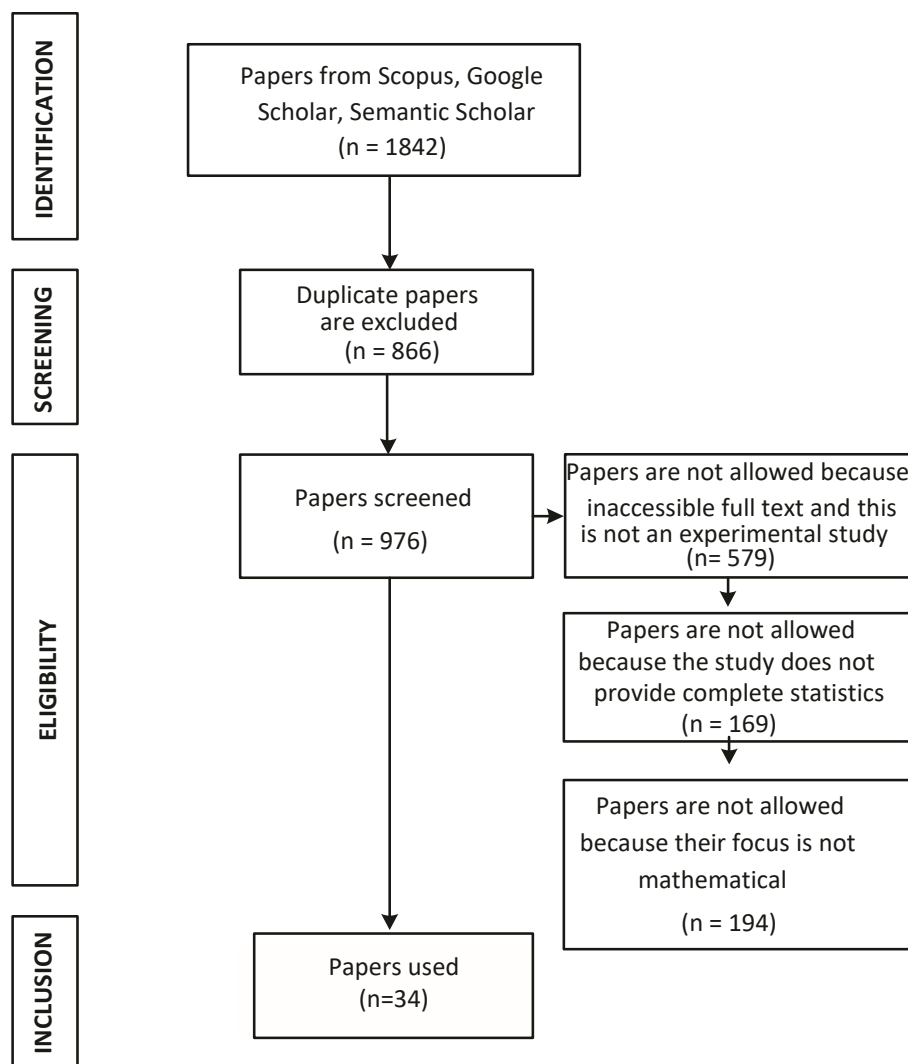
Table 1: Inclusion criteria for the investigative study

| No | Criteria                      | Inclusion                                                               |
|----|-------------------------------|-------------------------------------------------------------------------|
| 1  | Population                    | Global student population                                               |
| 2  | Intervention                  | Implementation of blended learning as the primary intervention strategy |
| 3  | Comparison                    | Traditional learning as the comparative baseline                        |
| 4  | Outcome                       | Mathematics achievement as the measured outcome                         |
| 5  | Study Design                  | Experimental design with a control group                                |
| 6  | Statistical Data Availability | Statistical data available for both experimental and control groups     |
| 7  | Peer-Reviewed Sources         | Scholarly journal articles or peer-reviewed conference publications     |
| 8  | Publication Year Range        | Published between 2014–2023                                             |

| No | Criteria                | Inclusion                            |
|----|-------------------------|--------------------------------------|
| 9  | Full-Text Accessibility | Full-text articles accessible online |

### 3.3 Literature Search and Selection

A literature search was carried out using several databases, including Scopus, Semantic Scholar, and Google Scholar, with keyword combinations such as "blended learning" and "math achievement," "blended learning" and "mathematical abilities," or "blended learning" and "math skills." This search retrieved 75 documents from Scopus, 247 from Semantic Scholar, and 1,520 from Google Scholar, all related to blended learning and math achievement. The document selection process followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Moher et al., 2009). Figure 3 outlines the detailed steps of the selection process.



**Figure 3: PRISMA diagram outlining the document selection process.**

Figure 3 illustrates the flow diagram of the literature search conducted following the PRISMA protocol. From 2014 to 2023, a total of 1,842 articles were initially retrieved through various meta-search engines. A data cleanup tool was then used to remove duplicates, resulting in the elimination of 790 duplicate articles. The tool failed to detect certain duplicates because of discrepancies in formatting, including variations in wording, number formatting, and the presentation of information. As a result, manual review and removal became necessary to ensure accuracy, leading to the identification and elimination of 76 duplicates. Afterward, the remaining articles were carefully reviewed, and those that fulfilled the predetermined inclusion and exclusion criteria were assessed.

### 3.4 Data Extraction

This meta-analysis used an extensive coding sheet to document essential parameters and variables crucial to the research, including the researcher's identity, publication year, class size, educational level, platform used, and geographic location. These data points were meticulously recorded in Google Sheets. Additionally, the coding form incorporated essential metrics such as sample size for both groups, as well as mean and standard deviation values. This thorough coding methodology was implemented to enhance the reliability and robustness of the research findings. To ensure accuracy, two experts in meta-analytic reviews were consulted to verify and validate the data. After recoding and reviewing the data, no discrepancies were found between the experts' codings and those of the researchers, thus confirming the integrity and precision of the data for this meta-analytic study.

### 3.5 Data Analysis

This meta-analysis utilized Hedge's  $g$  to compute effect size (Borenstein et al., 2021), given the limited sample sizes in the blended learning classes (Harwell, 2020). Effect sizes were categorized according to the guidelines of Cohen (Putri, Juandi and Turmudi, 2024):  $g=0.00-0.20$  indicating a weak effect,  $g=0.21-0.50$  denoting a modest effect,  $g=0.51-1.00$  representing a moderate effect, and  $g > 1.00$  indicating a robust effect. Additionally, the impact of blended learning on math achievement was analyzed using the Z test (Borenstein et al., 2021). The study also employed the Q Cochran test to examine the effects of class size, grade level, platform, and geographical location. The formulation of Hedge's  $g$  is detailed bellow (Borenstein et al., 2021):

$$g = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} \times \left(1 - \frac{3}{4df - 1}\right)$$

Heterogeneity among the included studies was evaluated using Cochran's Q test and the  $I^2$  statistic. The Q-test yielded a value of 257.379 with  $df = 33$  ( $p < 0.001$ ), indicating significant heterogeneity among the studies. Furthermore, the  $I^2$  statistic was calculated at 87.18%, suggesting that 87.18% of the observed variability in effect sizes was attributable to true differences across studies rather than random sampling error. This high degree of heterogeneity underscores the importance of examining moderator variables to account for the variability. Moderator analysis was conducted to explore factors such as class size, educational level, digital platform, and geographical region, providing insights into the conditions under which blended learning interventions are most effective.

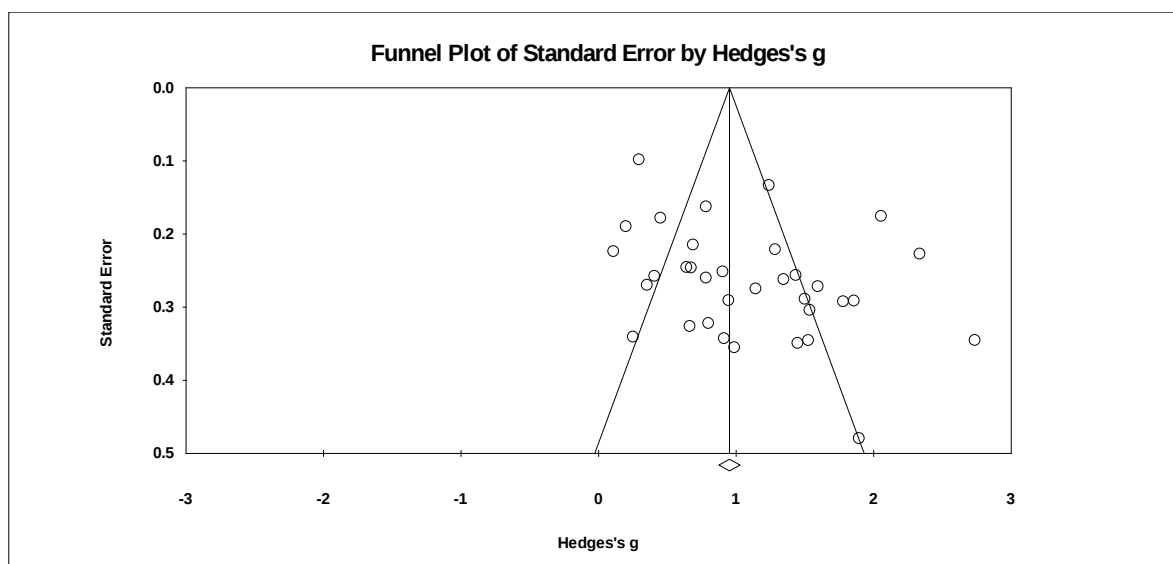
Evaluating publication bias and sensitivity is vital to guarantee the accuracy and stability of statistical data in major studies, as no research outcomes are completely unaffected by publication bias (Bernard et al., 2014). Funnel plots and the fill and trim method were employed in the analysis to evaluate publication bias (Harwell, 2020). Sensitivity analysis was performed using the "One study deleted" function within the Comprehensive meta-analysis (CMA) program.

## 4. Results

This meta-analytic review aims to provide a comprehensive examination of the impact of blended learning on mathematics achievement. It explores key aspects, including sensitivity analysis, publication bias, estimated effect size, and subgroup analysis. These components are systematically discussed in the following subsections to provide an in-depth understanding of the study's findings.

### 4.1 Sensitivity Analysis and Publication Bias

To verify phenomena of publication bias indication, it can be carried out by observing the dispersion of effect size data in the funnel plot (refer to Figure 4).



**Figure 4: The spread of effect size data in the funnel plot**

As shown in Figure 4, the effect size data in the funnel plot exhibited a symmetrical distribution. To justify the symmetry of the distribution, the test of trim and fill was conducted (refer to Table 2).

**Table 2: The outcomes of the fill and trim analysis**

|                        | Studies Trimmed | Effect Size in g | Lower Limit | Upper Limit | Q-value |
|------------------------|-----------------|------------------|-------------|-------------|---------|
| <b>Observed Values</b> |                 | 1.090            | 0.867       | 1.313       | 257.379 |
| <b>Adjusted Values</b> | 0               | 1.090            | 0.867       | 1.313       | 257.379 |

As shown in Table 2, There was no need to exclude any effect size data from the distribution, neither from the right side nor the left side. This justifies that there was really symmetrical distribution of effect size data in the funnel plot. Consequently, this interprets that there is no phenomenon of publication bias of the collection of effect size data (Tawaldi et al., 2023).

Researchers conducted a sensitivity analysis to examine the effect size data by identifying outliers within the range of the highest and lowest effect sizes. The findings revealed that the smallest effect size in g was 0.867, while the largest was 1.313. On the other hand, the estimated point in g of 34 effect size data was 1.090. As a consequence of these findings, the estimated point was located in the interval between 0.867 and 1.313, and moreover, there was no data that could be outlier. This interprets that there no phenomena which indicated sensitive data when there was a variation in the amount of data within the effect size collection (Putri, Juandi and turmudi, 2024).

#### 4.2 Estimated Effect Size

34 eligible documents included in this meta-analytic review generated 34 units of effect size in g and involved 2,996 students. The unit of effect size was heterogeneous in the perspective of direction, significance, and strength (refer to Table 3).

**Table 3: The results of calculations of effect size**

| Document                     | Effect Size in g Unit | P-value |
|------------------------------|-----------------------|---------|
| Lin, Tseng and Chiang (2016) | 0.356 [-0.174; 0.886] | 0.188   |
| Fazal and Bryant (2019)      | 0.298 [0.104; 0.491]  | 0.003   |
| Setyaningrum (2018)          | 0.454 [0.104; 0.805]  | 0.011   |
| Alsahhi et al. (2021)        | 2.058 [1.713; 2.404]  | 0.000   |
| Tong, Uyen and Ngan (2022)   | 0.691 [0.269; 1.113]  | 0.001   |
| Yaghmour (2016)              | 1.288 [0.853; 1.723]  | 0.000   |

| Document                                  | Effect Size in g Unit | P-value |
|-------------------------------------------|-----------------------|---------|
| Pertiwi et al. (2019)                     | 0.906 [0.412; 1.400]  | 0.000   |
| Noviyanti, Sugiharta and Farida (2019)    | 1.146 [0.606; 1.686]  | 0.000   |
| Nugraha, Astawa and Ardana (2019)         | 0.676 [0.193; 1.160]  | 0.006   |
| Mutaqin, Marethi and Syamsuri (2016)      | 1.539 [0.942; 2.136]  | 0.000   |
| Septiyan, Anriani and Hendrayana (2019)   | 0.410 [-0.096; 0.916] | 0.112   |
| Trisnayanti, Sariyasa and Suweken (2020)  | 1.861 [1.289; 2.433]  | 0.000   |
| Apsari (2020)                             | 0.112 [-0.328; 0.552] | 0.618   |
| Jayanti and Rahmawati (2017)              | 1.349 [0.835; 1.864]  | 0.000   |
| Khairiyah, Mulyono and Fauzi (2021)       | 0.255 [-0.414; 0.924] | 0.456   |
| Sudiarta and Sadra (2016)                 | 1.438 [0.934; 1.942]  | 0.000   |
| Nugraha, Astawa and Ardana (2019)         | 1.898 [0.957; 2.838]  | 0.000   |
| Zein et al. (2019)                        | 0.785 [0.274; 1.295]  | 0.003   |
| Satriani, Wangid and PA (2020)            | 1.598 [1.064; 2.132]  | 0.000   |
| Supriadi et al. (2014)                    | 2.740 [2.062; 3.417]  | 0.000   |
| Nida, Usodo and Saputro (2020)            | 0.785 [0.465; 1.104]  | 0.000   |
| Mashuri and Nasrum (2020)                 | 1.529 [0.851; 2.207]  | 0.000   |
| Ektafia, Fitri and Najibufahmi (2021)     | 1.783 [1.208; 2.357]  | 0.000   |
| Anwar and Setyaningrum (2021)             | 0.643 [0.161; 1.125]  | 0.009   |
| Ayuningtyas and Prastowo (2022)           | 1.451 [0.766; 2.137]  | 0.000   |
| Nasution, Sintia and Putri (2022)         | 0.991 [0.294; 1.688]  | 0.005   |
| Setiyani (2019)                           | 0.803 [0.170; 1.435]  | 0.013   |
| Darmono and Maryam (2019)                 | 0.949 [0.378; 1.520]  | 0.001   |
| Muncarno and Astuti (2021)                | 0.916 [0.243; 1.589]  | 0.008   |
| Pokorny (2019)                            | 1.242 [0.980; 1.505]  | 0.000   |
| Ayob et al. (2023)                        | 0.202 [-0.170; 0.575] | 0.287   |
| Indrapangastuti, Surjono and Yanto (2021) | 1.504 [0.936; 2.071]  | 0.000   |
| Angreanisita and Mastur (2021)            | 0.667 [0.027; 1.307]  | 0.041   |
| Seage and Türegün (2020)                  | 2.339 [1.892; 2.785]  | 0.000   |
| Estimated Effect Size                     | 1.090 [0.867; 1.313]  | 0.000   |

Table 3 shows that the estimated point for the 34 effect size data units in g was 1.090, indicating that the use of blended learning has a strong positive impact on students' math achievement. Additionally, the Z-test significance value was below 0.05, demonstrating that the implementation of blended learning had a significant effect on improving students' math performance. This suggests that incorporating blended learning into math instruction effectively enhances students' math achievement.

#### 4.3 Subgroup Analysis

The Q Cochrane test was applied to test some moderating factors (e.g., educational level, class capacity, digital platform, and geographical location) predicted in differentiating students' math achievement in the mathematics learning implementing blended learning (refer to Table 4).

**Table 4: The results of Q Cochrane test**

| Substantial Factor | Groups             | Effect Size in g Unit | P-value |
|--------------------|--------------------|-----------------------|---------|
| Educational Level  | Primary School     | 1.582                 | 0.004   |
|                    | Junior High School | 0.714                 |         |
|                    | Senior High School | 1.328                 |         |

| Substantial Factor    | Groups                | Effect Size in g Unit | P-value |
|-----------------------|-----------------------|-----------------------|---------|
|                       | College/University    | 1.313                 |         |
| Class Capacity        | Large Class           | 1.158                 | 0.291   |
|                       | Small Class           | 0.938                 |         |
| Digital Platform      | Edmodo                | 1.085                 | 0.000   |
|                       | GeoGebra              | 2.740                 |         |
|                       | Google Classroom      | 0.643                 |         |
|                       | LMS                   | 1.201                 |         |
|                       | Microsoft Mathematics | 0.949                 |         |
|                       | Moodle                | 0.955                 |         |
|                       | PowerPoint            | 1.288                 |         |
|                       | Schoology             | 1.378                 |         |
|                       | Video                 | 0.835                 |         |
|                       | WhatsApp              | 1.046                 |         |
| Geographical Location | Asia                  | 1.070                 | 0.605   |
|                       | America               | 1.308                 |         |
|                       | Europe                | 1.242                 |         |

Figure 5 illustrates the comparison of effect sizes across various moderating factors influencing students' mathematics achievement in blended learning environments. This visualization highlights the patterns and magnitudes of effect sizes for each factor, such as 'Educational Level,' 'Class Capacity,' 'Digital Platform,' and 'Geographical Location.' It complements the data presented in Table 4, providing a clearer and more intuitive understanding for readers.



**Figure 5: Effect sizes of moderating factors in blended learning**

From Table 4, it can be stated that two moderating factors (e.g., educational level and digital platform) significantly differentiated students' math achievement in the environment of blended learning. Educational level showed a strong impact, with the highest effect size observed at the primary school level ( $g = 1.582$ ,  $p = 0.004$ ), suggesting that blended learning may be particularly effective for younger students who benefit from foundational and interactive learning approaches. Digital platforms also played a crucial role, with GeoGebra demonstrating the largest effect size ( $g = 2.740$ ), highlighting the importance of dynamic, visually engaging tools in enhancing mathematical understanding. Meanwhile, there have not been adequate evidences to state that class capacity and geographical location significantly differentiated students' math achievement in the

environment of blended learning ( $p > 0.05$ ). This suggests that blended learning's effectiveness may be consistent across classroom sizes and regions, though further research is needed to explore these factors in detail, as discussed later.

## 5. Discussion

### 5.1 The Effectiveness of Blended Learning for Students' Math Achievement

This meta-analytic review demonstrates that the blended learning environment has a notably strong positive effect on enhancing students' math achievement. Furthermore, it indicates that implementing blended learning can significantly improve students' math achievement. Two other related meta-analytic studies similarly concluded that blended learning had a significant positive impact on students' success in math education (Mawardi, Budiningsih and Sugiman, 2023; Setiawan, Muhtadi and Hukom, 2022). Specifically, Mawardi, Budiningsih and Sugiman (2023) meta-analysis of 26 empirical studies found that blended learning had a significant positive effect ( $d = 1.01$ ;  $p = 0.00$ ) on students' math skills. Similarly, Setiawan, Muhtadi and Hukom (2022), in a meta-analysis of 36 empirical studies, showed that blended learning had a substantial positive effect ( $\Delta = 1.27$ ;  $p < 0.05$ ) on students' mathematical abilities. These findings reinforce the conclusion of this study, confirming that blended learning environments significantly enhance students' math achievement. This highlights the effectiveness of blended learning in enhancing students' math achievement.

Blended learning merges conventional in-class teaching with online educational activities. This method provides a flexible and interactive learning experience, enabling students to progress at their own speed and utilize a range of resources (Almasi and Zhu, 2018). Generally, Cronje and Van Zyl (2022) explained that blended learning consisted of face-to-face learning (teachers provide direct guidance, facilitate discussions, and address individual student needs) and online learning (students access interactive content, complete assignments, and receive personalized feedback). Blended learning offers several benefits for mathematics education. By combining face-to-face learning and online learning, it can cater to different learning styles, promote student engagement, and improve academic performance (Garrison and Kanuka, 2004). In detail, Hrastinski (2019) explained that blended learning has several advantages in mathematics, such as: (1) Improved student engagement: The use of technology and interactive activities fosters active participation, such as virtual simulations, gamified exercises, and collaborative tools like online discussion forums, which captivate students' attention and encourage deeper involvement in learning. (2) Increased student achievement: By incorporating multimedia resources and real-time feedback mechanisms, students gain a more comprehensive understanding of mathematical concepts, enabling them to achieve better outcomes in assessments. (3) Personalized instruction: Teachers can utilize data analytics from learning platforms to identify individual students' strengths and weaknesses, allowing for tailored lesson plans and adaptive learning paths that accommodate diverse learning styles and paces. (4) Flexibility and accessibility: Students can access recorded lectures, digital resources, and assignments anytime and anywhere, providing opportunities for review and reinforcement outside traditional classroom hours, particularly beneficial for students with varied schedules or geographical limitations. The advantages of blended learning enable mathematics teachers to optimize students' math achievement.

Within a blended learning environment, technology plays a critical role in enhancing mathematics education by providing interactive tools and resources. Tools such as online simulations, graphing calculators, and adaptive platforms support students in visualizing abstract concepts and engaging in problem-solving activities, fostering critical thinking and deeper comprehension (Lin, Tseng and Chiang, 2016; Vo, Zhu and Diep, 2017). Through tailored instruction and immediate feedback, blended learning effectively addresses individual learning needs, contributing significantly to improved mathematics achievement. Although the implementation of blended learning presents challenges, such as access to reliable technology, teacher training, and curriculum alignment (Ayob et al., 2023), this study focuses on its demonstrated benefits for mathematics education. By integrating traditional teaching methods with technological innovations, blended learning enhances student engagement, promotes personalized instruction, and positively influences mathematics achievement in diverse educational contexts (Tong, Uyen and Ngan, 2022).

### 5.2 A Variety of Students' Math Achievement in the Environment of Blended Learning

The current meta-analytic review has examined that two moderating factors, such as educational level and digital platform significantly differentiated students' math achievement in the environment of blended learning. Meanwhile, there has not been adequate evidence to conclude that class capacity and geographical location significantly differentiated students' math achievement in the blended learning environment. Each substantial factor is explained in the following subsections.

### *5.2.1 Educational level*

A variety of educational levels significantly differentiate students' math achievement in the environment of blended learning. This was line to Means et al. (2013) who showed that there has been existing evidence to conclude that the factor of educational level differentiated students' academic achievement in the environment of blended learning. Elementary school students benefit more from blended learning because interactive tools like GeoGebra cater to their developmental stage, enabling scaffolded learning and fostering engagement. This is supported by Piaget's theory of cognitive development, which emphasizes the importance of concrete experiences for younger learners. Visual and interactive resources in blended learning help establish foundational mathematical concepts, making the approach particularly effective at this level. In contrast, secondary school and college students engage with more abstract reasoning and complex problem-solving, which may require additional pedagogical strategies beyond blended learning's basic framework. These findings suggest that implementing blended learning at the elementary level maximizes its impact on students' math achievement, while its effectiveness diminishes as the need for advanced cognitive skills increases at higher educational levels.

### *5.2.2 Class capacity*

Class capacity did not significantly differentiate students' math achievement in blended learning environments. This finding aligns with Mawardi, Budiningsih and Sugiman (2023), who observed no significant differences in mathematical skills between students in large and small classes. These reviews show that class capacity is the factor which does not differentiate students' math achievement in the environment of blended learning. Interestingly, descriptive analysis indicates that the effect of blended learning was higher in large classes compared to small classes. This could be attributed to the collaborative nature of blended learning, which fosters peer interactions and leverages the collective knowledge within larger groups. From a theoretical standpoint, Vygotsky's social constructivism supports this observation. In larger classes, students have more opportunities to engage in collaborative learning and benefit from scaffolding provided by both peers and interactive digital tools. Blended learning platforms, such as Google Classroom and Moodle, enhance these interactions by providing structured activities and immediate feedback, ensuring that students remain engaged and supported, regardless of class size. Consequently, the integration of blended learning technologies not only mitigates challenges associated with larger classes but also optimizes students' math achievement by leveraging the social dynamics of learning.

### *5.2.3 Digital platform*

A variety of digital platforms significantly differentiated students' math achievement in the environment of blended learning. This was line to Mawardi, Budiningsih and Sugiman (2023) who revealed that the factor of media platforms significantly differentiated students' mathematical skills in the environment of blended learning. Specifically, the use of GeoGebra software in blended learning environments was found to have a greater effect on improving students' math achievement compared to other digital platforms, such as Edmodo, Google Classroom, LMS, Microsoft Mathematics, Moodle, PowerPoint, Schoology, Video Conference, and WhatsApp. GeoGebra's interactive features, such as real-time graphing, manipulation of equations, and dynamic modeling, enhance engagement and foster a deeper understanding of mathematical concepts. This is supported by Mayer's multimedia learning theory, which emphasizes that learning is most effective when visual and verbal materials are combined in a coherent and meaningful way (Mayer, 2009). GeoGebra stands out because it enables students to experiment directly with mathematical representations, providing immediate feedback and promoting active learning. Unlike text-based platforms, GeoGebra's ability to make abstract concepts more tangible makes it particularly impactful for enhancing students' math achievement.

### *5.2.4 Geographical location*

A variety of geographical location did not significantly differentiate students' math achievement in the environment of blended learning. This was similar to Schmid et al. (2023), who reported no significant differences in academic outcomes across various regions in blended learning contexts. However, descriptive analysis indicated that the effect of blended learning was highest in the United States compared to Asia and Europe. This may be attributed to better access to technology, advanced teacher training programs, and well-established educational policies supporting the integration of digital tools. The consistent effectiveness of blended learning across regions highlights its universal applicability. Nevertheless, these results emphasize critical implications for educational policy. Governments and institutions in regions with lower effects, such as Asia and Europe, could focus on addressing barriers to technology adoption and enhancing teacher professional

development. Providing affordable digital tools and equitable access to resources would help bridge these gaps. To scale blended learning initiatives and address broader disparities in mathematics achievement, public-private partnerships could be explored. These partnerships could provide funding and technical support for infrastructure development. Additionally, policymakers could develop nationwide frameworks to promote collaboration between institutions, enabling the sharing of best practices and resources. These efforts would ensure that the benefits of blended learning are accessible to students globally, regardless of their geographical or socioeconomic background.

### 5.3 Implications to Mathematics Education

The recent meta-analytic review offers several practical implications for mathematics education and future empirical studies. The findings indicate that blended learning significantly enhances students' math achievement. Specifically, using blended learning approaches, mathematics educators including teachers and lecturers can effectively boost student performance. The review highlights that incorporating GeoGebra software into blended learning environments yields a greater positive impact on math achievement compared to other digital platforms. GeoGebra stands out because of its dynamic and interactive features, such as real-time graphing, equation manipulation, and geometric modeling. These features allow students to visualize and interact with mathematical concepts, which enhances understanding and retention. Unlike text-based platforms, GeoGebra transforms abstract mathematical concepts into tangible, interactive models, making it particularly effective in improving math outcomes. Moreover, the review reveals that blended learning is more effective at the primary school level than at the secondary school and college/university levels. Thus, implementing blended learning in elementary schools is likely to be more beneficial for optimizing math achievement compared to higher educational levels. Therefore, mathematics educators should consider applying blended learning strategies, especially with tools like GeoGebra, to maximize student achievement in elementary education.

### 5.4 Limitations and Suggestions

This meta-analytic review has several limitations. Of the total studies identified, only 34 met the inclusion criteria and provided sufficient data for calculating effect sizes. The dataset mainly included studies from specific educational contexts, limiting its generalizability. Many studies also lacked detailed demographic information, such as socioeconomic backgrounds or prior exposure to digital learning tools, which could have enriched the analysis. Additionally, some relevant studies were inaccessible due to publisher restrictions or insufficient statistical reporting. Future research should ensure transparent data reporting and consider publishing in open-access platforms to improve accessibility and inclusion of diverse studies in future meta-analyses.

## 6. Conclusion

Blended learning significantly enhances students' math achievement worldwide. Moreover, its implementation has proven effective in improving math performance across Asia, America, and Europe from 2014 to 2023. Various moderating factors, including educational level and digital platforms, play significant roles in influencing students' math achievement in blended learning settings. Meanwhile, there is insufficient evidence to suggest that class capacity and geographical location significantly differentiate students' math achievement in blended learning settings.

## Acknowledgements

The authors express gratitude to the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia through the Directorate of Research and Community Service (DPPM) and Master's to Doctoral Program for Outstanding Bachelor's Graduates (PMDSU) Batch VII for providing financial support and the opportunity [Contract Number: 082/E5/PG.02.00.PL/2024].

**Ethics Statement:** This study did not require ethical approval as no human subjects were involved.

**AI Statement:** The authors confirm that no Artificial Intelligence tools were utilized in this study.

## References

- Aldalah, O. M. A., Shatat, F., and Ababneh, Z. W., 2019. The impact of blended learning on the development of the cognitive and metacognitive thinking skills in mathematics of the (ECT) students. *Journal of Institutional Research South East Asia*, 17(1).
- Almasi, M., and Zhu, C., 2018. Students' perceptions of social presence in blended learning courses in a tanzanian medical college. *International Journal of Emerging Technologies in Learning*, 13(9).

- Alsahhi, N. R., Al-Qatawneh, S., Eltahir, M., and Aqel, K., 2021. Does blended learning improve the academic achievement of undergraduate students in the mathematics course?: A case study in higher education. *Eurasia Journal of Mathematics, Science and Technology Education*, 17(3), 1–14. <https://doi.org/10.29333/EJMSTE/10781>
- Angreanisita, W., and Mastur, Z., 2021. Mathematical literacy seen from learning independency in blended learning with project based learning assisted by moodle. *Ujmer*, 10(2), 155–161. <http://journal.unnes.ac.id/sju/index.php/ujmer>
- Anwar, S., and Setyaningrum, W., 2021. Can blended learning help improve students' critical thinking skills? *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 10(2), 721. <https://doi.org/10.24127/ajpm.v10i2.3455>
- Apsari, N. P. D. M., 2020. Pengaruh blended learning berbasis video pembelajaran terhadap kemampuan pemecahan masalah ditinjau dari tingkat kecerdasan logis siswa [The influence of video-based blended learning on problem-solving ability in terms of students' logical intelligence levels]. *Suluh Pendidikan*, 18(1), 131–147. <https://doi.org/10.46444/suluh-pendidikan.v18i1.121>
- Ayob, H. H., Daleure, G., Solovieva, N., Minhas, W., and White, T., 2023. The effectiveness of using blended learning teaching and learning strategy to develop students' performance at higher education. *Journal of Applied Research in Higher Education*, 15(3), 650–662. <https://doi.org/10.1108/JARHE-09-2020-0288>
- Ayuningtyas, D. R., and Prastowo, A., 2022. Efektivitas model blended learning untuk meningkatkan kemampuan berpikir kritis matematis siswa sekolah dasar [The effectiveness of the blended learning model in enhancing elementary school students' mathematical critical thinking skills]. *Jurnal Basicedu*, 6(6), 9285–9293. <https://doi.org/10.31004/basicedu.v6i6.3512>
- Azid, N., Ali, R. M., El Khuluqo, I., Purwanto, S. E., and Susanti, E. N., 2022. Higher order thinking skills, school-based assessment and students' mathematics achievement: Understanding teachers' thoughts. *International Journal of Evaluation and Research in Education*, 11(1), 290–302. <https://doi.org/10.11591/ijere.v11i1.22030>
- Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., and Abrami, P. C., 2014. A meta-analysis of blended learning and technology use in higher education: from the general to the applied. *Journal of Computing in Higher Education*, 26(1), 87–122. <https://doi.org/10.1007/s12528-013-9077-3>
- Bonk, C. J., and Graham, C. R., 2012. *The handbook of blended learning: Global perspectives, local designs*. Wiley+ ORM.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R., 2021. *Introduction to meta-analysis*. John Wiley & Sons.
- Cao, W., 2023. A meta-analysis of effects of blended learning on performance, attitude, achievement, and engagement across different countries. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1212056>
- Cronje, J., and Van Zyl, I., 2022. WhatsApp as a tool for building a learning community. *Electronic Journal of e-Learning*, 20(3), pp296-312. <https://doi.org/10.34190/ejel.20.3.2286>
- Darmono, P. B., and Maryam, I., 2019. Pengaruh blended learning berbantuan microsoft mathematic terhadap kemampuan berpikir kritis mahasiswa [The influence of blended learning assisted by microsoft mathematics on students' critical thinking skills]. *PRISMA, Prosiding Seminar Nasional Matematika*, 2, 583–588.
- Ektafia, F., Fitri, A., and Najibufahmi, M., 2021. Penerapan e-learning berbasis schoology terhadap kemampuan pemahaman konsep matematika siswa SMP N 1 sragi [The implementation of schoology-based e-learning on the mathematical concept understanding ability of SMP N 1 Sragi students]. *Seminar Nasional Pendidikan Matematika*, 2(1), 255–262.
- Fardian, D., Suryadi, D., and Prabawanto, S., 2025. A praxeological analysis of linear equations in Indonesian mathematics textbooks: Focusing on systemic and epistemic aspect. *Journal on Mathematics Education*, 16(1), 225–254. <https://doi.org/10.22342/jme.v16i1.pp225-254>
- Fazal, M., and Bryant, M., 2019. Blended learning in middle school math. *Journal of Online Learning Research*, 5(1), 49–64.
- Garrison, D. R., and Kanuka, H., 2004. Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, 7(2), 95–105.
- Graham, C. R., 2006. Blended learning systems. *The Handbook of Blended Learning: Global Perspectives, Local Designs*, 1, 3–21.
- Gredler, M. E., 2011. Understanding Vygotsky for the classroom: Is it too late? *Educational Psychology Review*, 24(1), 113–131. <https://doi.org/10.1007/s10648-011-9183-6>
- Harwell, M., 2020. Growth in the amount of literature reviewed in a meta-analysis and reviewer resources. *Mid-Western Educational Researcher*, 32(1), 31–47.
- Helsa, Y., Suparman, Juandi, D., Turmudi, and Ghazali, M. B., 2023. A meta-analysis of the utilization of computer technology in enhancing computational thinking skills : Direction for mathematics learning. *International Journal of Instruction*, 16(2), 735–758.
- Hrastinski, S., 2019. What do we mean by blended learning? *TechTrends*, 63(5), 564–569. <https://doi.org/10.1007/s11528-019-00375-5>
- İDİL, Ş., GÜLEN, S., and DÖNMEZ, İ., 2024. What should we understand from PISA 2022 results? *Journal of Steam Education*, 7(1), 1–9. <https://doi.org/10.55290/steam.1415261>
- Indrapangastuti, D., Surjono, H. D., and Yanto, B. E., 2021. Effectiveness of the blended learning model to improve students achievement of mathematical concepts. *Journal of Education and E-Learning Research*, 8(4), 423–430. <https://doi.org/10.20448/journal.509.2021.84.423.430>
- Jayanti, J., and Rahmawati, R., 2017. Model pembelajaran generatif (MPG) berbantuan blended learning pada trigonometri untuk meningkatkan kemampuan pemecahan masalah matematis mahasiswa PGRI [Generative learning model (MPG) assisted by blended learning in trigonometry to improve mathematical problem-solving ability of PGRI students. *Jurnal Pendidikan Matematika*, 2(November), 82–97.

- Khairiyyah, A., Mulyono, and Fauzi, K. M. A., 2021. The learning effect of blended learning based on google class room and initial mathematics on mathematic representation and resilience of students in the Covid-19 pandemic. *Britain International of Linguistics Arts and Education (BLoLAE) Journal*, 3(1), 63–76. <https://doi.org/10.33258/biolae.v3i1.410>
- Kilpatrick, J., 2001. Understanding mathematical literacy: The contribution of research. *Educational Studies in Mathematics*, 47(1), 101–116. <https://doi.org/10.1023/A:1017973827514>
- Kintu, M. J., Zhu, C., and Kagambe, E., 2017. Blended learning effectiveness: the relationship between student characteristics, design features and outcomes. *International Journal of Educational Technology in Higher Education*. <https://doi.org/10.1186/s41239-017-0043-4>
- Kiviniemi, M. T., 2014. Effects of a blended learning approach on student outcomes in a graduate-level public health course. *BMC Medical Education*, 14(1), 1–7. <https://doi.org/10.1186/1472-6920-14-47>
- Leinwarnd, S. E., 2014. National council of teachers of mathematics. *Principles Ro Actions: Ensuring Mathematical Success for All*. Reston: VA: Author.
- Lin, Y.-W., Tseng, C.-L., and Chiang, P.-J., 2016. The effect of blended learning in mathematics course. *Eurasia Journal of Mathematics Science and Technology Education*. <https://doi.org/10.12973/eurasia.2017.00641a>
- Machumu, H. J., Zhu, C., and Sesabo, J. K., 2016. Blended learning in the vocational education and training system in Tanzania: Understanding vocational educators' perceptions. *International Journal of Multicultural and Multireligious Understanding*, 3(2), 30–45.
- Mashuri, S., and Nasrum, A., 2020. Efek pembelajaran tambahan menggunakan schoology pada mata kuliah kalkulus [The effect of supplementary learning using schoology in calculus courses]. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 9(3), 561–569.
- Mawardi, D. N., Budiningsih, C. A., and Sugiman., 2023. Blended learning effect on mathematical skills: A meta-analysis study. *Ingenierie Des Systemes d'Information*, 28(1), 197–204. <https://doi.org/10.18280/isi.280122>
- Mayer, R. E., 2009. *Multimedia Learning*. <https://doi.org/10.1017/cbo9780511811678>
- Means, B., Toyama, Y., Murphy, R., and Baki, M., 2013. The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, 115(March), 1–47.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Group\*, P., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–269.
- Muncarno, and Astuti, N., 2021. Pengaruh model pembelajaran blended learning terhadap kemampuan berpikir kritis matematika peserta didik sekolah dasar [The influence of the blended learning model on elementary school students' mathematical critical thinking skills]. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 10(4), 2784–2790.
- Mutaqin, A., Marethi, I., and Syamsuri, S., 2016. Model blended learning di program studi pendidikan matematika UNTIRTA [Blended learning model in the mathematics education study program at UNTIRTA]. *Jurnal Cakrawala Pendidikan*, 35(1).
- Nasution, E. Y. P., Sintia, L., and Putri, R., 2022. The effect of blended learning models assisted by video tutorials on students' critical thinking ability in mathematics learning. *Logaritma : Jurnal Ilmu-Ilmu Pendidikan Dan Sains*, 10(01), 1–22. <https://doi.org/10.24952/logaritma.v10i01.5186>
- Nguyen, T., Netto, C. L. M., Wilkins, J. F., Bröcker, P., Vargas, E. E., Sealfon, C. D., Puthipiroj, P., Li, K. S., Bowler, J. E., Hinson, H. R., Pujar, M., and Stein, G. M., 2021. Insights into students' experiences and perceptions of remote learning methods: From the COVID-19 pandemic to best practice for the future. *Frontiers in Education*, 6. <https://doi.org/10.3389/feduc.2021.647986>
- Nida, N. K., Usodo, B., and Saputro, D. R. S., 2020. The blended learning with Whatsapp media on mathematics creative thinking skills and math anxiety. *Journal of Education and Learning (EduLearn)*, 14(2), 307–314. <https://doi.org/10.11591/edulearn.v14i2.16233>
- Noviyanti, F., Sugiharta, I., and Farida, F., 2019. Analisis kemampuan pemecahan masalah matematis : dampak blended learning menggunakan edmodo [Analysis of mathematical problem-solving ability: the impact of blended learning using edmodo]. *Desimal: Jurnal Matematika*, 2(2), 173–180. <https://doi.org/10.24042/djm.v2i2.4035>
- Nugraha, D. G. A. P., Astawa, I. W. P., and Ardana, I. M., 2019. Pengaruh model pembelajaran blended learning terhadap pemahaman konsep dan kelancaran prosedur matematis [The influence of the blended learning model on concept understanding and mathematical procedural fluency]. *Jurnal Riset Pendidikan Matematika*, 6(1), 75–86. <https://doi.org/10.21831/jrpm.v6i1.20074>
- OECD., 2023. *PISA 2022 Results (Volume II)*. <https://doi.org/10.1787/a97db61c-en>
- Paas, F., Van Gog, T., and Sweller, J., 2010. Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review*, 22, 115–121.
- Pertiwi, A., Kariadinata, R., Juariah, J., Sugilar, H., and Ramdhani, M. A., 2019. Edmodo-based blended learning on mathematical proving capability. *Journal of Physics: Conference Series*, 1157(4). <https://doi.org/10.1088/1742-6596/1157/4/042001>
- Pokorný, M., 2019. Blended learning can improve the results of students in combinatorics and data processing. In 2019 international symposium on educational technology (ISET) (pp. 207–210). IEEE.
- Putri, A. D., Juandi, D., and Turmudi, T., 2024. Realistic mathematics education and mathematical literacy: a meta-analysis conducted on studies in Indonesia. *Journal of Education and Learning (EduLearn)*, 18(4), 1468–1476. <https://doi.org/10.11591/edulearn.v18i4.21650>

- Putri, A. D., Yerizon, Y., Arnellis, A., and Suherman, S., 2024. Development of realistic mathematics education-based teaching materials to increase students' mathematical literacy ability. *AIP Conference Proceedings*, 3024(1). <https://doi.org/10.1063/5.0204587>
- Samritin, S., Susanto, A., Manaf, A., and Hukom, J., 2023. A meta-analysis study of the effect of the blended learning model on students' mathematics learning achievement. *Jurnal Elemen*. <https://doi.org/10.29408/jel.v9i1.6141>
- Satriani, R. D., Wangid, M. N., and PA, P., 2020. Pengaruh edmodo terhadap pemahaman konsep matematika dan kemandirian belajar mahasiswa [The influence of edmodo on students' mathematical concept understanding and learning independence]. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 9(4), 1137. <https://doi.org/10.24127/ajpm.v9i4.3181>
- Schmid, R. F., Borokhovski, E., Bernard, R. M., Pickup, D. I., and Abrami, P. C., 2023. A meta-analysis of online learning, blended learning, the flipped classroom and classroom instruction for pre-service and in-service teachers. *Computers and Education Open*, 5(January), 100142. <https://doi.org/10.1016/j.caeo.2023.100142>
- Seage, S. J., and Türegün, M., 2020. The effects of blended learning on STEM achievement of elementary school students. *International Journal of Research in Education and Science*, 6(1), 133–140. <https://doi.org/10.46328/ijres.v6i1.728>
- Septiyan, I., Anriani, N., and Hendrayana, A., 2019. Perbandingan model discovery learning dan blended learning terhadap pencapaian [Comparison of discovery learning and blended learning models on achievement]. *Penelitian Pengajaran Matematika*, 1(2), 75–93.
- Setiawan, A. A., Muhtadi, A., and Hukom, J., 2022. Blended learning and student mathematics ability in Indonesia: A meta-analysis study. *International Journal of Instruction*, 15(2), 905–916. <https://doi.org/10.29333/iji.2022.15249a>
- Setiyani, S., 2019. Blended learning: the effectiveness of schoology based e-learning on mathematic communication ability. *Jurnal Kependidikan: Penelitian Inovasi Pembelajaran*, 3(2), 143–155. <https://doi.org/10.21831/jk.v3i2.23820>
- Setyaningrum, W., 2018. Blended learning: Does it help students in understanding mathematical concepts? *Jurnal Riset Pendidikan Matematika*. <https://doi.org/10.21831/jrpm.v5i2.21428>
- Staker, H., and Horn, M. B., 2012. *Classifying K–12 blended learning*.
- Sudiarta, I. G. P., and Sadra, I. W., 2016. Pengaruh model blended learning berbantuan video animasi terhadap kemampuan pemecahan masalah dan pemahaman konsep siswa [The influence of a blended learning model assisted by animated videos on students' problem-solving ability and concept understanding]. *Jurnal Pendidikan Dan Pengajaran*, 49(2), 48. <https://doi.org/10.23887/jppundiksha.v49i2.9009>
- Suparman, S., and Juandi, D., 2022. Self-Efficacy and Mathematical Ability: A Meta-Analysis of Studies Conducted in Indonesia. *Pedagogika*, 147(3), 26–57. <https://doi.org/10.15823/p.2022.147.2>
- Supriadi, N., Kusumah, Y. S., Sabandar, J., and Afgani, J. D., 2014. Developing high-order mathematical thinking competency on high school students ' through geogebra-assisted blended learning. *Mathematical Theory and Modelling*, 4(6), 57–66.
- Tawaldi, S., Nurlaelah, E., Juandi, D., and Suparman., 2023. Is mathematics anxiety related to mathematics learning? A meta-analysis. *MSCEIS 2021*, 090044, 1–10. <https://doi.org/https://doi.org/10.1063/5.0155846>
- Tong, D. H., Uyen, B. P., and Ngan, L. K., 2022. The effectiveness of blended learning on students' academic achievement, self-study skills and learning attitudes: A quasi-experiment study in teaching the conventions for coordinates in the plane. *Heliyon*, 8(12). <https://doi.org/10.1016/j.heliyon.2022.e12657>
- Trisnayanti, N. P. E., Sariyasa, and Suweken, G., 2020. Pengaruh model pembelajaran blended learning terhadap pemahaman konsep dan motivasi belajar siswa [The effect of blended learning model on students' concept understanding and learning motivation]. *Jurnal MathEdu: Mathematic Education Journal*, 3(3), 1–8.
- Vo, H. M., Zhu, C., and Diep, N. A., 2017. The effect of blended learning on student performance at course-level in higher education: A meta-analysis. *Studies in Educational Evaluation*, 53, 17–28.
- Yaghmour, K. S., 2016. Effectiveness of blended teaching strategy on the achievement of third grade students in mathematics. *Journal of Education and Practice*, 7(5), 65–73.
- Zein, M., M. Nuh, Z., Dardiri, D., Jasril, J., Candra, R. M., Hanafi, I., and Thahir, M., 2019. Hybrid learning in mathematics learning: Experimental study in SMA Negeri 1 Pekanbaru. *Malikussaleh Journal of Mathematics Learning (MJML)*, 2(2), 56–60. <https://doi.org/10.29103/mjml.v2i2.2009>

# Virtual – Augmented Reality (VAR) for Science Learning: Development and Impact on Students' HOTS Skills

Iwan Maulana<sup>1</sup>, Siswandari<sup>2</sup>, Gunarhadi<sup>2</sup> and Agus Efendi<sup>2</sup>

<sup>1</sup>Education Science – Universitas Sebelas Maret, Indonesia

<sup>2</sup>Sebelas Maret University, Indonesia

[maulana\\_93@student.uns.ac.id](mailto:maulana_93@student.uns.ac.id)

[siswandari@staff.uns.ac.id](mailto:siswandari@staff.uns.ac.id)

[gunarhadi@staff.uns.ac.id](mailto:gunarhadi@staff.uns.ac.id)

[agusefendi@staff.uns.ac.id](mailto:agusefendi@staff.uns.ac.id) (corresponding author)

<https://doi.org/10.34190/ejel.23.1.3733>

An open access article under [CC Attribution 4.0](#)

**Abstract:** This study focuses on the role of innovation in educational technology, especially the use Virtual – Augmented Reality (VAR) at the junior high school level. The COVID-19 pandemic has widened the existing gap associated with technology and learning, calling for the need to provide interactive learning tools to develop students' higher-order thinking skills (HOTS). This study aimed to design and determine the effectiveness of VAR learning tools on students' HOTS skills. The research method used was Research and Development (R&D) using the Alessi and Trollip model, which includes three steps: planning, design, and development. The research population consisted of 242 junior high school students from Surakarta City, Indonesia, who were recruited through cluster sampling. Observations, interviews, and product feasibility questionnaires were used as instruments. The results showed that the students' HOTS skills improved significantly with the use of VAR-integrated learning tools in science subjects. The learners' mean scores on the post-test had a marked rise as opposed to the pre-test, signifying the effectiveness of the medium in greatly augmenting HOTS skills. This study suggests that VAR-based learning tools can be used as an effective means to overcome the digital divide and improve the quality of education by developing students' higher-order thinking skills, which are essential for 21st century learning.

**Keywords:** VAR, Science learning, HOTS, Technology, Skills

## 1. Introduction

The COVID-19 pandemic has caused significant tragic changes in human paradigms and habits. It goes without saying that This event fundamentally affected global education, requiring an immediate shift from physical to online or e-learning (Rakhmetov et al., 2022). However, this transition has not been universal, as students from families with low socioeconomic status (SES) face considerable challenges in obtaining the necessary digital tools and stable internet connections, which hinders their ability to participate in online learning (Bismala and Manurung, 2021). Students from families with lower socioeconomic status often face challenges in accessing digital devices and the Internet, which restricts their participation in e-learning activities (Yao et al., 2022). In addition, many teachers are found to have low digital competency, which affects their ability to teach in an online environment. This problem exists because not all teachers have a reasonable level of digital skills, and they need to transition to a new digital environment (Garad, Al-Ansi and Qamari, 2021). These issues highlight the urgent need for alternative learning innovations that can bridge the digital divide and enhance learning experiences for all students.

One potentially effective alternative is the implementation of Virtual – Augmented Reality (VAR) technologies in the teaching of science subjects because they provide engaging learning experiences beyond the reach of conventional digital facilities (Rukayah et al., 2022; Kamińska et al., 2019). As stated previously, Augmented Reality (AR) technology is an enhancement of the real world with virtual elements in real-time interactions, in which users are provided with digital information that interacts with their environment (Septinaningrum et al., 2022). Virtual Reality technology refers to the ability to immerse the user in an interactive virtual world, severing the limitations imposed by the real world, and permitting the user to engage with the virtual environment without any restrictions (Raja and Priya, 2021). In this study, we introduce VAR as a new interactive media that integrates both Virtual Reality (VR) and Augmented Reality (AR) into a single platform. Unlike conventional VR or AR applications, VAR provides users with an immersive learning experience by leveraging the strengths of both technologies. These immersive tools have revolutionized the educational landscape by providing innovative ways to enhance learning experiences (Elmqaddem, 2019). In addition, VAR can foster higher-order thinking skills (HOTS) by motivating learners to analyze, evaluate, and create in virtual environments on essential skills needed in science education and learning in the 21st century (Lee et al., 2022). Nevertheless, there is little

empirical evidence for integrated VAR technology to improve students' HOTS in junior secondary level science classes. This highlights the necessity for additional investigation of how VAR can close existing educational gaps and aid in the development of vital intellectual abilities in learners.

Looking at some of the information above, it cannot be denied that the use of AR and VR in education has revolutionized teaching and learning methods, making the learning experience more interactive, impressive, and fun (Saravanan et al., 2022). In addition, AR and VR help in real contextual learning in history, for example, where students can visit historical sites to learn, while simultaneously helping in remote education for people who are in distant places and are able to use supporting technological devices (Serin, 2020). As stated, AR and VR technologies have become indispensable in the ongoing educational revolution alongside Information Communication Technology (ICT) since they have improved the overall quality of education and the effectiveness with which learners achieve the set educational outcomes. Nonetheless, there is still a need to systematically investigate the effectiveness of these technologies across diverse distances and their relevance to the essential cognitive skills of critically analyzing and reasoning, known as HOTS.

To successfully integrate VAR technology into educational systems, institutions face multiple challenges, including operating within a constrained budget, which can make acquiring the necessary software and equipment exceedingly difficult (Scavarelli, Arya and Teather, 2021). Support and training from auxiliary technical staff are also required. Many teachers lack the ability to effectively utilize VAR tools, which poses a challenge, as educators must be trained to build their confidence and skills (Yusof, 2019; Oleksiuk and Oleksiuk, 2020). The adoption rates at the beginning tend to be low, as research has shown that learners rarely utilize VAR laboratories for engineering subjects, signifying the early phases of innovation dissemination (Yanto et al., 2022). While these findings highlight technical and financial barriers to VAR adoption, they do not fully account for pedagogical constraints or sociocultural factors that may influence the effectiveness of VAR-based learning. Therefore, additional studies are needed to explore these dimensions in greater depth.

Alignment with national educational objectives greatly affects acceptance and attitudes, indicating that teachers' feelings are pivotal as well (Bima, Saputro and Efendy, 2021). The need for constant assessment and modifications to ensure that VAR meets educational goals and its subsequent integration into existing educational systems adds to the already existing complexity (Shepa et al., 2021; Wang et al., 2023). In contrast to past research that has mostly examined barriers to implementation, few studies have looked at the equity of benefits from VAR learning within different socioeconomic contexts and the impact of these diverse contexts on students. To meet these challenges, there must be sufficient funds, adequate training, infrastructure improvement, and support to enable successful VAR integration into education. In addition, there is still no answer to the question of which practices are most effective in the use of VAR aimed at maximizing students' educational achievements.

Studies have revealed that the adoption and effectiveness of VAR technologies in educational settings are greatly impeded by inherent complications. In Wibowo et al.'s (2021) study, VAR adoption was linked to other factors in secondary schools, such as the allocation and management of resources, the presence of adequate teacher support, and the attitudes of teachers towards the technology, indicating the presence of other influences (Wibowo et al., 2021). Munje and Jita (2020) suggested that in Africa, the lack of adequate training, insufficient infrastructure, and VAR adoption are all interconnected with a more general lack of adequate power supply. There is also an inability to systematically assess the appropriateness of a given technology to the schools or regions that compose the problem. These studies are helpful in clarifying the logistical and financial impediments to adopting VAR technologies, but they miss how different VAR pedagogies can be employed to develop HOTS. This deficiency marks a shift towards a more comprehensive reconsideration of teaching strategies designed to optimize the outcome of VAR in different classroom environments. The proposed solution emphasizes the development of holistic strategies that address the complex interrelationship between inadequate training, infrastructure, and user-centered strategies, while providing thorough evaluations of VAR integration into educational practices. Instead of checking the plausibility of adopting VAR technology, further investigations should focus on the impact of different teaching styles on the integration of such technologies, especially in developing HOTS for learners from different educational settings.

Although there are numerous challenges that some countries may face in implementing new technologies, such as AR and VR, several studies have shown that in some developed countries, AR and VR are being utilized for educational activities. A comprehensive review conducted by previous researchers successfully explored deep applications of VR in the educational context. Research has indicated that there is potential for VR to create engaging and interactive learning environments that can enhance knowledge retention and student motivation

(Strand, 2020). In line with this, another study asserted that VR has an impact on science education, noting that VR can significantly enhance students' attitudes and learning motivation, especially in immersive virtual environments designed to improve students' learning outcomes (Al-Amri, Osman and Musawi, 2020). However, there is a lack of consensus on the extent to which motivation and engagement directly contribute to HOTS development, as some studies have suggested that increased engagement does not necessarily translate into deeper cognitive processing. This discrepancy suggests that further research is needed to clarify the relationship between student motivation, engagement, and HOTS acquisition in VAR-based learning. Furthermore, a meta-analysis study has also highlighted the advancements in AR in education over the past decade, with a focus on its implementation in STEM-based learning to support scientific literacy and enhance student engagement and understanding (Pedaste, Mitt and Jürivete, 2020). AR and VR technology can also be said to be flexible because they can be accessed and integrated via smartphones (Muñoz-Saavedra, Miró-Amarante, and Domínguez-Morales, 2020; Bukhori et al., 2019), where currently almost the majority of students already have smartphones. Despite these advantages, there is still limited research investigating how mobile-based AR and VR applications compare with full-scale VAR systems in fostering HOTS. This distinction is critical, as mobile technology may provide a more accessible, yet potentially less immersive, learning experience.

One important skill for students is higher-order thinking skills (HOTS), which comprise the core cognitive activities of analysis, evaluation, and synthesis. These skills support the achievement of complex problem-solving activities and advanced thinking skills. These skills allow students to transfer skills, solve problems, and engage in critical and creative thinking (Astrid and Hasanah, 2022). As a goal of modern education, the acquisition of HOTS has accelerated in the present millennial era, aiming to enhance students' ability to analyze, evaluate, and create (Tong et al., 2022). For example, in mathematics education, instruments for assessing the HOTS have been developed to measure these skills. These instruments were tested for reliability and their difficulty index to be used in judging the students' ability to analyze (C4), evaluate (C5), and create (C6), as described in Bloom's taxonomy (Zainil et al., 2023; Muhayimana, Kwizera, and Nyirahabimana, 2022). Within the framework of teaching English as a foreign language, HOTS integrated into examinations is aimed at developing high school students' critical and creative thinking skills. Examination analysis showed that there are many questions that seek to assess critical thinking skills, which are then designed to test creative thinking and problem-solving abilities (Permana et al., 2019; Arifin, 2020). In addition, the application of the HOTS in elementary schooling is still in its early stages, but it shows good prospects, especially with the development of valid, practical, and reliable assessment instruments that allow teachers to support the development of these skills among fifth graders (Parmiti, Antara and Wulandari, 2022). The integration of HOTS in its dimensions of planning, implementation, and assessment enables the learners' activities to be not only designed to be effortless, but also stimulating and meaningful to ensure higher-order thinking is developed.

Incorporating HOTS into educational settings is crucial for preparing students to tackle complex problems with enhanced cognitive skills. Higher-order thinking skills are not limited to knowledge recall; they encompass problem-solving, critical thinking, creative thinking, argumentation, and decision-making, which are fundamental for 21st century skills (Yusof, 2019). Several studies have revealed that HOTS-based assessment tools notably improve students' cognitive, affective, and psychomotor skills, allowing them to comprehend and apply scientific concepts (Ramos, Dolipas and Villamor, 2013). For example, HOTS-based worksheets in trigonometry have proven to be valid, practical, and effective in enhancing problem-solving abilities among high school students (Maryani et al., 2021; Tanjung, Nababan and Sa'dijah, 2020). In the context of science education, HOTS play a crucial role in helping students grasp abstract concepts, conduct experiments effectively, and cultivate a scientific mindset, which is fundamental for innovation and technological advancement. Because of these expectations, infusing higher-order thinking skills (HOTS) can no longer be considered optional when preparing learners for an ever-changing world. The development of VAR technologies provides a new avenue for improving HOTS through more engaging and convincing active learning, which ensures a higher level of comprehension and participation. VAR allows experiential learning through simulation and live interaction, which enables students to investigate scientific processes and use critical thinking to explain them in ways that are not possible in traditional teaching (Oleksiuk and Oleksiuk, 2020).

Another investigation highlighted the prospect of AR in depicting science concepts that are highly abstract and can pose difficulties to students with various educational backgrounds and teaching experiences (Yang, 2023). In addition, self-directed learning, creative thinking, critical thinking, and knowledge creation efficacy skills were positively affected by the use of AR under QIMS, particularly for low-achieving students (Syawaludin, Gunarhadi and Rintayati, 2019). Along the same lines, Wehrmann and Zender (2024) noted that VAR provides useful opportunities for immersive learning and helps overcome gaps in understanding concepts, as well as boosting

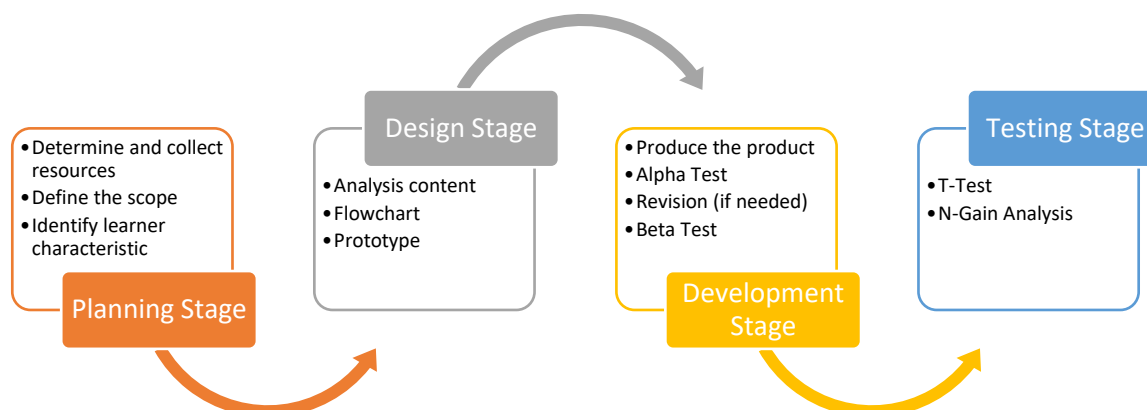
the interest of students, especially in science education. Combining these technologies transforms education with VAR, which can solve well-known issues related to traditional teaching methods by providing simpler and more attractive ways to present subjects (Averbeck et al., 2019; Averbeck et al., 2024). The results change the perception of science education by using VR, AR, and VAR technology for junior high school children, which provides hope for better interactive and immersive learning experiences.

Drawing from several analyses of previous studies, it appears that, despite their benefits, the application of VAR technologies continues to face significant challenges, particularly in developing countries. The widening digital divide exacerbated by the COVID-19 pandemic has created significant barriers, particularly in rural areas that have limited infrastructure and connectivity. Furthermore, the lack of digital literacy among teachers and students from low socioeconomic status (SES) backgrounds serves as a barrier to the effective implementation of VAR-based learning. While later studies emphasized the impact of AR and AI technologies on the improvement of motivational and conceptual understanding, adoption is hindered by financial barriers, inadequate training, and difficulty integrating these technologies into preexisting educational systems (Mariscal et al., 2020; Wehrmann and Zender, 2024). Apart from these gaps, other outstanding issues include the absence of pertinent research on the application of VAR in education and the development of concepts and functioning of science education at the junior high school level, especially the integration of higher-order thinking skills and science education. There is also a lack of research on the use of VAR in educational settings with limited resources that face issues, such as low levels of infrastructure development and digital literacy. Therefore, there is a pressing need for well-defined steps to enable the effective use of VAR in education so that it is not focused merely on knowledge acquisition but on the development of HOTS, which is critical for 21st-century learners.

To address this gap, this study examines whether the use of VAR improves HOTS in junior high school science education. More specifically, it designs and assesses a problem-solving, analytical thinking, and cognitive development-enhancing learning model that is VAR-based. The research is directed by the question, "Does the use of VAR technology improve the HOTS in junior high school science education?". This study is part of an effort to address the educational digital divide, with a particular focus on how VAR facilitates the strengthening of HOTS as a crucial 21st century competency. VAR technology's impact on devising strategies for HOTS is evaluated to address the gap within educational literature and transform its findings into practical guidance for the use of immersive learning technologies.

## 2. Method

This study utilizes the Research and Development (R&D) approach, which is consistent with the study's objective of creating a VAR, which is suitable for enhancing higher-order thinking skills (HOTS) among junior high school students (Gall, Gall and Borg, 2003). This study adopts the development model proposed by Alessi and Trollip (Alessi and Trollip, 2001). Overall, this development model comprises three main stages: planning, design, and development (Yaniawati et al., 2021). The scope and limitations of the product were determined during the planning stage. The design stage involved the design of content and documentation. The development stage involved the production of media, validation of media feasibility, and testing of media to measure the impact of the developed product. All the steps for this research and development summarized in figure 1.



**Figure 1: Research and Development Model**

The participants of this study were high school students and teachers who taught the subject of natural science. Therefore, the subjects involved in this research consisted of 242 students from six junior high schools in Solo

Raya, including Junior High School (SMP) Kalam Kudus Surakarta, SMP Tarakanita Sukoharjo, SMP Widya Wacana Surakarta, SMP Negeri 3 Surakarta, SMP Negeri 4 Surakarta, and SMP Negeri 1 Masaran Sragen. The students were selected using a cluster sampling technique (Rejeginingsih, Sudiyanto and Budiarto, 2022). Cluster sampling was chosen to ensure diverse representations across multiple schools while maintaining the practical feasibility of implementing the VAR-based learning model. This method allows for efficient selection of participants within naturally occurring classroom groups, reducing logistical complexity compared to simple random sampling. For the development phase of the product feasibility assessment, 24 students, three teachers, two media experts, and two subject matter experts were selected to evaluate the quality of the product. Schools were chosen based on their relevance to the study and representativeness of different school environments (Saputri, Rukayah and Indriayu, 2018).

Data collection was conducted concurrently with product development activities using non-test techniques, including observations, interviews, and product feasibility questionnaires (Budiyo, 2017). The instruments used included stratified Likert-like scale questionnaires derived from Dinayusadewi and Agustika (2020) to measure the feasibility of VAR products, which were estimated using a 4-point like scale where 1 means 'Strongly Disagree' and 4 means 'Strongly Agree'. The neutral point for the response set was eliminated to ensure that all the respondents offered their opinions. Moreover, an observation checklist was used to assess the learners' participatory level during the lesson, while the teachers' interviews gave us some impression on the effectiveness of the lesson media qualitatively (Setiawan, Pusporini and Dardjito, 2020). Data collection techniques to measure students' HOTS skills were performed through tests (Fuad et al., 2017). The tests included topics students were learning in science during junior high school and provided a scope of research on assessing HOTS for junior high school students (Rofiah, Aminah and Sunarno, 2018; Widiyawati et al., 2021; Dasilva et al., 2019).

The validity of the assessment was determined using content validity through expert evaluations from the field of education (Surbhi, 2019). The expert judgment process ensured that the instrument was aligned with Bloom's taxonomy and the cognitive levels required for the HOTS assessment. Moreover, construct validity was verified with factor analysis, confirming that the instrument synergistically measured the separate components of analysis (C4), evaluation (C5), and creation (C6) of higher-order thinking skills (HOTS) within their respective constituent parts. The outcomes shown in Table 1 illustrate that all parameters of the indicators of higher order thinking skills (HOTS) were valid and fulfilled the requirements.

**Table 1: Validity and Reliability Test Results of HOTS Instrument for Science Learning**

| Indicator                | Validity Method                            | Results      | Reliability Method                      | Cronbach's Alpha |
|--------------------------|--------------------------------------------|--------------|-----------------------------------------|------------------|
| <b>Analysis (C4)</b>     | Expert Judgment (Content Validity)         | Valid        | Internal Consistency (Cronbach's Alpha) | 0.78             |
| <b>Evaluation (C5)</b>   | Factor Analysis (Construct Validity)       | Valid        | Internal Consistency (Cronbach's Alpha) | 0.81             |
| <b>Creation (C6)</b>     | Expert Judgment (Content Validity)         | Valid        | Internal Consistency (Cronbach's Alpha) | 0.85             |
| <b>Overall HOTS Test</b> | Combined Expert Judgment & Factor Analysis | Highly Valid | Overall Reliability                     | 0.81 (Reliable)  |

Simultaneously, the reliability of the instrument was checked using Cronbach's alpha (internal consistency). The obtained reliability coefficient was 0.81, which means that the instrument is reliable for measuring students' HOTS skills (Widyaningsih et al., 2020). The reliability scores of each of the HOTS indicators ranged from 0.78 to 0.85, which is acceptable, providing assurance of the consistency of most students' HOTS assessment results over different attempts to administer the tests. This thorough validation and reliability analysis increased the trustworthiness of the instrument in measuring the impact of VAR-based learning on students' HOTS development in science education.

Both descriptive and inferential statistical methods were used to analyze the collected data. The interviews and observations of the learning activities were analyzed using descriptive analysis (Patten and Newhart, 2018). The observations examined included student engagement, problem solving, collaboration with VAR, and interaction with the content that reflects cognitive engagement and learning behavior that provide value. Thematic analysis of interview data was performed through coding, categorization, and identifying patterns regarding VAR implementation, students' perceptions, and teaching effectiveness. Statistical Descriptive Analysis is also

relevant when considering the feasibility of media VAR, as validated by the Teacher, Media Expert, and Content Expert. The assessment of media feasibility involved computing the total score values to find the mean score percentages (Septinaningrum et al., 2022). For the VAR to be declared suitable as a learning medium, table 1 shows that the feasibility level analysis must obtain a minimum percentage score of 63%. table 2 shows the conversion criteria for the feasibility of VAR.

**Table 2: VAR eligibility criteria**

| Percentage | Qualification  | Decision      |
|------------|----------------|---------------|
| 81 – 100%  | Very good      | Very Eligible |
| 61 – 80%   | Good           | Eligible      |
| 41 – 60%   | Enough         | Less Eligible |
| 21 – 40%   | Deficient      | Not Eligible  |
| 1 – 20%    | Very Deficient | Rejected      |

This study employed a pre-test and post-test control group design to evaluate the effectiveness of VAR-based learning in enhancing Higher-Order Thinking Skills (HOTS). The experimental group used VAR-based learning, whereas the control group followed conventional methods, allowing for a direct comparison of learning outcomes.

Based on this design, the research hypotheses are formulated as follows:

- **H<sub>0</sub> (Null Hypothesis):** VAR-based learning has **no significant effect** on students' HOTS compared to conventional methods.
- **H<sub>1</sub> (Alternative Hypothesis):** VAR-based learning has a **significant positive effect** on students' HOTS compared to conventional methods.

To test these hypotheses, paired sample t-tests were used to analyze improvements within each group, while independent sample t-tests were used to compare post-test scores between the groups. Additionally, N-gain analysis was used to measure the magnitude of learning improvement. Table 3 shows the pre/post-test design used in this study.

**Table 3: Research Design: Pre-Test and Post-Test Control Group**

| Group        | Pre-Test | Treatment                     | Post-Test |
|--------------|----------|-------------------------------|-----------|
| Experimental | ✓        | VAR-Based Learning Model      | ✓         |
| Control      | ✓        | Conventional Learning Methods | ✓         |

To assess learning effectiveness, an inferential analysis was conducted using a t-test and N-gain analysis (Supianti et al., 2022). The t-test was used to test the research hypothesis, analyzing improvements within groups (paired sample t-test) and between groups (independent sample t-test) to determine whether VAR-based learning significantly impacted HOTS. N-Gain analysis measured normalized learning gains based on the following criteria, as displayed in table 4.

**Table 4: N-Gain Criteria**

| N-Gain (g)         | Interpretation |
|--------------------|----------------|
| $g > 0.7$          | High           |
| $0.3 < g \leq 0.7$ | Medium         |
| $g < 0.3$          | Low            |

Adapted from (Winarni and Purwandari, 2019)

The VAR product was specifically developed for this study using the Alessi and Trollip R&D model, integrating augmented reality (AR) and virtual reality (VR) simulations to enhance spatial and conceptual visualization. The key features include interactive 3D models, real-time feedback, and problem-based scenarios designed to foster HOTS development. These methodological refinements provide a clear justification for the research design, statistical analyses, and interventions used, ensuring an in-depth assessment of VAR's impact of VAR on HOTS.

### 3. Result and Discussions

#### 3.1 Results of the Planning Stage

Planning this research and development involves examining student characteristics and needs, identifying learning activities, and establishing the scope of VAR products. Student characteristics, instructor needs, and present learning activities were assessed through observation and interviews. At *SMP Kalam Kudus Surakarta*, *SMP Tarakanita Sukoharjo*, and *SMP Negeri 3 Surakarta*, teachers mostly used textbooks and presented content via LCD projectors with minimally interactive PowerPoint presentations of text and a few examples. Thus, students seemed bored and indifferent to teacher questions, creating a teacher-centered learning environment.

The teacher provides explanations and resources, followed by group assignments during the core learning activities. Teacher-prepared tasks prompt group discussions and presentations. Due to the lack of systematic group activity preparation, students talk a lot and bother their peers, making talk unproductive. In addition, some groups assigned tasks to one member. Administrative difficulties that decrease time for creativity and lack of knowledge about the newest technology breakthroughs that can be used as learning media have prevented teachers from developing effective interactive media for learning. Non-interactive media such as text and images bores and demotivates the students. Therefore, students are more engaged and science lessons should deliver knowledge and interesting experiences.

Additionally, semi-structured interviews with junior high natural science teachers yielded additional data. About three topics teachers were interviewed to learn about learning, instructional media, and classroom teaching practices. The data from oral interviews with junior high science subject teachers will be used as a benchmark for student and teacher needs for innovation in VAR instructional media for science learning. The interview results from *SMP Kalam Kudus*, *SMP Tarakanita Sukoharjo*, and *SMP Negeri 3 Surakarta* science instructors are as follows: Teachers reported that student participation in science learning has decreased in recent years. These teachers taught science through lectures and assignments. As most pupils are passive learners, these strategies are poor at engaging with them. Some teachers employ debate and problem-based learning but struggle to create an engaged and compelling learning environment.

According to the teachers, media such as textbooks, PowerPoint presentations, and the Internet are commonly used in ICT-based learning media. However, the use of ICT-based learning media is still limited, and has not been fully utilized to create more interactive learning experiences. Teachers recognize that ICT-based learning media have the potential to help students better understand natural sciences (i.e., *IPA*) materials, but limitations in time and knowledge of current technology pose obstacles to its implementation. Teachers also agree that VAR-based learning media have great potential for application in science learning. Interviews and observations at *SMP Kalam Kudus Surakarta*, *SMP Tarakanita Sukoharjo*, and *SMP Negeri 3 Surakarta* highlight the urgent need for VAR-based learning media that students can independently use. Current methods, which are dominated by lectures and assignments, have decreased student involvement and participation. The existing ICT-based media are limited and less interactive. Teachers concur that enhancing students' critical thinking skills requires interactive media. Implementing VAR in science learning can foster a more engaging and efficient environment that optimally stimulates students' critical thinking skills.

Considering the great potential of VAR, which has not been widely implemented in the context of science learning, the development of this medium is expected to replicate real-life situations and present materials dynamically through videos and animations. This research is crucial for addressing these needs and directing the development of innovative learning media in line with the latest technological advancements.

#### 3.2 Results of the Design Stage

The design phase involved preparing the content and program outline for a VAR product. This stage includes a comprehensive analysis of the previous results to minimize errors during development and testing. The material for the augmented reality media is sourced from the national curriculum guidebook for grade VIII, semester II, including the lesson plan (i.e., *RPP*) and curriculum, which encompasses Core Competencies and Basic Competencies. These competencies, derived from the analysis of augmented reality media needs, focus on knowledge, skills, and attitudes with an emphasis on knowledge. The contents cover the human respiratory and excretory systems, as shown in table 5.

Table 5: Content Competencies for VAR Material in Science Learning

| Basic Competencies |                                                                                                                                                                   | Indicator |                                                                |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|----------------------------------------------------------------|
| 3.9.               | Analyzing the human respiratory system and understanding disorders of the respiratory system, as well as efforts to maintain the health of the respiratory system | 3.9.1     | Identify respiratory organs                                    |
|                    |                                                                                                                                                                   | 3.9.2     | Understand the mechanics of breathing                          |
|                    |                                                                                                                                                                   | 3.9.3     | Understand various disorders of the respiratory system         |
| 3.10.              | Analyzing the excretory system in humans and understanding disorders of the excretory system and efforts to maintain the health of the excretory system           | 3.10.1    | Analyze the structure and function of the excretory system     |
|                    |                                                                                                                                                                   | 3.10.2    | Analyze disorders of the excretory system                      |
|                    |                                                                                                                                                                   | 3.10.3    | Analyze efforts to maintain the health of the excretory system |

### 3.3 Development Stage Results

The development stage is the implementation of the previous stages, and some of the activities carried out during this stage are developing the initial product that is truly complete in terms of its components. Subsequently, validation is carried out by experts, and if there is anything that needs to be improved, it is immediately fixed. There was then a validation stage by the subject teachers and students. The results of the development stage are as follows.

The next step involves an initial evaluation of the product developed by experts. Two media experts assessed Visual Appearance, Interactivity, Design Consistency, Accessibility, and Technological Compatibility. Two subject matter experts evaluated Content Quality and Learning Effectiveness. Additionally, three practitioners, specifically science teachers, will assess the product's alignment with the teaching practices and practical applications. Overall, evaluations by experts and teachers during the alpha test deem the product 'eligible' for use as learning media and for proceeding to the next stage. The feasibility evaluation results are listed in Table 6.

Table 6: Alpha Test results

| No                       | Expert and Teacher Validator | Average Validation Results (%) | Qualification | Decision      |
|--------------------------|------------------------------|--------------------------------|---------------|---------------|
| 1.                       | Media I                      | 86%                            | Very good     | Very Eligible |
| 2.                       | Media II                     | 91%                            | Very good     |               |
| 3.                       | Material I                   | 81%                            | Very good     | Very Eligible |
| 4.                       | Material II                  | 89%                            | Very Good     |               |
| 5.                       | Teacher I                    | 75%                            | Good          | Very Eligible |
| 6.                       | Teacher II                   | 83%                            | Very good     |               |
| 7.                       | Teacher III                  | 92%                            | Very good     |               |
| Average Total Alpha Test |                              | 85%                            | Very good     | Very Eligible |

The Alpha Test results in Table 4 show that expert and teacher validation confirmed that the developed media and materials were excellent and highly suitable for use. Media I and II scored 86% and 91%, respectively, and both were deemed excellent. Materials I and II received 81% and 89%, respectively, and were rated as excellent. Teacher assessment revealed Teacher I at 75% (good), Teacher II at 83% (very good), and Teacher III at 92% (very good). The overall average from the Alpha Test was 85%, qualifying as very good, indicating that the media was very eligible. Validators provided suggestions for improvement: 1) enhance the home display with more attractive buttons and colors, 2) add navigation buttons, and 3) include diverse learning sources beyond textbooks or worksheets.

The next step involved usability testing to evaluate the functionality of the application from the perspective of high school students. The usability testing results inform the final product distribution and assess the impact of the VAR product. Conducted on 24 high school students, the usability testing revealed an average student response score of 88%, qualifying as "Very Good." The detailed results are presented in Table 7.

Table 7: Beta Test Results

| No                      | Test Type                       | Average Validation Results (%) | Qualification |
|-------------------------|---------------------------------|--------------------------------|---------------|
| 1.                      | Preliminary Trial (5 Students)  | 82%                            | Very good     |
| 2.                      | Play field trial (12 students)  | 88%                            | Very good     |
| 3.                      | Operational trial (24 students) | 95%                            | Very good     |
| Average Total Beta Test |                                 | 88%                            | Very good     |

The beta test results indicate that the VAR learning media product has been tested at several stages with very positive outcomes. A preliminary test involving five students yielded an average validation of 82% with a "very good" qualification. The main field test involving 12 students obtained an average of 88%, also with a 'very good' qualification. Furthermore, the operational test involving 24 students showed very satisfactory results, with an average of 95% remaining in the "very good" qualification. Overall, the total average of the beta test reached 88% with a 'very good' qualification.

### 3.4 Testing Phase Results

During several classroom learning meetings, the impact or influence of the product on improving students' high-order thinking skills (HOTS) in junior high school was measured. The testing of Students' HOTS skills were done using pretest and posttest methods with a control group. The difference in the average post-test scores was then analyzed using an Independent-sample T-test to determine the significance of the difference between the two classes. The results of the t-tests are presented in Table 8.

Table 8: T-test results

| Group              | Number of Students | Average | Standard Deviation | t-statistic | p-value |
|--------------------|--------------------|---------|--------------------|-------------|---------|
| Experimental Class | 121                | 87.714  | 9.948              | 14.283      | 0.000   |
| Control Class      | 121                | 70.52   | 10.26              | 3.12        | 0.0019  |

With a p-value of 0.001985, which is smaller than 0.05, we rejected the null hypothesis ( $H_0$ ). This means that there was a significant influence on increasing students' HOTS skills between the experimental and control classes.

After conducting a t-test, it was found that the VAR media product significantly improved students' Higher Order Thinking Skills (HOTS). To assess the effectiveness of the product, an N-Gain score analysis was performed by comparing pretest and post-test scores from both the experimental and control groups. Table 9 shows that the experimental group achieved an average N-Gain score of 0.714 (high range), whereas the control group achieved an average N-Gain score of 0.088 (low range). These results suggest that the developed product is more effective than traditional classroom media in enhancing students' HOTS. The Detailed N-Gain test results are presented in Table 9.

Table 9: N-Gain of Students' HOTS Skills

| Details | N-Gain Experimental Class | N-Gain Control Class |
|---------|---------------------------|----------------------|
| count   | 121.000                   | 121.000              |
| mean    | 0.714                     | 0.088                |
| std     | 0.288                     | 1.241                |
| min     | -0.713                    | -12.885              |
| 25%     | 0.557                     | -0.008               |
| 50%     | 0.740                     | 0.209                |
| 75%     | 0.891                     | 0.461                |
| max     | 1.275                     | 1.013                |

Implementing VAR media in experimental science classes significantly boosts students' higher-order thinking skills (HOTS), as shown by an average N-Gain score of 0.714, indicating a high improvement. In contrast, the

control class without VAR media scored an average N-Gain of 0.088, which was categorized as low. This demonstrates that VAR media is highly effective in enhancing HOTS in students.

#### **4. Discussions**

Research shows that using VAR technology in junior high science classes improves higher-order thinking skills. The t-test and N-gain analysis demonstrated that the experimental class using VAR media improved more than the control class using conventional media. VAR-based technology-supported learning activities boost students' grasp of complicated concepts and develop HOTS-related abilities, such as analysis and evaluation. Therefore, VAR technology is effective when integrated into a planned teaching method relevant to the content being taught. The use of technology in learning activities must be adjusted and aligned with the strategy or learning model used during the learning process as well as its sophistication.

The research findings showed a significant difference in HOTS skills between the experimental and control classes. The results confirm this, with a higher average post-test score in the experimental group (87.714) than in the control group (70.52), with a t-statistic of 14.283 and a p-value of 0.000, indicating a statistically significant effect of VAR-based learning. Observations and motivation questionnaires further revealed greater student engagement, active interaction with VAR content, and improved HOTS skills in the experimental group. Students who learned using VAR media were more enthusiastic and actively engaged in learning activities (Ansari, KG and Baby, 2023). They were more likely to ask questions, participate in discussions, and show a high level of interest in the material being studied, indicating that VAR technology not only enhances cognitive skills but also the affective aspects of learning (Rudnik, 2023; Almaguer et al., 2023).

The findings of this research cannot be separated from those of comprehensive preliminary studies. As is well known, the analysis of needs in the development of digital learning products plays a crucial role in ensuring that the resulting product is effective and efficient in supporting the learning process (Ambarsari et al., 2021). Through a needs analysis, developers deeply understand what end users, in this case, learners and educators, need to meet the intended instructional goals (Rejekiningsih et al., 2023). Budiarto, Rejekiningsih, and Sudiyanto (2021) highlighted the personal, skill, organizational, and facility requirements for the creation of a digital learning aid needed during instruction. Moreover, needs analysis also prevents developed products from being erroneous and discrepant (Rosalina and Suhardi, 2020). Hence, fully knowing the user requirements together with the capability of digital media allows developers to accomplish effective learning products that satisfy the requirements of contemporary technological and educational development.

These findings are relevant to the field of education, specifically junior high school science education. VAR technology can be an effective option for boost students' interest and involvement in lessons (Saravanan et al., 2022). Furthermore, it helps students acquire the ability to think critically and creatively, which is very important for the 21st century (Nicolaou et al., 2019; Syawaludin et al., 2019). With VAR technology, learning becomes more interactive and enjoyable, which can improve students' overall learning outcome.

The current results corroborate earlier works that argue in favor of the use of VR and AR technologies for educational purposes. For instance, the use of mobile AR apps for project-based learning has been proven to have far-reaching benefits for the innovation, creativity, problem-solving skills, and information literacy of students, especially in chemistry education (Chen, Huang and Chou, 2017). The application of Virtual Reality (VR) and Augmented Reality (AR) in the teaching of Natural Science (IPA) has remarkably increased students' 21st-century competencies. Studies show that the use of VR and AR technology has improved students' academic performance and satisfaction in several subjects and areas, particularly in science education. Moreover, integrating VR and AR into the educational process may contribute to forming students' digital skills, literacy, and other skills of the twenty-first century that are needed today (Hughes and Maas, 2019; Hughes and Maas, 2018).

Moreover, using VR and AR in education can improve critical thinking, creativity, communication, and collaboration (4C skills), which are fundamental to tackling issues in the modern world. With the help of VR and AR technologies, educators can develop interesting and integrated learning environments where students can acquire skills that meet the needs of time (Lee et al., 2022; Monterubbianesi et al., 2022). Thus, the use of AR and VR technologies in teaching science not only aids learners in comprehending the subject matter but also equips them with the skills they need to thrive in the future (Rudnik, 2023; Dilmen and Atalay, 2021).

In the context of 21st century education, the use of VAR technology can also help students prepare for deep and interactive learning experiences that develop adaptability, problem-solving, and technology skills that are highly needed in today's digital era (Safri and Sheikh, 2022; Rejekiningsih et al., 2023). Therefore, the implementation

of VAR in IPA learning not only enhances the quality of education, but also helps create a generation that is ready to face future challenges with relevant and up-to-date skills (Singh, 2024; Fu, 2021).

Although the benefits of VAR technology are obvious, its actual implementation poses a challenge. Software and equipment costs can be crippled by budget-constrained institutions. Furthermore, effective application of the technology necessitates specialized technical assistance, as well as thorough educator training. The lack of adequate infrastructure, especially in rural regions where the Internet and technological devices are scarce, adds to the challenges of implementation. These issues can be addressed through collaborative participation in technology infrastructure improvements and teacher training. The results of this study are intended to assist educational media designers, educators, and teachers adopting VAR technology, while emphasizing the importance of infrastructure aid and teacher professional development concerning technology use in education.

## 5. Conclusion and Suggestions

In general, the findings of this study show that VAR technology has great prospects for improving higher-order thinking skills (HOTS) of junior high school students in science education. While implementation concerns need to be addressed, VAR can be incorporated to foster active and meaningful learning. This study aims to develop innovative educational media and contribute to the growing concern for improving education through technology integration pedagogy. Subsequent studies should examine the use of VAR in fields other than science education to include mathematics, social sciences, and language learning. In addition, further research on long-term learning retention, adaptability of students, and readiness of teachers in VAR environments is needed. Varied approaches will assist educational practitioners, and VAR could benefit educational levels and contexts. Moreover, educational policymakers and institutions should concentrate on teacher training as well as the provision of infrastructure and tools that will permit the injection of VAR into different educational contexts.

## References

- Al-Amri, A., Osman, M. and Musawi, A. Al, 2020. The effectiveness of a 3D-virtual reality learning environment (3D-VRLE) on the omani eighth grade students' achievement and motivation towards physics learning. *International Journal of Emerging Technologies in Learning*, 15(5), pp.4–16. <https://doi.org/10.3991/IJET.V15I05.11890>.
- Alessi, S.M. and Trollip, S.R., 2001. *Multimedia for learning: Methods and development*. Allyn & Bacon.
- Almaguer, C.A.G., Acosta, A.C.A., Lopez, A.A., Jimenez, O.R.R., Cardoso, C.C. and Ramirez, C.Z., 2023. Augmented, Virtual and Immersive Reality as Learning Support in TEC21 Educational Model. In: *2023 Future of Educational Innovation-Workshop Series Data in Action*. [online] IEEE. pp.1–5. <https://doi.org/10.1109/IEEECONF56852.2023.10104699>.
- Ambarsari, R., Sartono, E.K.E., Mustadi, A., Zubaidah, E., Jhon, W. and Rafsanjani, M.A., 2021. Needs Analysis for the Development of Electronic Story Calender Media to Improve Spirit of Nationalism. *Journal of Education Research and Evaluation*, [online] 5(3), pp.398–405. Available at: <https://ejournal.undiksha.ac.id/index.php/JERE/article/view/33281>.
- Ansari, A.K., KG, S.S. and Baby, B.C., 2023. Virtual Reality and Augmented Reality in Education. *International Journal for Research in Applied Science and Engineering Technology*, [online] 11(3), pp.2014–2018. <https://doi.org/10.22214/ijraset.2023.49825>.
- Arifin, S., 2020. The Role of Critical Reading to Promote Students' Critical Thinking and Reading Comprehension. *Jurnal Pendidikan dan Pengajaran*, 53(3), p.318. <https://doi.org/10.23887/jpp.v53i3.29210>.
- Astrid, A. and Hasanah, A., 2022. Integrating Higher Order Thinking Skills (HOTS) Into English Language Teaching for Elementary School Students: Teachers' Perspectives and Challenges. *3L: Language, Linguistics, Literature*, 28(3), pp.217–230. <https://doi.org/10.17576/3L-2022-2803-14>.
- Averbeck, F., Leifeling, S., Müller, K. and Schoenfelder, T., 2024. Virtual Reality in Social Work Teaching - Two Approaches to 360° Videos and Collaborative Working. *Electronic Journal of e-Learning*, 22(3), pp.111–123. <https://doi.org/10.34190/ejel.21.6.3225>.
- Bima, M., Saputro, H. and Efendy, A., 2021. Virtual Laboratory to Support a Practical Learning of Micro Power Generation in Indonesian Vocational High Schools. *Open Engineering*, 11(1), pp.508–518. <https://doi.org/10.1515/eng-2021-0048>.
- Bismala, L. and Manurung, Y.H., 2021. Student satisfaction in e-learning along the covid-19 pandemic with importance performance analysis. *International Journal of Evaluation and Research in Education*, 10(3), pp.753–759. <https://doi.org/10.11591/ijere.v10i3.21467>.
- Budiarto, M.K., Rejekiingsih, T. and Sudiyanto, S., 2021. Students' opinions on the need for interactive multimedia development for entrepreneurship learning. *International Journal of Evaluation and Research in Education (IJERE)*, [online] 10(4), p.1290. <https://doi.org/10.11591/ijere.v10i4.21411>.
- Budiyono, 2017. *Pengantar Metodologi Penelitian Pendidikan*. UNS Press.
- Bukhori, B., Said, H., Wijaya, T. and Nor, F.M., 2019. The effect of smartphone addiction, achievement motivation, and textbook reading intensity on students' academic achievement. *International Journal of Interactive Mobile Technologies*, 13(9), pp.66–80. <https://doi.org/10.3991/ijim.v13i09.9566>.

- Chen, C.H., Huang, C.Y. and Chou, Y.Y., 2017. Effects of augmented reality-based multidimensional concept maps on students' learning achievement, motivation and acceptance. *Universal Access in the Information Society*, 18(2), pp.257–268. <https://doi.org/10.1007/s10209-017-0595-z>.
- Dasilva, B.E., Ardiyati, T.K., Suparno, Sukardiyono, Eveline, E., Utami, T. and Ferty, Z.N., 2019. Development of Android-based Interactive Physics Mobile Learning Media (IPMLM) with scaffolding learning approach to improve HOTS of high school students. *Journal for the Education of Gifted Young Scientists*, 7(3). <https://doi.org/10.17478/jegys.610377>.
- Dilmen, I. and Atalay, N., 2021. The Effect of the Augmented Reality Applications in Science Class on Students' 21st Century Skills and Basic Skills. *Journal of Science Learning*, 4(4), pp.337–346. <https://doi.org/10.17509/jsl.v4i4.32900>.
- Dinayusadewi, N.P. and Agustika, G.N.S., 2020. Development Of Augmented Reality Application As A Mathematics Learning Media In Elementary School Geometry Materials. *Journal of Education Technology*, 4(2), p.204. <https://doi.org/10.23887/jet.v4i2.25372>.
- Elmqaddem, N., 2019. Augmented Reality and Virtual Reality in Education. Myth or Reality? *International Journal of Emerging Technologies in Learning (IJET)*, 14(03), p.234. <https://doi.org/10.3991/ijet.v14i03.9289>.
- Fu, L., 2021. Research on the Teaching Model of Animation Professional Class Based on AR/VR Technology and 5G Network. *Wireless Communications and Mobile Computing*, 2021, pp.1–10. <https://doi.org/10.1155/2021/1715909>.
- Fuad, N.M., Zubaidah, S., Mahanal, S. and Suarsini, E., 2017. Improving junior high schools' critical thinking skills based on test three different models of learning. *International Journal of Instruction*, 10(1), pp.101–116. <https://doi.org/10.12973/iji.2017.1017a>.
- Gall, M.D., Gall, J.P. and Borg, W.R., 2003. *Educational Research\_ An Introduction (7th Edition)*.
- Garad, A., Al-Ansi, A.M. and Qamari, I.N., 2021. The role of e-learning infrastructure and cognitive competence in distance learning effectiveness during the covid-19 pandemic. *Cakrawala Pendidikan*, 40(1), pp.81–91. <https://doi.org/10.21831/cp.v40i1.33474>.
- Huang, K.-T., Ball, C., Francis, J., Ratan, R., Boumis, J.K. and Fordham, J., 2019. Augmented Versus Virtual Reality in Education: An Exploratory Study Examining Science Knowledge Retention When Using Augmented Reality/Virtual Reality Mobile Applications. *Cyberpsychology Behavior and Social Networking*, 22(2), pp.105–110. <https://doi.org/10.1089/cyber.2018.0150>.
- Hughes, J. and Maas, M., 2018. Developing 21st Century Competencies of Marginalized Students Through the Use of Augmented Reality (AR). *Learning Landscapes*, 11(1), pp.153–169. <https://doi.org/10.36510/learnland.v11i1.929>.
- Kamińska, D., Sapiński, T., Wiak, S., Tikk, T., Haamer, R.E., Avots, E., Helmi, A., Ozcinar, C. and Anbarjafari, G., 2019. Virtual reality and its applications in education: Survey. *Information (Switzerland)*, 10(10). <https://doi.org/10.3390/info10100318>.
- Lee, T., Wen, Y., Chan, M.Y., Azam, A.B., Looi, C., Taib, S.F.B.M., Ooi, C.H., Huang, L., Xie, Y. and Cai, Y., 2022. Investigation of virtual & augmented reality classroom learning environments in university STEM education. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2022.2155838>.
- Mariscal, G., Jiménez, E., Vivas-Urias, M.D., Redondo-Duarte, S. and Moreno-Pérez, S., 2020. Virtual reality simulation-based learning. *Education in the Knowledge Society*, 21. <https://doi.org/10.14201/eks.20809>.
- Maryani, I., Prasetyo, Z.K., Wilujeng, I., Purwanti, S. and Fitriawanati, M., 2021. HOTS Multiple Choice and Essay Questions: A Validated Instrument to Measure Higher-order Thinking Skills of Prospective Teachers. *Journal of Turkish Science Education*, 18(4), pp.674–690. <https://doi.org/10.36681/tused.2021.97>.
- Monterubbianesi, R., Monterubbianesi, R., Tosco, V., Tosco, V., Vitiello, F., Vitiello, F., Orilisi, G., Orilisi, G., Fraccastoro, F., Fraccastoro, F., Putignano, A., Putignano, A., Orsini, G. and Orsini, G., 2022. Augmented, Virtual and Mixed Reality in Dentistry: A Narrative Review on the Existing Platforms and Future Challenges. *Applied Sciences*, 12(2), p.877. <https://doi.org/10.3390/app12020877>.
- Muhayimana, T., Kwizera, L. and Nyirahabimana, M.R., 2022. Using Bloom's taxonomy to evaluate the cognitive levels of Primary Leaving English Exam questions in Rwandan schools. *Curriculum Perspectives*, 42(1), pp.51–63. <https://doi.org/10.1007/s41297-021-00156-2>.
- Munje, P.N. and Jita, T., 2020. The impact of the lack of ICT resources on teaching and learning in selected South African primary schools. *International Journal of Learning, Teaching and Educational Research*, 19(7), pp.263–279. <https://doi.org/10.26803/IJLTER.19.7.15>.
- Muñoz-Saavedra, L., Miró-Amarante, L. and Domínguez-Morales, M., 2020. Augmented and virtual reality evolution and future tendency. *Applied Sciences (Switzerland)*, 10(1). <https://doi.org/10.3390/app10010322>.
- Nicolaou, C., Matsiola, M. and Kalliris, G., 2019. Technology-Enhanced Learning and Teaching Methodologies through Audiovisual Media. *Education Sciences*, [online] 9(3), p.196. <https://doi.org/10.3390/educsci9030196>.
- Oleksiuk, V.P. and Oleksiuk, O.R., 2020. Exploring the potential of augmented reality for teaching school computer science. In: *CEUR Workshop Proceedings*. pp.91–107.
- Parmiti, D.P., Antara, I.G.W.S. and Wulandari, I.G.A.A.M., 2022. The Effectiveness of E-Scrapbook Media Containing HOTS Questions on Science Learning Outcomes of Elementary School Students. *Journal of Education Research and Evaluation*. <https://doi.org/10.23887/jere.v6i3.52078>.
- Patten, M.L. and Newhart, M., 2018. Descriptive and Inferential Statistics. In: *Understanding Research Methods*. <https://doi.org/10.4324/9781315213033-66>.
- Pedaste, M., Mitt, G. and Jürivete, T., 2020. What is the effect of using mobile augmented reality in K12 inquiry-based learning? *Education Sciences*, <https://doi.org/10.3390/educsci10040094>.

- Permana, T.I., Hindun, I., Rofi'ah, N.L. and Nur Azizah, A.S., 2019. Critical Thinking Skills: The Academic Ability, Mastering Concepts, and Analytical Skill of Undergraduate Students. *Jpbi (Jurnal Pendidikan Biologi Indonesia)*, 5(1), pp.1–8. <https://doi.org/10.22219/jpbi.v5i1.7626>.
- Raja, M. and Priya, G.G.L., 2021. An Analysis of Virtual Reality Usage through a Descriptive Research Analysis on School Students' Experiences: A Study from India. *International Journal of Early Childhood Special Education*, 13(2), pp.990–1005. <https://doi.org/10.9756/INT-JECSE/V13I2.211142>.
- Rakhmetov, M., Sadvakasova, A., Saltanova, G. and Yessekenova, A., 2022. Usage and effectiveness of educational platforms in Kazakhstan during the Covid-19 pandemic. *World Transactions on Engineering and Technology Education*, 20(3), pp.226–231.
- Ramos, J.L.S., Dolipas, B.B. and Villamor, B.B., 2013. Higher Order Thinking Skills and Academic Performance in Physics of College Students : A Regression Analysis. *International Journal of Innovative Interdisciplinary Research*, (4), p.48. <https://doi.org/ISSN 1839-9053>.
- Rejekiingsih, T., Maulana, I., Budiarto, M.K. and Qodr, T.S., 2023. Android-based augmented reality in science learning for junior high schools: Preliminary study. *International Journal of Evaluation and Research in Education*, 12(2). <https://doi.org/10.11591/ijere.v12i2.23886>.
- Rejekiingsih, T., Sudiyanto, S. and Budiarto, M.K., 2022. The Utilization of Computer-Based Interactive Multimedia in Improving Entrepreneurial Attitudes of High School Students. *JPI (Jurnal Pendidikan Indonesia)*, 11(1), pp.1–9. <https://doi.org/10.23887/jpi-undiksha.v11i1.37031>.
- Rofiah, E., Aminah, N.S. and Sunarno, W., 2018. PENGEMBANGAN MODUL PEMBELAJARAN IPA BERBASIS HIGH ORDER THINKING SKILL (HOTS) UNTUK MENINGKATKAN KEMAMPUAN BERPIKIR KRITIS SISWA KELAS VIII SMP/MTs. *INKUIRI: Jurnal Pendidikan IPA*, [online] 7(2), p.285. <https://doi.org/10.20961/inkuiri.v7i2.22992>.
- Rosalina, S.S. and Suhardi, A., 2020. NEED ANALYSIS OF INTERACTIVE MULTIMEDIA DEVELOPMENT WITH CONTEXTUAL APPROACH ON POLLUTION MATERIAL. *INSECTA: Integrative Science Education and Teaching Activity Journal*, 1(1), p.93. <https://doi.org/10.21154/insecta.v1i1.2107>.
- Rovithis, E., Floros, A., Moustakas, N., Vogklis, K. and Kotsira, L., 2019. Bridging audio and augmented reality towards a new generation of serious audio-only games. *Electronic Journal of e-Learning*, 17(2), pp.144–156. <https://doi.org/10.34190/JEL.17.2.07>.
- Rudnik, Y., 2023. The Use of Augmented Reality and Virtual Reality Technologies in Teaching Foreign Languages. *Educological Discourse*, 43(1), pp.165–183. <https://doi.org/10.28925/2312-5829.2023.110>.
- Rukayah, Daryanto, J., Atmojo, I.R.W., Ardiansyah, R., Saputri, D.Y. and Salimi, M., 2022. Augmented Reality Media Development in STEAM Learning in Elementary Schools. *Ingenierie des Systemes d'Information*, 27(3), pp.463–471. <https://doi.org/10.18280/isi.270313>.
- Safri, N.M. and Sheikh, U.U., 2022. Issues and challenges of technology-enhanced learning during the Covid-19 era: a case study. *World Transactions on Engineering and Technology Education*, 20(2), pp.89–94.
- Saputri, D.Y., Rukayah, R. and Indriayu, M., 2018. Need Assessment of Interactive Multimedia Based on Game in Elementary School: A Challenge into Learning in 21st Century. *International Journal of Educational Research Review*, 3(3), pp.1–8. <https://doi.org/10.24331/ijere.411329>.
- Saravanan, S., Saravanan, D.S.K., Parthasarathy, S. and Parthasarathy, D.S., 2022. Meaningful Learning through Virtual Reality and Augmented Reality Technological Environments. *International Journal for Research in Applied Science and Engineering Technology*. <https://doi.org/10.22214/ijraset.2022.44015>.
- Scavarelli, A., Arya, A. and Teather, R.J., 2021. Virtual reality and augmented reality in social learning spaces: a literature review. *Virtual Reality*, 25(1), pp.257–277. <https://doi.org/10.1007/s10055-020-00444-8>.
- Septinaningrum, Hakam, K.A., Setiawan, W. and Agustin, M., 2022. Developing of Augmented Reality Media Containing Grebeg Pancasila for Character Learning in Elementary School. *Ingenierie des Systemes d'Information*, 27(2), pp.243–253. <https://doi.org/10.18280/isi.270208>.
- Serin, H., 2020. Virtual Reality in Education from the Perspective of Teachers. *Revista Amazonia Investiga*, 9(26), pp.291–303. <https://doi.org/10.34069/ai/2020.26.02.33>.
- Setiawan, A., Pusporini, W. and Dardjito, H., 2020. Observation instrument for student social attitude in primary schools: Validity and reliability. *Jurnal Penelitian dan Evaluasi Pendidikan*, 24(1). <https://doi.org/10.21831/pep.v24i1.31868>.
- Shepa, M.J., Serevina, V. and Astra, I.M., 2021. Development of virtual reality-based learning media on electromagnetic wave radiation material. In: *Journal of Physics: Conference Series*. <https://doi.org/10.1088/1742-6596/1876/1/012088>.
- Singh, K.D., 2024. QoS-enhanced Load Balancing Strategies for Metaverse-infused VR/AR in Engineering Education 5.0. *Computer Applications in Engineering Education*, 32(3). <https://doi.org/10.1002/cae.22722>.
- Strand, I., 2020. Virtual reality in design processes-a literature review of benefits, challenges, and potentials. *FormAkademisk*, 13(6), pp.1–19. <https://doi.org/10.7577/FORMAKADEMISK.3874>.
- Supianti, I.I., Yaniawati, P., Osman, S.Z.M., Al-Tamar, J. and Lestari, N., 2022. DEVELOPMENT OF TEACHING MATERIALS FOR E-LEARNING-BASED STATISTICS MATERIALS ORIENTED TOWARDS THE MATHEMATICAL LITERACY ABILITY OF VOCATIONAL HIGH SCHOOL STUDENTS. *Infinity Journal*, 11(2), pp.237–254. <https://doi.org/10.22460/infinity.v11i2.p237-254>.
- Surbhi, S., 2019. Difference between Descriptive and Inferential Statistics. *Regression Analysis: An Intuitive Guide*.

- Syawaludin, A., Gunarhadi, G. and Rintayati, P., 2019. Development of Augmented Reality-Based Interactive Multimedia to Improve Critical Thinking Skills in Science Learning. *International Journal of Instruction*, [online] 12(4), pp.331–344. <https://doi.org/10.29333/iji.2019.12421a>.
- Tanjung, H.S., Nababan, S.A. and Sa'dijah, C., 2020. Development of assessment tools of critical thinking in mathematics in the context of hots. *Advances in Mathematics: Scientific Journal*, 9(10), pp.8659–8667. <https://doi.org/10.37418/amsj.9.10.91>.
- Tong, L.C., Tong, L.C., Rosli, M.S., Rosli, M.S., Saleh, N.S. and Saleh, N.S., 2022. Enhancing HOTS using Problem-Based Learning and Digital Game in the Context of Malaysian Primary School. *International journal of interactive mobile technologies*. <https://doi.org/10.3991/ijim.v16i02.27677>.
- Wang, Q., Zhang, Q., Sun, W., Boulay, C., Kim, K. and Barmaki, R.L., 2023. A scoping review of the use of lab streaming layer framework in virtual and augmented reality research. *Virtual Reality*, [online] 27(3), pp.2195–2210. <https://doi.org/10.1007/s10055-023-00799-8>.
- Wehrmann, F. and Zender, R., 2024. Inclusive Virtual Reality Learning: Review and 'Best-Fit' Framework for Universal Learning. *Electronic Journal of e-Learning*, 22(3), pp.74–89. <https://doi.org/10.34190/ejel.21.6.3265>.
- Wibowo, F.C., Nasbey, H., Sanjaya, L.A., Darman, D.R., Ahmad, N.J. and Ismail, H.N., 2021. The technology of interactive book augmented reality (IBAR) for facilitating student 21-century skills. *Journal of Theoretical and Applied Information Technology*, 99(22), pp.5276–5286.
- Widiyawati, Y., Widiyawati, Y., Nurwahidah, I., Nurwahidah, I., Sari, D.S., Sari, D.S., Masykuri, M., Masykuri, M., Masykuri, M., Budiyanto, C.W., Budiyanto, C.W. and Budiyanto, C.W., 2021. The 21 st century science learning: HOTS and digital literacy among junior high school students in Semarang, Indonesia. *Journal of Physics: Conference Series*. <https://doi.org/10.1088/1742-6596/1842/1/012081>.
- Widyaningsih, S.W., Yusuf, I., Prasetyo, Z.K. and Istiyono, E., 2020. Online Interactive Multimedia Oriented to HOTS through E-Learning on Physics Material about Electrical Circuit. *JPI (Jurnal Pendidikan Indonesia)*, [online] 9(1), pp.1–14. <https://doi.org/10.23887/jpi-undiksha.v9i1.17667>.
- Winarni, E.W. and Purwandari, E.P., 2019. The effectiveness of turtle mobile learning application for scientific literacy in elementary school. *Journal of Education and e-Learning Research*, 6(4). <https://doi.org/10.20448/journal.509.2019.64.156.161>.
- Yang, B., 2023. Virtual Reality and Augmented Reality for Immersive Learning: A Framework of Education Environment Design. *International Journal of Emerging Technologies in Learning (IJET)*. <https://doi.org/10.3991/ijet.v18i20.44209>.
- Yaniawati, P., Supianti, I.I., Fisher, D. and Sa'adah, N., 2021. Development and effectiveness of mobile learning teaching materials to increase students' creative thinking skills. *Journal of Physics: Conference Series*, 1918(4), p.2234. <https://doi.org/10.1088/1742-6596/1918/4/042081>.
- Yanto, D.T.P., Kabatiah, M., Zaswita, H., Jalinus, N. and Refdinal, R., 2022. Virtual Laboratory as A New Educational Trend Post Covid-19: An Effectiveness Study. *Mimbar Ilmu*, 27(3). <https://doi.org/10.23887/mi.v27i3.53996>.
- Yao, Y., Wang, P., Jiang, Y.J., Li, Q. and Li, Y., 2022. Innovative online learning strategies for the successful construction of student self-awareness during the COVID-19 pandemic: Merging TAM with TPB. *Journal of Innovation and Knowledge*, 7(4). <https://doi.org/10.1016/j.jik.2022.100252>.
- Yusof, Y.B.M., 2019. 21 st Century Learning is Not Merely ICT. *International Research Journal of Education and Sciences*, [online] 3(1), pp.18–23. Available at: <<https://www.masree.info/wp-content/uploads/2019/11/IRJES-VOL-3-ISSUE-1-ARTICLE-5.pdf>>.
- Zainil, M., Kenedi, A.K., Indrawati, T. and Handrianto, C., 2023. The influence of a STEM-based digital classroom learning model and high-order thinking skills on the 21st-century skills of elementary school students in Indonesia. *Journal of Education and e-Learning Research*, 10(1), pp.29–35. <https://doi.org/10.20448/jeelr.v10i1.4336>.

# Quality of e-Learning in Nepalese Universities During the COVID-19 Pandemic

Ratna Mani Nepal<sup>1</sup>, Shyam Guragain<sup>2</sup>, Jiwnath Ghimire<sup>3</sup> and Bimal Khadka<sup>4</sup>

<sup>1</sup>Center for Nepal and Asian Studies (CNAS), Tribhuvan University, Kathmandu, Nepal

<sup>2</sup>Tri-Chandra Multiple Campus, Tribhuvan University, Kathmandu, Nepal

<sup>3</sup>Department of Community and Regional Planning, Iowa State University, USA

<sup>4</sup>Department of Rural Development, Patan Multiple Campus, Tribhuvan University, Kathmandu, Nepal

[ratna.nepal@cnas.tu.edu.np](mailto:ratna.nepal@cnas.tu.edu.np)

[shyam.guragain@trc.tu.edu.np](mailto:shyam.guragain@trc.tu.edu.np) (Corresponding Author)

[jghimire@iastate.edu](mailto:jghimire@iastate.edu)

[khadka.bimal@gmail.com](mailto:khadka.bimal@gmail.com)

<https://doi.org/10.34190/ejel.23.1.3955>

An open access article under [CC Attribution 4.0](#)

**Abstract:** Covid 19 pandemic triggered changes in various socio-economic sectors that were already in pace. Among them, the digitalization of higher education faced one of the profound shifts due to the pandemic. The application of digital education is gaining momentum and is likely to continue until the long future after Covid 19. However, there are limited assessments of its quality, especially in the countries in the global south. The purpose of this study was to assess the quality of online learning in Nepalese universities during the pandemic using a customized SERVQUAL model with six dimensions, namely, tangibility, assurance, responsiveness, empathy, and online-class. The online-class dimension is added in this study, which includes conditions of e-learning such as electricity supply, internet service, electronic devices, time schedule, thus characterizing the conditions in a developing country. Specifically, this study focused students' perceived quality of e-learning in universities in Nepal. A survey was conducted between April to July 2021 among 451 graduates and undergraduates who participated in online classes during the pandemic from six universities to explore their satisfaction with the quality of digital education. The survey included 27 questions related to six dimensions: reliability, responsiveness, assurance, empathy, tangibility, and online-class. The data was analyzed quantitatively using the open-source platform Jamovi (version 2.3.28). The results showed that students expressed satisfaction with the quality of online education of e-learning during the pandemic. However, the study found variations in student perceptions based on demographic factors, highlighting differences in satisfaction levels across diverse student groups. This result implies that despite its increasing application, e-learning needs to accommodate a more inclusive approach to improve its quality and satisfy learners from all socio-economic backgrounds. A blended mode of teaching could be more effective in the changing context. This should be paralleled with instructors as well as support staff training. At the same time, the limitations of online-class can be addressed in collaboration with government agencies and universities which further enhances the quality of e-learning.

**Keywords:** e-Learning, Digital education quality, Service quality assessment, COVID-19, Nepal, Higher education

## 1. Introduction

The education sector has been digitalized, promoting online teaching and learning during the COVID-19 pandemic. As the pandemic progressed, digitalization was accompanied by designing new systems and infrastructures, with students' learning outcomes assessed online (Eltahir, Alsalhi, and Al-Qatawneh, 2022). The digitalization is significantly advanced in the higher education sector (Márquez-Ramos, 2021). Although distance higher education is gaining momentum and is likely to continue, there are limited assessments of its service quality. Existing studies have primarily focused on the overall and descriptive assessments of online learning (Al Rawashdeh et al., 2021) or the challenges of inclusivity across different societal groups (Devkota, 2021; Muthuprasad et al., 2021; Nayak and Alam, 2022). Therefore, as much as studies are necessary, to examine its quality and scrutinize its usefulness. The quality of online higher education is a particular concern in developing countries like Nepal, where infrastructures for its proliferation —such as electricity, internet access, and technology—are hindered by issues related to production, access, economic constraints, and human resources (Upadhayaya et al., 2021).

In recent years, higher education has increasingly been considered a service industry (Davies, 2021; Larson, 2009). Higher education institutions (HEIs) are shifting from simply delivering knowledge to offering a range of services to students, such as academic support, career guidance, counseling, physical and mental health support, and financial assistance (Dugenio-Nadela et al., 2023). With the engagement of the private sector in education,

Pavlov and Joy (2018) argue that education has evolved into a market economy, contributing to national revenue, job creation, knowledge production, patent generation, and intellectual property rights. The sector is further flourishing due to increased access and the integration of new technologies in online learning (World Bank, 2024). Thus, studies claim that online education is the future of learning (Chan, Bista and Allen, 2021). With the proliferation of online education during Covid 19 and its aftermath, it is imperative to explore the quality of service in such a thriving industry.

Despite the unprecedented proliferation of digital learning during and after the Covid 19 pandemic, a paucity of studies analyzes the factors determining its quality. This is a less evaluated issue in higher education institutions in developing countries (Zaw and Hlaing, 2024). Available literature barely explored the quality of digital education at the university level. For example, Poudel (2021) applied the Service Quality (SERVQUAL) model to examine the quality of digital education at the college level. Similarly, Adhikari (202) explored non-academic factors of quality in a campus in Western Nepal. Some studies in the pre-pandemic era assessed the quality of distance learning but also at the college level (Baniya, 2016). These studies barely analyze the quality of digital education as perceived by university students on a wider scale. At the same time, studies also highlight issues such as poor digital infrastructure, the digital divide, and its impact on access and outcome of digital learning in developing countries (Al Rawashdeh et al., 2021). However, there is less concern if digital infrastructure, which we conceptualize as 'online learning' in this study, impacts the quality of education and students' satisfaction. This is concerning because, in developing countries, a deep digital divide exists, implying a state of unequal distribution of electricity and internet facilities, poor access to computers or laptops, and a lack of technical know-how (Muthuprasad, Aiswarya, Aditya & Jha, 2021). In this study, we assume that online learning is an additional dimension determining the remaining dimensions of quality digital education, such as responsiveness, assurance, tangible, empathy, and reliability as described by the SERVQUAL model.

The purpose of this study is to measure the quality of digital education in Nepalese universities during the Covid 19 pandemic. It has examined the students' perception and analyzed whether their satisfaction is reciprocal to the quality of digital education in Nepal. The main research questions this study addressed are as follows:

*RQ1: How did the university students experience the quality of digital education in Nepalese universities during the Covid 19 pandemic?*

*RQ2: To what extent do online teaching and students' demographics play a role in forming a perception of digital education quality in Nepal?*

The present study explores students' perceptions of the quality of online education during the Covid 19 pandemic in Nepal. It assumes that students' quality assessments will provide valuable feedback to the key stakeholders of HEIs to improve digital education and guide policy reforms if needed. For this study, e-learning (synonymously used with online learning, digital education, distance learning (ODL), digital learning, and internet-based learning) is defined as a learning environment that uses communication technologies, such as mobile devices or computers, to deliver lectures, learning activities, and students' assessments via the internet.

This article is divided into seven sections. The following section discusses conceptual and theoretical models, along with a review of relevant past studies. This follows with sections on methods and materials, results and findings, discussion, conclusion and limitations, and future research potential.

## **2. Review of Literature**

Four themes of literature are dominant in this field. First, service quality models that focus on the theoretical and practical aspects of the SERVQUAL (service quality) model. Second, the body of literature explores why digital education is essential and what factors determine its quality. Third, literature that applies the SERVQUAL model to examine the quality of e-learning before and after the COVID-19 pandemic is analyzed. Fourth, studies on the quality of e-learning focusing on Nepal's higher education institutions (HEIs) are considered.

### **2.1 Service Quality Models**

Scholars agree that customer satisfaction is crucial for service industries to sustain their business (Parasuraman, Zeithaml, and Berry, 1985; Sumi and Kabir, 2021). Consumer satisfaction, in turn, influences customer loyalty and trust (Uppal, Ali and Gulliver, 2017). However, the satisfaction is heavily dependent on service quality. Various models such as the Nordic model (Ghotbabadi and Baharun, 2012), SERVQUAL (Parasuraman, Zeithaml, and Berry, 1985, 1988), SERVPERF (Cronin and Taylor, 1992), perceived service quality (Grönross, 1984) and total quality management (Rahman and Nasrin, 2024) are in practice to assess service quality. This study adopts the SERVQUAL model to assess the service quality of online learning in higher education.

SERVQUAL is a widely adopted model for examining the quality of service from the consumer's perspective. Developed in 1985 by a group of business scholars, it measures the gap between customer expectations and perceptions across five key service dimensions: tangibles, reliability, responsiveness, assurance, and empathy (Parasuraman, Zeithaml, and Berry, 1985, 1988). These dimensions are collectively represented by the acronym "RATER." Reliability implies the industry's capacity to execute the service correctly and honor its promises. Assurance encompasses the awareness and respect of personnel and their capability to stimulate confidence and trust. Tangibility includes physical evidence of the service, such as offices, equipment, managers, and communication materials. Empathy represents communication, access, and understanding of the customer. Responsiveness refers to the promptness of personnel to deliver facilities. The model is popular among researchers because it is flexible and can be applied across a wide range of research areas, including the education sector (Uppal, Ali, and Gulliver, 2017). Despite their broader representation and application in various sectors of the service industry, these dimensions cannot be translated into the education sector and need modifications before applying to the quality of digital education. Similarly, the SERVQUAL model does not cover the aspects of digital education that represent its basic infrastructures, such as electricity, internet, device ownership, space for learning, etc. In this study, these aspects represent a new dimension 'online-class' in aggregate.

## 2.2 Quality of e-Learning

Past studies suggest that while the digitalization of higher education demands continuity, it must be accompanied by quality control (El Mhouti, Nasseh, and Erradi 2013; UNICEF, 2024). Studies argue that the increasing application of digital technologies must be balanced with the quality they offer to students. The value of digitalized higher education is also a concern for the business community (Larson, 2009), managers (Rizos, Sfakianaki, and Kakouris, 2022), and parents (World Bank, 2024). Sumi and Kabir (2021) argue that implementing an operative e-learning system requires concerted efforts from all concerned.

Studies highlight the positive role of digitalization in higher education, arguing that constant quality measurement supports its further improvement. Majid et al. (2022) showed that online learning increased motivation, yielded satisfactory outcomes, and enhanced the teaching and learning environments. Sousa and Mourao (2022) emphasized that the digitalization of higher education should involve multi-stakeholder participation in decision-making to empower students by facilitating and promoting their learning processes. Haleem et al. (2022) argued that digital technologies improve student performance and enhance the learning environment. Naim (2021) found that quality standards helped the university improve its online courses, align course materials with learning outcomes, and offer a systematic approach to evaluating and improving course quality. Overall, digitalization enhanced education quality, met international standards, and improved teaching and learning experiences.

## 2.3 SERVQUAL Models for e-Learning Quality Assessments

It is critical to differentiate the quality of e-learning before and after the Covid 19 pandemic to identify factors associated with students' perceived quality and to find research needs.

Application of SERVQUAL model in education is diverse. Sumi and Kabir (2021) applied a modified SERVQUAL model with seven dimensions: reliability, assurance, tangibility, empathy, responsiveness, learning content, and materials. The study found that reliability, assurance, and empathy significantly impacted student satisfaction, while responsiveness did not. Learning materials strongly influenced students' perceptions of service quality. Wantara (2022) found that, except for assurance, students' perceptions of the other four SERVQUAL dimensions were significant in evaluating digital education at the University of Trunojoyo Madura Indonesia, during the Covid 19 pandemic. Agrawal, Verma, and Malhotra (2021) identified empathy, responsiveness, reliability, and web content as significant factors affecting student satisfaction. Limbu and Pham (2023) found that system quality and instructor/course material quality are substantive factors of student satisfaction. Despite their wider coverage of students' perception of digital education and the application of the SERVQUAL model, these studies pay less attention to the local factors that determine e-learning quality.

A few studies concerned this. In a study, Saleem et al. (2022) found that situational factors, such as poor access to technology and inconsistent electricity supply, significantly impacted online education quality in Pakistan. A need for innovative approaches to make digital learning more effective and qualified is suggested by Behera et al. (2023). Ramírez-Hurtado et al. (2021) maintain that fine-tuned collaboration among teachers, students, and administrative staff can enhance online learning quality in any context. These findings, nevertheless, do not hint

polymakers and concerned stakeholders in developing countries about the measures to be undertaken to address the quality of digital education.

Pre-pandemic studies on the quality of digital education highlight similar aspects as pandemic era literature and so leave similar gaps. Udo et al. (2011) applied a revised SERVQUAL model, including the website content as a new one to examine the factors affecting the quality of distance learning. The authors found that, except for reliability, the other four dimensions played a significant role in learning. Dursun, Oskaybas, and Gokmen (2013) found disproportionate weight given by students to the SERVQUAL dimensions. Responsiveness received the lowest score, followed by reliability, empathy, tangibles, and credibility. Pham et al. (2019) found an association between the quality of e-learning and the support service quality in HEIs in the pre-pandemic era. These studies, as stated before, do not suggest measures that could affect the quality of digital education in a pandemic-like situation.

## **2.4 Quality Perception Across Students' Demography**

The Covid 19 pandemic literature barely concern students' demography as the factor determining the quality of digital education. A few pre-pandemic studies explored the relationship between students' demography and their perception of the quality of digital education. Arthar-Nyarko, Twoli and Khatiti (2017) found that students' age, gender, marital status, and geographic distance had a significant relationship with e-Learning satisfaction. Min and Khoo (2014) also found that gender and levels of study were significantly associated with students' perception of the service quality of higher education. Dursun, Oskaybas, and Gokmen (2013) identified significant relationships between students' profession, gender, marital status, age, and place of study, and the perceived quality of digital education services. Richardson, Long, and Woodley (2003) found significant relationships between factors such as age, discipline, hearing status, prior qualifications, and perceived quality of education. Studies also found a non-linear relationship between students' age and gender, and service quality (Ilias et al., 2008). The present study explores the relationship between the students' demography and their perception towards the quality of digital education.

## **2.5 Research Context: Digital Education Quality in Nepal**

In Nepal, universities offered online courses since the commencement of the Online and Distance Learning Policy in 2007 (Upadhyaya et al., 2021). The digitalization trend sharply increased amid the Covid 19 pandemic. Nearly 85% of learners benefited through this mode across various levels of education (Sapkota, 2023). Nevertheless, very limited studies examined the quality aspect of the new mode of learning applying the SERVQUAL measures. Poudel (2021) examines all five service quality dimensions as prescribed by Parasuraman, Berry, and Zeithaml (1988) and their relationship with students' satisfaction in three colleges in the Chitwan district of Nepal. The study found that the empathy dimension is significantly associated with students' satisfaction regarding service quality provided by their colleges. The present study expands SERVQUAL measures at the university level and examines how university students perceived the quality of e-learning during the Covid-19 pandemic.

Other studies attempt to analyze the quality of digital education but do not fully apply SERVQUAL measures. For example, Upadhyaya et al. (2021) identified quality as one of the major components influencing students' perceptions of course delivery and participation, with significant differences based on students' residence (rural vs. urban) and internet infrastructure and electricity availability. Sharma et al. (2020) examined four dimensions of ODL—learner dimensions, instructor characteristics, technological characteristics, and course management and coordination and found that all four dimensions are positively correlated with students' satisfaction with online learning, with female students reporting higher satisfaction than male students. Adhikari (2024) investigated non-academic aspects of service quality.

Some studies examined the quality of distance education in Nepal in the pre-Covid 19 era. For instance, Baniya (2016) found empathy and responsiveness as significant predictors of service quality at management schools. A study by Shaky, Sharma, and Thapa (2017) noted that the quality of education through digital means should be analyzed in terms of accessibility, learning flexibility, and the ability to meet educational demands across diverse regions of Nepal.

Past studies show a paucity of literature focusing on the quality of e-learning offered by higher education institutions during the Covid 19 period. A notable gap exists in this area regarding the use of the SERVQUAL model, specifically in Nepal. Additionally, studies are rarely concerned 'online-class' as a predictor of service quality. Understanding the performance of distance education during the pandemic is crucial, as it serves as a substantial factor in predicting the future quality of e-learning. Students' perceptions and satisfaction with digital learning technologies would better predict the future of digital education.

### 3. Methods and Materials of the Study

This study investigated the quality of digital education from students' perspectives using the customized SERVQUAL model. The research focused on students' online learning experiences during the Covid 19 pandemic, applying 22 Likert scale perceptions (breakdown in table no. 1 below) to measure students' perceptions as suggested by the model and adding online-class as an additional dimension. The online-class dimension is taken as a moderating variable that mediates the other five dimensions of the SERVQUAL model. It consisted of 5 Likert scale statements, including factors such as supply of electricity, internet facility, availability of digital technologies, etc., which are not addressed by the five dimensions in the SERVQUAL model and are also critical in the case of a developing country like Nepal. The quality of e-learning was assessed with students' gender, marital status, employment status, and study location. Additionally, the relationship between online-classes and perceived service quality was also tested.

#### 3.1 Hypotheses

*H1: There is a significant difference in the perceived quality of e-learning based on gender.*

To maintain the quality of e-learning, both genders may not perceive it at the same level. The study by Dursun, Oskaybas, and Gokmen (2013) found no significant difference in the perceived and expected service quality of online learning based on gender. However, the context of Nepal is different, given the social, cultural, and rural-urban differences, which might have a significant role in the accessibility of education among genders.

*H2: There is a significant difference in the perceived quality of e-learning based on marital status.*

One advantage of online learning is that students can attend classes from home, regardless of their marital status (Dursun, Oskaybas, and Gokmen, 2013). This study evaluates whether this is the case among higher education students in different universities in Nepal.

*H3: There is a significant difference in perceived service quality of e-learning based on study location.*

The study location also influences the quality of e-learning. In rural areas, scarce internet access can reduce the effectiveness of online learning (Giri, 2021; Chaudhary et al., 2022). Thus, this study tests the hypothesis that there is a significant difference in the perceived service quality of e-learning based on study location (urban vs. rural areas).

*H4: There is a significant difference in the perceived quality of e-learning based on employment status.*

According to Nepal et al. (2024), employed individuals showed higher interest in online learning than unemployed people. This study tests the hypothesis that there is a significant difference in the perceived service quality of online learning based on employment status.

*H5: There is a significant relationship between the perceived quality of e-learning and online-class participation.*

Past studies found an association between online-class and service quality (Kanan et al., 2023; Tj & Tanuraharjo, 2020). This study tests the hypothesis that there is a significant relationship between the perceived service quality of digital education and online-class participation.

#### 3.2 Research Design

The study employed a quantitative design to assess the quality of e-learning from the student's perspective (Fig. 1). The SERVQUAL model, developed by Parasuraman, Zeithaml, and Berry (1985, 1988), was customized to measure perceived service quality across five dimensions, with an additional dimension for online-class. This approach, focusing solely on perceived service quality, has been similarly applied in studies by Al-Mushasha and Nassuora (2012), Uppal, Ali and Gulliver (2017), and Pham et al. (2019) in the context of e-learning. The SERVQUAL model is flexible and has been adapted to measure service quality in various fields. Due to the pandemic in Nepal, the government imposed a lockdown, which forced educational institutions for long educational vacations. It affected the academic calendar, forcing universities to apply internet-based learning as an alternative to continuing education. So, expected service quality and the gap between expected and perceived service quality defined by the model were irrelevant. So, this study only considered perceived service quality.

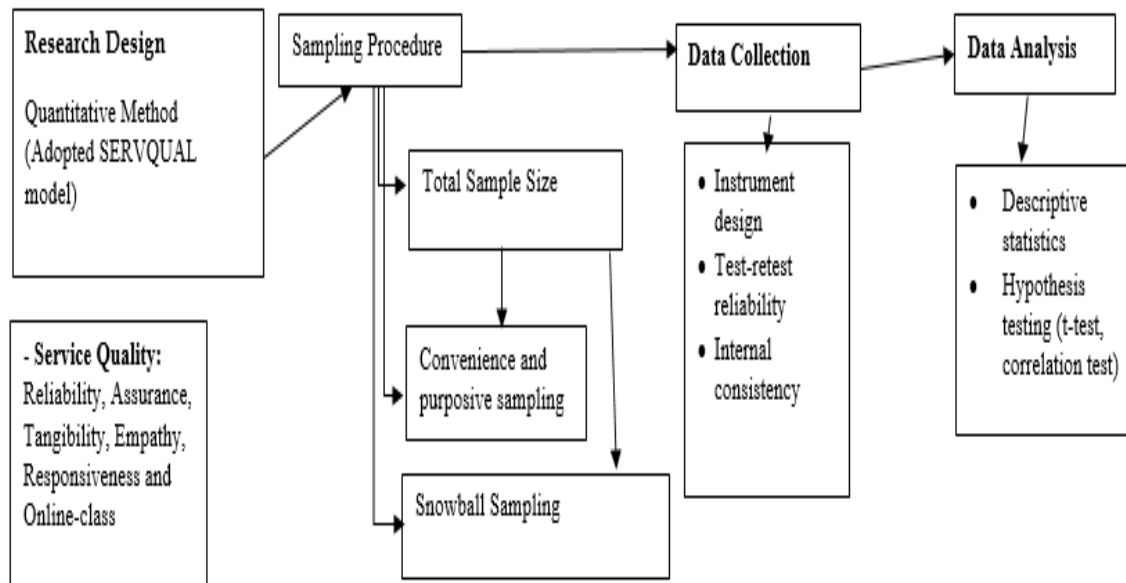


Figure 1: Research Design and Steps

### 3.3 Population and Sampling Procedure

The study targeted students who participated in online classes at universities in Nepal during the first and second Covid 19 lockdown periods. Therefore, the students who took classes during the first and second national lockdowns were the population of the study. The total number of students who joined online classes across Nepalese universities was unknown. Given the restrictions on human mobility, a snowball method was employed to select the sample.

Using convenience and purposive sampling methods, students who were in direct contact with the researcher through social media platforms and institutional associations were reached out for the study, and through the snowball sampling method, the researcher requested university faculty and administrators to distribute the survey questionnaire among students who were taking online classes during the pandemic. The rationale behind choosing convenience and purposive as well as snowball sampling is, first, due to the lockdown, all the universities were forced to adopt online teaching and learning. Second, reaching out to the students directly for data collection was impossible, and the researcher had to rely on their networks. These methods may introduce selection bias, as respondents may not fully represent the broader student population, which was also the researcher's concern during the research period. To address this, researchers diversified participants from multiple universities across Nepal (Fig. 2). For the reduction of possible bias, researchers ensure balanced representation from rural and urban municipalities and exclude data from small areas but include respondents from various places. This approach aimed to enhance the study's validity by capturing diverse student experiences and minimizing bias.



Figure 2: Selected Universities and Survey Response Counts

### 3.4 Data Collection

A survey questionnaire was distributed to online platforms as this approach is considered appropriate, as recommended by Kelley-Quon (2018) and Mohajan (2020). The survey was conducted among students pursuing bachelor's and MA in six universities, as shown in Figure 2, across Nepal between April and July 2021 during national lockdowns for Covid 19. Viber and WhatsApp for those in direct contact with the researchers and already engaged in online classes. For other students, questionnaires were distributed through the university faculty and staff. Given the lockdown constraints, snowball sampling was deemed suitable as it allows researchers to reach individuals with specific characteristics who might be difficult to identify otherwise (Parker, Scott and Geddes, 2019). The first nationwide lockdown began on 24 March 2020 and lasted until 21 July 2020, during which all public and private institutions, including higher education, were closed. In response, online learning was initiated on 28 March 2020.

Approximately 650 questionnaires were distributed, and 512 responses were returned. Of these, 19 were found to be duplicate responses, which were retained for reliability testing. The study population is unknown due to the difficulty of tracking the total number of students in online classes during the lockdowns. After careful evaluation, 451 responses met the inclusion criteria, while 42 were excluded. Thus, 451 valid responses were analyzed in the study.

### 3.5 Data Analysis

To develop the scale, dimensions were initially selected based on the model discussed by Parasuraman, Zeithaml, and Berry (1985, 1988) for identifying service quality. An additional dimension focusing on online-class was adopted from the works of Uppal, Ali, and Gulliver (2017) and Pham et al. (2019). Based on these dimensions, an initial set of 42 Likert scale statements was created. The scale ranged from 1 (strongly disagree) to 5 (strongly agree).

The open-source platform Jamovi (version 2.3.28) was utilized to analyze the data. The survey included 27 questions related to six dimensions: reliability, responsiveness, assurance, empathy, tangibility, and online-class. Since a Likert scale was used across dimensions, an average (mean score) was calculated by adding rating scores across all responses divided by the total responses. These were the scores for six dimensions. Using these scores, descriptive statistics were drawn. Additionally, mean and standard deviation values were computed based on gender, place of origin, marital status, and employment status to compare service quality across different groups. To identify differences between these groups, t-tests were conducted.

Furthermore, correlation coefficients were calculated to analyze the relationship strengths and directions among dimensions. The significance of these relationships was assessed using the p-value, with a standard threshold of 95% confidence level. Descriptive statistics were used to summarize the data because mean and standard deviations provide the average scores of each dimension with their average deviations from the means. This summary is used to explain the level of perceived quality on each dimension and is also used for comparison based on gender, marital status, employment status, and studying place. For hypothesis testing, an independent t-test was used to test the significant difference in scores of dimensions of service quality based on gender, marital status, employment status, and study place. Correlation analysis and significant tests were performed to test the relationship between online classes and perceived quality.

### 3.6 Instrument Development and its Reliability and Validity

A two-step process was employed to ensure the validity of the questionnaire because an additional dimension, online-class was added. Content validity was analyzed for validation of that dimension as well as statements related to SERVQUAL. First, four experts with at least five years of experience in online teaching (since 2015) were consulted for validity analysis. In the second step, seven students who had completed at least one online course were selected. Initially, the researchers discussed the questionnaire separately with each group. Following adjustments based on their feedback, a combined meeting was conducted online, following the approach suggested by Sumi and Kabir (2021). Based on these discussions, 22 statements related to service quality and five statements related to online-class were finalized for the study.

Cronbach's Alpha was conducted to assess the internal consistency of each SERVQUAL dimension and the online-class dimension. The results indicated that the SERVQUAL dimensions—tangibility, reliability, assurance, and empathy—had Alpha values ranging from 0.7 to 0.8, considered acceptable (Doll et al. 1995). However, the Alpha value for the responsiveness dimension was 0.577, which falls into questionable range. Despite this, the

overall Cronbach's Alpha for SERVQUAL (22 items) and the online-class dimension (5 items) were 0.88 and 0.818, respectively, indicating good internal reliability (Doll et al., 1995). Individual values are presented in Table 1.

**Table 1: Cronbach's Alpha**

| Dimensions            | Number of Items | Cronbach's Alpha | Intra-class Correlation |
|-----------------------|-----------------|------------------|-------------------------|
| <b>Tangible</b>       | 4               | 0.709            | 0.983                   |
| <b>Reliability</b>    | 5               | 0.704            | 0.937                   |
| <b>Responsiveness</b> | 4               | 0.577            | 0.961                   |
| <b>Assurance</b>      | 4               | 0.706            | 0.989                   |
| <b>Empathy</b>        | 5               | 0.752            | 0.944                   |
| <b>Online class</b>   | 5               | 0.818            | 0.984                   |

An intra-class correlation (ICC) analysis was conducted using 50 samples with a test-retest method, considering a 15-day interval between tests. Test-retest was conducted to ensure the reliability of data obtained from respondents, and intra-class correlation was calculated to ensure the reliability of data collected from respondents. The analysis was performed in SPSS 27, using a two-way mixed-effects model with absolute agreement. The ICC values ranged from 0.931 to 0.989, indicating excellent reliability, as suggested by Koo and Li (2016).

#### 4. Results and Findings

The demographic distribution of respondents is shown in Table 2. 54.99% of the respondents were female, while 45.01% were male, whereas 82.71% of the respondents were unmarried (single), and 17.29% were married. Most respondents were working (68.51% of the total respondents) and 53.66% of the respondents were studying online from rural areas (villages), whereas 46.34% were studying from urban areas (cities). Graduate students were higher than undergraduates in response where 56.76% of the respondents were pursuing graduate (master's) studies, while 43.24% were at the undergraduate (bachelor's) level.

**Table 2: Demographic Information (n=451)**

| Variable                                      | Frequency | Percent (%) |
|-----------------------------------------------|-----------|-------------|
| <b>Gender</b>                                 |           |             |
| Female                                        | 248       | 54.99       |
| Male                                          | 203       | 45.01       |
| <b>Marital Status</b>                         |           |             |
| Unmarried                                     | 373       | 82.71       |
| Married                                       | 78        | 17.29       |
| <b>Employment status</b>                      |           |             |
| Unemployed                                    | 309       | 68.51       |
| Employed                                      | 142       | 31.49       |
| <b>Geographic location of the respondents</b> |           |             |
| Rural area (village)                          | 242       | 53.66       |
| Urban area (city)                             | 209       | 46.34       |
| <b>Pursuing Degree</b>                        |           |             |
| Undergraduate (bachelor)                      | 195       | 43.24       |
| Graduate (master)                             | 256       | 56.76       |

The average values for the five service quality attributes, tangibles, reliability, responsiveness, assurance, and empathy, ranged from 3.58 to 3.86, where 0 was the minimum and 5 was the maximum, with standard deviations between 0.576 and 0.646. They were averaged to calculate service quality value. This resulted in an overall average perceived service quality score of 3.72 with a standard deviation of 0.473. The average value for online-class was 3.43, with a standard deviation of 0.781, as shown in Table 3 (also in figure 3).

**Table 3: Descriptive Statistic of Dimensions of Service Quality and Online-Class**

| Dimensions             | Mean        | Standard deviation | Coefficient of Variance |
|------------------------|-------------|--------------------|-------------------------|
| Tangible               | 3.59        | 0.646              | 17.99%                  |
| Responsiveness         | 3.58        | 0.631              | 17.63%                  |
| Reliability            | 3.83        | 0.576              | 15.04%                  |
| Assurance              | 3.86        | 0.624              | 16.17%                  |
| Empathy                | 3.75        | 0.624              | 16.64%                  |
| <b>Service quality</b> | <b>3.72</b> | <b>0.473</b>       | <b>12.72%</b>           |
| <b>Online class</b>    | <b>3.43</b> | <b>0.781</b>       | <b>22.77%</b>           |

The zero-order correlation among tangible, reliability, responsiveness, assurance, and empathy was weak and positive, but the p-value shows that these relations were significant. Details are shown in Table 4(also in figure 3).

**Table 4: Correlation Matrix of Dimensions of Service Quality**

|                       | Tangible       | Reliability    | Responsiveness | Assurance      | Empathy        | Service quality |
|-----------------------|----------------|----------------|----------------|----------------|----------------|-----------------|
| <b>Tangible</b>       | <b>1</b>       |                |                |                |                |                 |
| <b>Reliability</b>    | <b>0.653**</b> | <b>1</b>       |                |                |                |                 |
| <b>Responsiveness</b> | <b>0.406**</b> | <b>0.407**</b> | <b>1</b>       |                |                |                 |
| <b>Assurance</b>      | <b>0.485**</b> | <b>0.553**</b> | <b>0.467**</b> | <b>1</b>       |                |                 |
| <b>Empathy</b>        | <b>0.341**</b> | <b>0.431**</b> | <b>0.467**</b> | <b>0.521**</b> | <b>1</b>       |                 |
| <b>Online Class</b>   | <b>0.466**</b> | <b>0.483**</b> | <b>0.330**</b> | <b>0.526**</b> | <b>0.434**</b> | <b>0.591 **</b> |

Note: \*\* indicates that significant at 0.01

The average perceived quality between males and females was slightly different but more than 3, and the p-value shows that the difference was not significant, whereas the value was higher for married respondents but significant compared to unmarried respondents. Similarly, the correlation between online class and service quality was 0.591 with a p-value of 0.000, indicating that the relation is positively moderate and significant. Correlation results are shown in Table 5.

**Table 5: Service Quality of Digital Education and its Significance**

| Category          | Mean | SD    | SE     | t-value | p-value |
|-------------------|------|-------|--------|---------|---------|
| <b>Female</b>     | 3.71 | 0.402 | 0.0255 | 0.425   | 0.671   |
| <b>Male</b>       | 3.73 | 0.549 | 0.0385 |         |         |
| <b>Married</b>    | 3.69 | 0.480 | 0.0248 | 2.693   | 0.007   |
| <b>Unmarried</b>  | 3.85 | 0.420 | 0.0476 |         |         |
| <b>Rural</b>      | 3.68 | 0.475 | 0.031  | 1.814   | 0.070   |
| <b>Urban</b>      | 3.76 | 0.468 | 0.032  |         |         |
| <b>Unemployed</b> | 3.70 | 0.401 | 0.026  | -0.947  | 0.344   |
| <b>Employed</b>   | 3.75 | 0.500 | 0.042  |         |         |

The average perceived quality of online learning between respondents from rural and urban areas was slightly different but more than 3, and the p-value shows that the difference was insignificant. Similarly, the perceived quality of online learning among married respondents was little more than unmarried, and test statistics (p-value=0.007) are significant.

The results of the hypothesis tests are summarized in Table 6. The p-values for differences in perceived service quality based on gender, study location, and employment status were all greater than 0.05, indicating that these differences were not statistically significant. As a result, H1, H3 and H4 were rejected. The p-value for the difference in perceived quality based on marital status was less than 0.05, indicating a significant difference,

leading to the acceptance of H2. H5 regarding a significant and positive relationship between service quality and online classes was also accepted.

**Table 6: Hypotheses Testing**

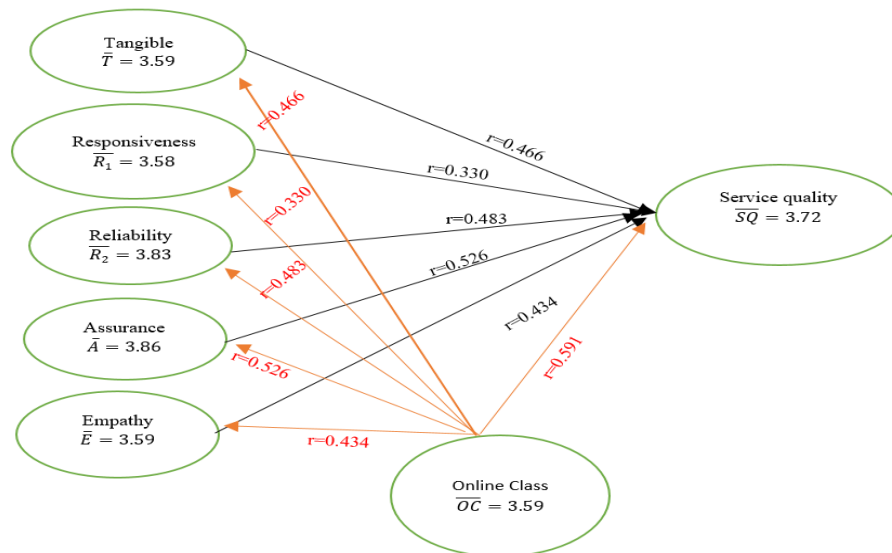
|    | Statement                                                                                         | Test Value                | Sig.  | Result      |
|----|---------------------------------------------------------------------------------------------------|---------------------------|-------|-------------|
| H1 | There is a significant difference in the perceived quality of e-learning based on gender.         | 0.425                     | 0.671 | H1 Rejected |
| H2 | There is a significant difference in the perceived quality in e-learning based on marital status. | 2.693                     | 0.007 | H2 Accepted |
| H3 | There is a significant difference in perceived quality in e-learning based on study place.        | 1.814                     | 0.070 | H3 Rejected |
| H4 | There is a significant difference in perceived quality in e-learning based on employment status.  | -0.947                    | 0.344 | H4 Rejected |
| H5 | There is a significant relationship between perceived quality and online-classes.                 | 0.591 (correlation value) | 0.000 | H5 Accepted |

## 5. Discussion

This study found that student's perception of the quality of e-learning during the pandemic in Nepal was moderate and that they were satisfied with the service. It indicated that students agreed on the quality of online education during the pandemic. The range of elements of service quality was moderate (between 3 and 4), implying that there was room for further improvement in the quality of digital education. Assurance had the highest value, followed by reliability, empathy, tangible, and responsiveness. In particular, the subject matter of teaching, lecture delivery, understanding, and interaction received low scores. This finding is similar to the findings of past studies (Al-Mushasha and Nassuora 2012; Nayak and Alam 2022; Alqahtani et al. 2022). The higher score given to assurance means that students were satisfied with teachers' and staff's services. However, responses also indicated that besides lecture delivery, instructors must use discussion and presentation methods to make lectures more engaging.

The study results indicate no significant difference in perception of digital service quality concerning students' gender and employment status. Thus, males and females do not perceive significant differences in service quality. However, there was a significant difference between married and unmarried respondents and respondents learning from rural and urban areas. This finding is similar to the findings of past studies (Agrawal, Verma, and Malhotra, 2021; Dursun, Oskaybas, and Gokmen, 2013; Limbu and Pham, 2023; Richardson, Long, and Woodley, 2003). Udo et al. (2011) confirmed that besides reliability, students' perceived quality was impacted by four other dimensions. In a study, Sumi and Kabir (2021) found that students' ratings of all dimensions were harmonious, and there was a significant effect on perceived service quality. Students' positive perception of the quality of online education services during the Covid 19 pandemic confirms its usefulness and perceived benefit.

Regarding the new online-class dimension, the score was lower than other dimensions of the SERVQUAL model (Fig. 3). Statements on this dimension included factors such as electricity supply, internet speed, availability of digital technologies, and space for the study, which received a moderate response. Past studies also show that students' motivation for online learning in the Global South was affected by the lack of electricity supply, access to the internet, and multimedia devices (Agrawal, Verma, and Malhotra, 2022; Chaudhary et al. 2022). This implies that HEIs must address electricity and internet facilities both within the institution and on the students' side. Implementing a mechanism to gather feedback from parents may help HEIs to ensure the quality of online education. Since the perception varies by place of study (rural and urban), HEIs may also collaborate with government agencies to address issues rural students face, such as electricity, internet, and digital technology.



**Figure 3: Mean and Correlation Coefficients of Dimensions of SERVQUAL and Online class**

The findings concerning the overall dimensions of the SERVQUAL model, precisely the new online-class dimension, reveal that higher education institutions across developing countries, including Nepal, could rethink higher education policies as well as the pedagogical avenues. Blended mode of learning could be more effective to satisfy the students' needs. Policymakers need to address the digital divide between rural and urban areas. Managers in the universities should train both instructors and staff on aspects of e-learning pedagogy, such as engagement with students, develop feedback system, and interactive classes. A major dimension of the new policy should be to enhance the quality of online-class. This would significantly improve quality of e-learning in universities.

## 6. Conclusion

This study concludes that Nepalese students positively responded to the switch to e-learning during Covid 19 despite emergency measures, implying their satisfaction with the quality of the service. Among five SERVQUAL dimensions representing the quality of the education service, students rated high responsiveness, followed by reliability, tangibility, assurance, and responsiveness. Online class dimension was another predictor of the quality of e-learning that precisely reflects the condition of service delivery in Nepal. It was the precondition to mend the e-learning service through the other five dimensions. The quality of distance education also depends on students' demography. This study further concludes that HEIs in Nepal need specific programs to improve the quality of e-learning, focusing primarily on the items represented by online-class dimension. HEIs need to collaborate with government agencies to maintain and improve online teaching facilities. HEIs need to organize faculty training programs. Instructors need to be dressed up and combine lectures and group discussion methods. Mechanisms to establish contacts and get feedback responses from students and other stakeholders seem helpful.

### 6.1 Limitations and Future Research Implications

Even though this study expanded the knowledge of the student's perception of the quality of e-learning, it has several limitations. Similarly, prospects for future research remain.

First, this study is based on samples from six different universities in Nepal. Therefore, generalizations of e-learning quality may be overstatements because they may differ in the case of individual universities. This might be true because universities in Nepal vary regarding history, number of students, teachers and administrative staff, and location. Each factor influences the quality of distance education. Any further studies could pick one of the HEIs and examine the quality.

Second, this research partially adopted SERVQUAL measures, i.e., 22 items of Likert scales of perception, but scales for expectation measurement are left. We argue that expectation is not an applicable component to measure the quality of digital education during the Covid 19 pandemic because the online learning mode was an emergency measure that students had no other options to retain in the education system. Assessing the gap

between expectation and perception could reveal an accurate picture of digital education's quality. Future research should consider this.

Third, this study contextualized the Covid 19 pandemic to assess the quality of e-learning. It has not been considered the pre- and post-Covid 19 periods to examine the quality as digitalization is gaining momentum post-pandemic, either solo or blended mode, its quality concerns in all contexts and periods. Therefore, any future studies could compare pre-pandemic era and post-pandemic era data and examine an accurate picture of the quality of e-learning and predict its future.

**Ethics Statement:** Participants in this study were informed before including in this study. They gave full approval of using the information for the purpose of this study. The survey was also approved by the ethics committee of Center for Nepal and Asian Studies (CNAS), Tribhuvan University (TU).

**AI Statement:** Authors declare that artificial intelligence was not used to prepare this study.

## References

- Adhikari, S., 2024. Service quality and students' satisfaction in higher education institutions: Nepali students' experiences and perceptions. *Prithvi Academic Journal*, 7 (2), pp.101-110. doi.org/10.3126/paj.v7i1.65767
- Agrawal, P., Verma, A. and Malhotra, S., 2021. An analysis of perceived service quality and students satisfaction of e-learning during Covid 19 in higher education institution. *The Online Journal of Distance Education and e-Learning*, 9(3), pp. 341-352. <https://www.researchgate.net/publication/354150349>.
- Al Rawashdeh, A. Z. et al., 2021. Advantages and disadvantages of using online learning in university education: Analyzing students' perspectives. *Electronic Journal of Online Learning*, 19, pp. 107–117. doi: 10.34190/ejel.19.3.2168
- Al-Mushasha, N. F. and Nassuora, A. B., 2012. Factors determining e-learning service quality in Jordanian higher education environment. *Journal of Applied Sciences*, 12(14), pp.1474-1480. doi: 10.3923/jas.2012.1474.1480
- Alqahtani, M.A. et al., 2022. Investigating students' perceptions of online learning use as a digital tool for educational sustainability during the COVID-19 pandemic. *Frontier in Psychology*, 13, pp. 886272. doi: 10.3389/fpsyg.2022.886272
- Angell, R. J., Heffernan, T. and Megicks, P., 2008. Service quality in postgraduate education. *Quality Assurance in Education*, 16(3), pp. 236-254. doi.org/10.1108/09684880810886259
- Arthar-Nyarko, E., Twoli, N. W. and Khatiti, D., 2017. Learner demographic, resource Characteristics, and responsiveness to eLearning delivery in selected distance education institutions in Ghana. *International Journal of Education and Research*, 5 (8), pp. 13-24. <https://www.ijern.com/journal/2017/August-2017/02.pdf>
- Baniya, R., 2016. Relationship between perception of service quality and students' satisfaction – A case study of a management school. *Journal of Education and Research*, 6(2), pp. 41-60. <https://www.learntechlib.org/p/208733/>
- Behera, A. K., Sousa de, R. A., Oleksik, V., Dong, J. and Fritzen, D., 2023. Student perceptions of remote learning transitions in engineering disciplines during the COVID-19 pandemic: A cross-national study. *European Journal of Engineering Education*, 48(1), pp.110-142. doi: 10.1080/03043797.2022.2080529
- Chan, R. Y., Bista, K. and Allen, R. M., 2021. Is online and distance learning the future in global higher education? The faculty perspectives during COVID-19. In: Chan RY, Bista K and Allen RM (eds) *Online Teaching and Learning in Higher Education during COVID-19*. New York: Routledge, pp.1-14. doi: 10.4324/9781003125921-2
- Chaudhary, G. P., Khadka, R. M., Lamichhane, A., Bhawana, D., Das, N., Tharu, N. R., Karki, K. D., and Pandey, J., 2022. Impact of COVID-19 pandemic on learning status of student in Nepal. *Journal of Education and Health Promotion*, 11(4), pp.314-334. doi.org/10.4103/jehp.jehp\_354\_22
- Cronin, J. and Taylor, S. A., 1992. Measuring service quality: A reexamination and extension. *Journal of Marketing*, 56, pp. 55-67. doi: 10.2307/1252296
- Davies, S., 2021. Universities need to realize they're in the service industry. Available at: <https://www.timeshighereducation.com/opinion/universities-need-realise-theyre-service-industry> (accessed 7 April 2024).
- Devkota, K. R., 2021. Inequalities reinforced through online and distance education in the age of COVID-19: the case of higher education in Nepal. *International Review of Education*, 67 (1-2), pp. 145–165. doi: 10.1007/s11159-021-09886-x
- Doll, W. J., Raghunathan, T. S., Lim, J. S., and Gupta, Y. P., 1995. A confirmatory factor analysis of the user information satisfaction instrument. *Information Systems Research*, 6(2), pp. 177-188. doi:10.1287/isre.6.2.177
- Dugenio-Nadela et al., 2023 Service quality and students' satisfaction in higher education institution. *Journal of Human Resource and Sustainability Studies*, 11 (4), pp. 858-870. doi: 10.4236/jhrss.2023.114049
- Dursun, T., Oskaybas, K. and Gokmen, C., 2013. The quality of the service of the distance education. *Procedia - Social and Behavioral Sciences*, 103, pp. 1133 – 1151. doi: 10.1016/j.sbspro.2013.10.441
- El Mhouti, A., Nasseh, A. and Erradi, M. 2013. How to evaluate the quality of digital learning resources? *International Journal of Computer Science Research and Application*, 3 (3), pp. 27-36.
- Eltahir, M. E., Alsalhi, N. R. and Al-Qatawneh, S. S., 2022. Implementation of e-exams during the COVID-19 pandemic: A quantitative study in higher education. *PloS one*, 17(5), pp. e0266940. doi.org/10.1371/journal.pone.0266940
- Ganagalla, S., 2023. The Importance of Educational Evaluation in Ensuring Quality Teaching. *Research & Reviews: Research Journal of Educational Studies*, 9(2), pp. 7-9. doi: 10.4172/JES.9.2.004

- Gautam, D. K. and Gautam, P. K., 2021. Transition to online higher education during COVID-19 pandemic: turmoil and way forward to developing country of South Asia-Nepal. *Journal of Research in Innovative Teaching and Learning*, 14, pp. 93–111. doi: 10.1108/JRIT-10-2020-0051
- Ghotbabadi, A. R., Baharun, R. and Feiz, S., 2012. A review of service quality models. In: 2nd International Conference on Management, 11th-12th June 2012. Langkawi Kedah, Malaysia. Available at: [https://www.researchgate.net/publication/230669329\\_A\\_REVIEW\\_OF\\_SERVICE\\_QUALITY\\_MODELS](https://www.researchgate.net/publication/230669329_A_REVIEW_OF_SERVICE_QUALITY_MODELS) (accessed 7/27/2024).
- Giri, S., 2021. Online education in Nepal: Prospects and challenges. *International Journal of Science and Research*, 10, pp. 1368-1375. doi:10.21275/SR21318103310.
- Grönross, C., 1984. A service quality model and its implications. *European Journal of Marketing*, 18(4), pp. 36-44. doi: 10.1108/EUM00000000004784
- Haleem, A., Javaid, M., Quadri, M. A., & Rajiv, S., 2022. Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers*, 3, pp. 275–285. doi.org/10.1016/j.susoc.2022.05.004
- Ilias, A., Abu Hassan, H. F., & Rahman, R. A., (2008) Student Satisfaction and Service Quality: Any Differences in Demographic Factors? *International Business Research*, 1 (4), pp. 131-143.
- Kajenthiran, K. and Karunanithy, M., 2015. Service Quality and Student Satisfaction: A Case Study of Private External Higher Education Institutions in Jaffna, Sri Lanka. *Journal of Business Studies*, 1 (2), pp. 46-61.
- Kanan, M. et al., 2023. Online education and managing service quality with the challenges of COVID 19: The case of University of Business and Technology (UBT) Saudi Arabia). *Applied Mathematics and Information Science*, 17(2), pp. 201-207. doi: 10.18576/amis/170201
- Koo, T. K. and Li, M. Y., 2016. A guideline of selecting and reporting intra-class correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), pp. 155–163. doi.org/10.1016/j.jcm.2016.02.012
- Larson, R. C., 2009. Education: Our most important service sector service science. *Social Science*, 1(4), pp. i-iii. doi: 10.1287/serv.1.4.i
- Limbu, Y. B. and Pham, L., 2023. Impact of e-learning service quality on student satisfaction during the COVID-19 pandemic: A systematic review. *Knowledge Management & E-Learning*, 15(4), pp. 523–538. doi.org/10.34105/j.kmel.2023.15.030
- Majid, A. H., Sanjari, A. A., Abdul, G. and Al-Abaidi, R. 2022. The Role of Digital Education in Improving the Quality of Education. *Res Militaris*, 12, pp. 3923-3923.
- Márquez-Ramos, L., 2021. Does digitalization in higher education help to bridge the gap between academia and industry? An application to COVID-19. *Industry and Higher Education*, 35(6), pp. 630-637. doi.org/10.1177/0950422221989190
- Min, S. and Khoon, C. C., 2014. Demographic Factors in the Evaluation of Service Quality in Higher Education: A Structural Equation Model (SEM) Approach. *International Journal of Marketing Studies*, 6 (1), pp. 90-102. doi:10.5539/ijms.v6n1p90
- Mohajan, H. K., 2020. Quantitative research: A successful investigation in natural and social sciences. *Journal of Economic Development, Environment and People*, 9(4), pp. 50-79.
- Muthuprasad, T., Aiswarya, S., Aditya, K. S., & Jha, G. K., 2021. Students' perception and preference for online education in India during COVID-19 pandemic. *Social Sciences and Humanities Open*, 3, pp. 100101. doi: 10.1016/j.ssaho.2020.100101
- Naim, A., 2021. Application of Quality Matters in Digital Learning in Higher Education. *Texas Journal of Multidisciplinary Studies*, 1, pp. 3-12.
- Nayak, K. V. and Alam, S., 2022. The digital divide, gender and education: Challenges for tribal youth in rural Jharkhand during COVID-19. *Decision*, 49, pp. 223–237. doi: 10.1007/s40622-022-00315-y
- Nepal, R. M., Khadka, B., Guragain, S., & Ghimire, J., 2024. Interest and motivation of disadvantaged students towards online learning during the COVID-19 pandemic in Nepal. *Frontiers in Education*, 9, pp. 1356279. doi: 10.3389/educ.2024.1356279
- Parasuraman, A., Zeithaml, V. A. and Berry, L. L., 1985. A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49, pp. 41–50.
- Parasuraman, A., Zeithaml, V. A. and Berry, L. L., 1988. SERVQUAL: A Multiple-Item Scale for Measuring Customer Expectations of Service. *Journal of Retail*, 64, pp. 12–40.
- Parker, C., Scott, S. and Geddes, A., 2019. *Snowball Sampling*. SAGE Research Methods Foundations. (In Press). Available at: 211022791.pdf (core.ac.uk)
- Pavlov, O. and Hoy, F., 2018. Higher education as a service. In: 44th Annual Conference of the Eastern Economic Association, Boston, 1-4 March, 2018, pp. 1-22. Available at: [https://www.researchgate.net/publication/374155603\\_Higher\\_Education\\_as\\_a\\_Service](https://www.researchgate.net/publication/374155603_Higher_Education_as_a_Service)
- Pham, L., Limbu, Y. B., Trung, B. K., Nguyen, H. T., & Pham, H., 2019. Does e-learning service quality influence e-learning student satisfaction and loyalty? Evidence from Vietnam. *International Journal of Educational Technology in Higher Education*, 16(1), pp. 1-26. doi: 10.1186/s41239-019-0136-3
- Poudel, S. P., 2021. Service quality provided by colleges in chitwan and its impact on student satisfaction. *Nepalese Journal of Management Research*, 1 (1), pp. 15-21. <http://balkumarcollge.edu.np/journal>
- Rahman, M. M. and Nasrin, S., 2024. Perceived service quality at higher education institutions: A study on the success factors of total quality management practices in Bangladesh. *Social Sciences & Humanities Open*, 10, pp. 100997. doi.org/10.1016/j.ssaho.2024.100997

- Ramírez-Hurtado, J. M., Hernandez-Díaz, A. G., Lopez-Sanchez, A. D. and Perez-Leon, V. E., 2021. Measuring online teaching service quality in higher education in the COVID-19 environment. *International Journal of Environmental Research and Public Health*, 18, pp. 2403. doi.org/10.3390/ijerph18052403
- Richardson, J. T. E., 2005. Students' perceptions of academic quality and approaches to studying in distance education. *British Educational Research Journal*, 31(1), pp. 7-27. doi: 10.1080/0141192052000310001
- Richardson, J. T. E., Long, G. L. and Woodley, A., 2003. Academic Engagement and Perceptions of Quality in Distance Education. *Open Learning: The Journal of Open, Distance and eLearning*, 18 (3), pp. 223-244. doi: 10.1080/0268051032000131008
- Rizos, S., Sfakianaki, E. and Kakouris, A., 2022. Quality of administrative services in higher education. *European Journal of Educational Management*, 5(2), pp. 115-128. doi.org/10.12973/eujem.5.2.115
- Saleem, F., Al Nasrallah, W., Malik, A. M., & Rehman, S. U., 2022. Factors affecting the quality of online learning during COVID-19: Evidence from a developing economy. *Frontier in Education*, 7, 847571. doi: 10.3389/feduc.2022.847571
- Sapkota, P. P., 2023. E-learning during COVID-19 lockdown among college students of Nepal: Opportunities and challenges. *KMC Journal*, 5(2), pp.104-118. doi: 10.3126/kmcj.v5i2.58233
- Schleicher, A. 2020. The impact of COVID-19 on education: Insights from education at glance 2020. Available at: <https://www.oecd.org/education/the-impact-of-covid-19-on-education-insights-education-at-a-glance-2020.pdf> (Accessed March 21, 2023)
- Shakya, S., Sharma, A. and Thapa, K. B., 2017. State Education System with e-learning in Nepal: Impact and Challenges. *Journal of the Institute of Engineering*, 2017, 13(1), pp. 10-19.
- Sousa, M. J. and Mourão, T., 2022. Metrics and indicators of online learning in higher education. In: Ullah A et al. (eds) *Proceedings of International Conference on Information Technology and Applications. Lecture Notes in Networks and Systems* (vol 350). Singapore: Springer, pp 719-729. doi.org/10.1007/978-981-16-7618-5\_61
- Sumi, R. S. and Kabir, G., 2021 Satisfaction of E-Learners with Electronic Learning Service Quality Using the SERVQUAL Model. *Journal of Open Innovation and Technology Marketing and Complexity*, 7 (2), pp. 1-27. doi.org/10.3390/joitmc7040227
- Tj, H. W. and Tanurahrarjo, H. H., 2020. The effect of online learning service quality on student satisfaction during COVID19 pandemic in 2020. *Journal Manajemen Indonesia*, 20(3), 240-251. doi: 10.25124/jmi.v20i3.3520
- Udo, G. J., Bagchi, K. K. and Kirs, P. J., 2011. Using SERVQUAL to assess the quality of e-learning experience. *Elsevier*, 27(3), pp. 1272-1283. doi.org/10.1016/j.chb.2011.01.009
- UNICEF 2024. Top 10 reasons digital learning succeeds or fails: Understanding how we can unlock the potential of digital learning. Available at: <https://www.unicef.org/innocenti/top-10-reasons-digital-learning-succeeds-or-fails> (accessed 18 May, 2024).
- Upadhayaya, P. R., Sharma, B., Gnawali, Y. P., & Belbase, S., (2021) Factors influencing graduate students perceptions' of online and distance learning in Nepal. *Turkish Online Journal of Distance Education-TOJDE*, 22(3), pp. 236-269.
- Uppal, M. A., Ali, S. and Gulliver, S. R., 2017. Factors determining e-learning service quality. *British Journal of Educational Technology*, 49(3), pp. 412-426.
- Wantara, P., 2022. Measuring service quality of lecturer in COVID-19 condition. *European Journal of Business and Management Research*, 7 (1), pp. 253-259. doi: 10.24018/ejbmr.2022.7.1.1285
- Watson, R., 2015. Quantitative research. *Nursing Standard*, 29(31), pp. 44-8. doi: 10.7748/ns.29.31.44.e8681
- World Bank 2024. Remote learning during COVID-19: Lessons from today, principles for tomorrow. Available at: <https://www.worldbank.org/en/topic/edutech/brief/how-countries-are-using-edtech-to-support-remote-learning-during-the-covid-19-pandemic> (accessed 23 July 2024).
- Zaw, W.M., and Hlang, S. S., 2024. Bridging the educational gap: The role of digital learning platforms in developing countries. *International Journal of Digital Education*, 1(1), pp.11-15. doi: <https://doi.org/10.61132/ijed.v1i1.122>