Assessment Validity in the Era of Generative AI Tools

Eltahir Kabbar and Bashar Barmada

https://doi.org/10.34074/proc.240105
Correspondence: Eltahir Kabbar, ekabbar@unitec.ac.nz

This publication may be cited as:

Kabbar, E., & Barmada, B. (2024). Assessment Validity in the Era of Generative AI Tools. In H. Sharifzadeh (Ed.), *Proceedings: CITRENZ 2023 Conference, Auckland, 27–29 September* (pp. 26–33). ePress, Unitec. https://doi.org/10.34074/proc.240105

Contact:
epress@unitec.ac.nz
www.unitec.ac.nz/epress/
Unitec
Private Bag 92025, Victoria Street West
Tāmaki Makaurau Auckland 1142
Aotearoa New Zealand

ISBN: 978-1-99-118344-6

ePress
Research with impact
Unitec

Unitec is a business division of Te Pūkenga – New Zealand Institute of Skills and Technology

ABSTRACT

Generative AI tools, a recent disruptive educational technology, are expected to change how education is delivered and administered. This study proposes a risk identification framework to support educators in identifying assessment integrity risks caused by generative AI tools. The framework also suggests possible actions to mitigate these risks. The proposed framework uses four factors (Assessment Type, AI Knowledge, Course Level, and Bloom's Taxonomy Cognitive Domain Level) to identify the risks associated with an assessment resulting from the usage of generative AI tools. It is critical to have such a framework to ensure the integrity of assessments while the education industry adapts to the generative AI tools era.

KEYWORDS

Artificial intelligence, generative AI tools, assessment validity, assessment evaluation framework

INTRODUCTION

Over the past few decades, various information communication technologies (ICTs) have significantly impacted teaching and learning in education. In recent years, many educators and educational institutes have been interested in applying educational technologies as they improve the learner experience (Flavin, 2012). Our working definition of educational technologies is informed by Bajpai and Leedham (1970), who take a holistic view of applying different systems to support teaching and learning.

The recent adoption of several educational technologies by educational institutes has disrupted the conventional way of learning and forced changes to their practices (Cassidy, 2023). These technologies have profoundly impacted education, transforming how students learn and teachers teach. As technology evolves, education will likely continue to be disrupted and changed in new and exciting ways.

This paper proposes a framework to support educators in designing and developing valid assessments that meet the regulatory authorities' requirements while addressing the challenges posed by the most recent disruptive technology (generative AI tools) and its risk to the integrity of assessments. The proposed framework is discussed in line with the assessment types approved by the New Zealand Qualifications Authority (NZQA) while being generic enough to be used as a blueprint for other education systems. This framework aims to enable educators to produce valid assessments that meet the learning-outcome requirements while being resilient to the threat posed by misusing generative AI tools.

DISRUPTIVE TECHNOLOGIES IN EDUCATION

In the 1980s, personal computers (PCs) became increasingly affordable and accessible, leading to a widespread adoption of PCs by educational institutes. Learners had to acquire computer skills to access digital educational resources quickly. Simultaneously, educators had to change how they delivered course material and assessment practices to take advantage of PC technology. The emergence of the internet in the 1990s transformed education by providing access to vast amounts of information, enabling online communication and collaboration. The

widespread use of the internet led to the development of online courses and distance learning programmes (Flavin, 2012; Gejendhiran et al., 2020).

In the early 2000s, learning management systems (LMSs) were introduced, providing a centralised platform for managing and delivering educational content and allowing for more personalised and interactive learning experiences (Buckley & Doyle, 2014; Oliveira et al., 2016). In addition, the use of LMSs enhanced course resource accessibility by students (Hew, 2015), enabling course instructors to customise students' learning experiences, moving away from the one-size-fits-all model (Berking & Gallagher, 2016). The introduction of LMSs also enabled educators to create streamlined assessment and feedback processes (Dahlstrom et al., 2014), and allowed students better visibility and control of their progress (Zabolotniaia et al., 2020), among other benefits.

Furthermore, the evolution of internet-enabled handheld devices, such as smartphones and tablets in the 2010s, enhanced learners' access to learning resources (Ait-Hroch & Ibrahimi, 2024). This disruptive technology introduced a mobile-learning (m-learning) aspect to digital learning (Sharples et al., 2010). In addition, it enabled learners to collaborate and communicate better with peers and educators (Hrastinski, 2008), and to have instant access to educational resources (Sheng et al., 2010), and better engagement with course content (Hamari et al., 2014; Salhab & Daher, 2023). An exciting pedagogical impact of this technology is the gamification of learning and the anticipated impact this may have on education (Flavin, 2012; Hamari et al., 2014).

In recent years, artificial intelligence and machine learning have developed adaptive learning platforms to personalise each student's learning experience (Chen et al., 2019), based on individual learning patterns and preferences (Kuppusamy, 2019). With the surge of devices enabled by the internet of things (IoT), more data can be collected and analysed to boost the learning capabilities of learners, based on a better understanding of their behaviours.

GENERATIVE AI TOOLS

Generative AI tools, also called AI chatbots, are the most recent types of disruptive technology in education. These tools are not standard search engines; they take the user prompt and use an intelligent algorithm based on natural language modelling to analyse the request, then they search through an extensive database of information, such as articles, books and websites, to formulate a response (Agar, 2023). Because they rely on AI, generative AI tools can learn over time and continuously improve their responses. Further, Abd-Elaal et al. (2019) assert that AI chatbots, such as grammar tools, use other assistive tools to enhance their decisions.

How Organisations React to Disruptive Technologies

Some academic institutes consider the use of AI without explicit permission to be a breach of assessment integrity. The response from such institutes is to ignore or fight AI usage. For example, an educational institute may develop tools that can detect AI-generated content, (Abd-Elaal et al., 2019); the University of Melbourne, the University of New South Wales and Waikato University have already adopted tools to detect the use of AI in student work (University of Waikato, n.d.).

However, others, such as the University of Sydney, who acknowledge generative AI as a tool to improve the student learning experience but are reluctant to adopt such tools before thorough testing and verification, have taken a different approach and adopted a wait-and-see strategy before committing either way.

A third group, including such institutes as the University of Wollongong and Massey University (Massey University, 2023), seem to embrace AI and consider its use in education to be beneficial to the learning process, emphasising that students need to learn how to use it in the right way (Saunders, 2023).

NZQA Assessment Types

The New Zealand Qualifications Authority (NZQA) is the official New Zealand body that awards secondary- and tertiary-level qualifications. They ensure that the education providers' programmes match corresponding qualification requirements and are recognised nationally and internationally. NZQA lists various assessment tools and approaches, classified into five main categories: Oral, Written, Practical, Verification and Other. Table 1 shows the NZQA categories and some of their assessments. For the complete list of NZQA assessments, the reader is advised to check (NZQA, n.d.). Depending on the qualification level, assessments are assigned to learners, and with the help of educational frameworks, such as Bloom's Taxonomy (Forehand, 2005), the tasks of assessments are defined.

Table 1. NZQA assessment categories and some of their assessment methods.

Written evidence	Practical evidence	Oral evidence	Verification evidence	Other evidence
- Reports - Portfolio - Tests - Assignments - Tables/Charts/Forms - Booklets - Worksheet	- Project - Simulation - Demonstration - Model - Posters	- Presentations - Questions/Answers - Interviews - Speech	- Feedback	- Prior Knowledge

THE PROPOSED APPROACH: A GENERIC FRAMEWORK

Generative AI tools are here to stay, disrupting the education process and current assessment methods. However, there is no unified approach to managing the impact of generative AI tools on the education process, particularly on assessment integrity. This paper proposes a framework to support educators in deciding the risk level AI poses for existing assessments. Furthermore, this paper will identify changes that can be used to improve assessment methods and practices, and increase resilience against the recent development of AI. The framework shown in Figure 1 encourages educators to consider four factors when assessing the level of risk posed by generative AI tools to the current assessment practices and take appropriate action to address the identified risk level. The four factors to consider are Assessment Type, AI Knowledge, Course Level and Bloom's Taxonomy Cognitive Domain Level.

Assessment Type

NZQA assessment types fall within five categories, as shown in Table 1. The risks associated with assessments vary depending on the assessment type. For example, the risks associated with the integrity of written reports as evidence of learning due to the illegitimate use of AI tools are potentially higher than in-class presentations or tests carried out in a controlled environment. Therefore, educators must consider the assessment type in their risk-assessment evaluation.

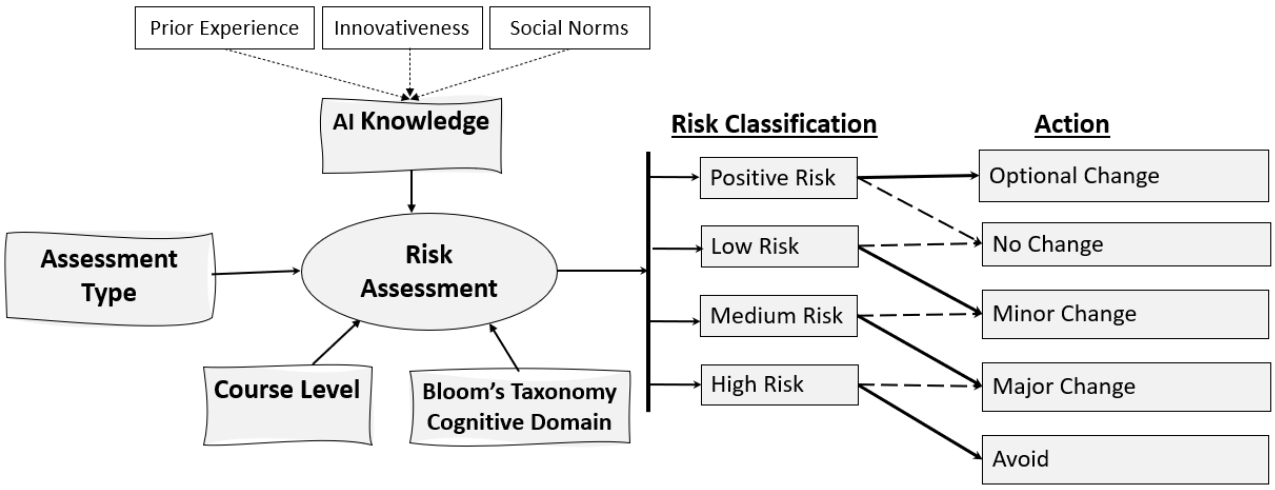

Figure 1. Proposed risk-evaluation framework to mitigate the impact of generative AI tools on assessment validity.

AI Knowledge

The second factor in the framework is AI Knowledge, which focuses on the educator's awareness of the different generative AI tools, their usage in an educational context, and their functionality. The educators' AI knowledge is also shaped by their prior experience, innovativeness, and the social norms and practices of their organisation. An adequate AI knowledge level is critical for educators in reaching accurate risk-assessment outcomes.

Course Level

The third factor in the framework is Course Level. Academic institutes offer different qualifications at different levels. The assessments associated with each course are usually designed to assess learning at the designated level using the appropriate assessment type. While some assessment types pose a significant risk at certain levels, the same assessment type could be less risky at a different level; therefore, the course level must be considered during the risk assessment. For example, a written report at Level 5, where students are expected to recall facts and basic concepts, poses a relatively lower risk compared to a written report at Level 7, where students are expected to draw connections among ideas in a literature review domain.

Bloom's Taxonomy Cognitive Domain Level

The fourth factor to be considered is the Bloom's Taxonomy Cognitive Domain Level embedded in the course learning outcome(s). Often, a course has different learning outcomes, and various assessment types are used to collect evidence of learning. Therefore, the course assessment's integrity risks can vary depending on the cognitive domain level associated with each assessed task. Generally, the lower the Bloom's Taxonomy domain level, the higher the risk, as various generative AI tools can be illegitimately used to produce work that would require students to work at the Remember/Understand levels compared to the Evaluate/Create levels. Therefore, educators must consider the Bloom's Taxonomy level during the risk assessment.

While considering that each of the four factors discussed provides a reasonable risk-level assessment, the accuracy of the risk-level assessment can only be achieved by considering all elements.

The outcome of the risk-assessment exercise leads to one of the four risk-classification statuses described in Figure 1. Assessments with a 'Positive Risk' status are valid; however, educators may opt to change the requirements to take advantage of the opportunities presented by generative AI tools. We recommend educators make changes to their

assessments to improve the overall student learning experience and quality of the assessment with the existence of generative AI tools while maintaining the validity of the assessment.

In addition, assessments with a 'Low Risk' status are also valid; therefore, no changes are required because generative AI tools have little to no impact on the assessment validity. We recommend educators make minor changes to their assessments, leading to more resilient assessments to mitigate the risk of generative AI tools.

Assessments with a 'Medium Risk' status require greater changes, such as modifying/updating the assessment instrument from written reports to written assessments to be completed under a controlled environment, to ensure the assessment is resilient against generative AI tools (as shown in Figure 2). While the required changes may not affect the core of the assessment requirements, we believe these changes will significantly improve the assessment's resilience against generative AI tools, moving it to the Green Zone.

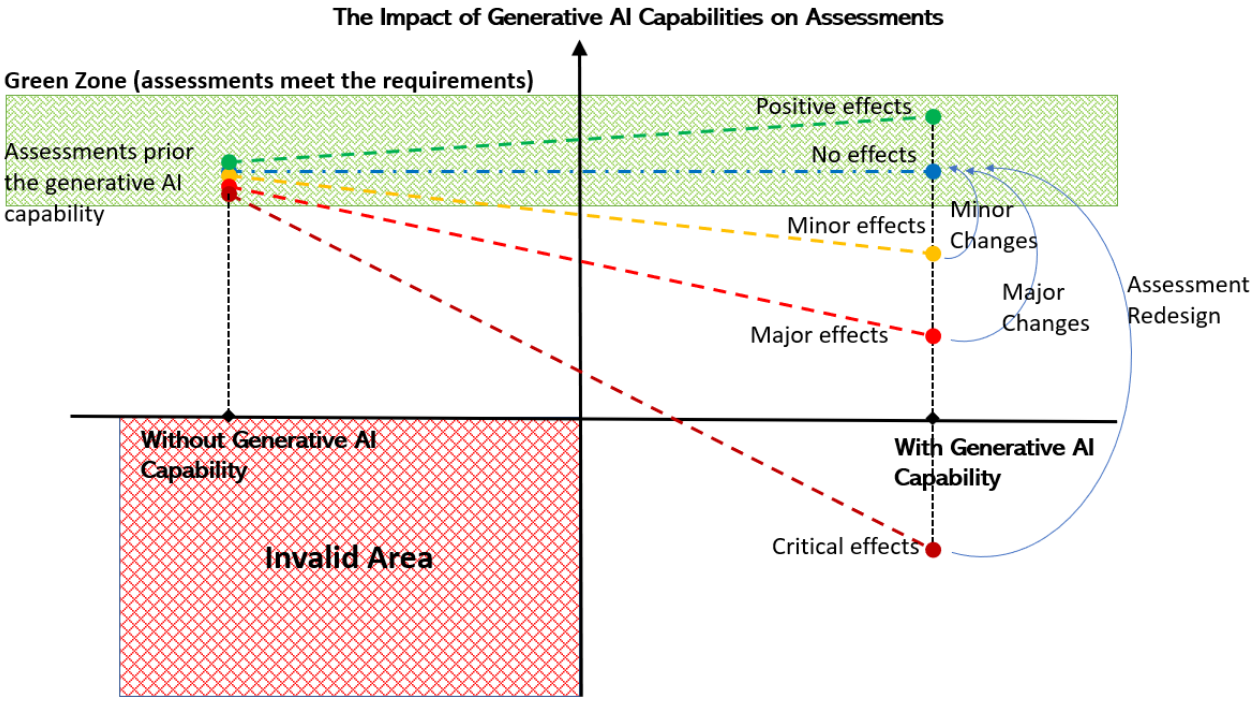

Figure 2. The impact of generative AI capabilities on assessments.

Finally, we recommend avoiding the assessments with 'High Risk' status, or at least undertaking significant changes to the assessment requirements and type, leading to a complete redesign. Educators are encouraged to change assessments that can be easily solved by generative AI tools, such as written research reports or even developing software, and include presentation/viva components in these assessments to allow assessors to judge better the authenticity of the work submitted. We acknowledge that the approval of changes that impact a programme can be complex, and the redesign takes time; hence, the option exists to make significant changes.

CONCLUSION

Disruptive technologies have constantly challenged the norms and brought exciting opportunities to many industries, including education. Generative AI tools are one of the most recent disruptive technologies in education. This technology is here to stay, and students around the globe are taking advantage of it to learn and complete the required assessments. Academic institutes react differently to the challenges posed by generative AI tools, particularly to assessment integrity. Some educational institutes have banned the use of these technologies in assessments, while others consider them to be valuable tools for learning and actively set policies to promote them.

Some academic institutes have yet to decide on a clear approach to the use of generative AI tools, leaving it to the academic staff to decide. Nevertheless, the impact of generative AI tools on assessments makes educators nervous, because of the potential implications on assessment integrity and ensuring that an assessment's learning outcomes are met without compromising its integrity.

The paper has introduced a framework to assist educators in navigating and managing the risks posed by generative AI tools to ensure assessment validity during our transition into the new education norm. We believe that generative AI tools are a game changer; the proposed model will help educators maintain assessment validity and integrity during the transition, but ultimately, a rethink of how we assess course learning outcomes is required.

REFERENCES

Abd-Elaal, E-S., Gamage, S. H. P. W., & Mills, J. E. (2019). Artificial intelligence is a tool for cheating academic integrity. *AAEE Conference Proceedings*. https://aaee.net.au/wp-content/uploads/2020/07/AAEE2019_Annual_Conference_paper_180.pdf

Agar, N. (2023). *The AIs are coming: Will ChatGPT create a future of bullsh*t (jobs)?* The University of Waikato. https://www.waikato.ac.nz/news-opinion/media/2023/the-ais-are-coming-will-chatgpt-create-a-future-of-bullsht-jobs

Ait-Hroch, A., & Ibrahimi, A. (2024). The place of e-learning in language learning: Its contribution to the enhancement of language proficiency. In M. Khaldi (Ed.), *Innovative instructional design methods and tools for improved teaching* (pp. 341–362). IGI Global. https://doi.org/10.4018/979-8-3693-3128-6.ch017

Bajpai, A. C., & Leedham, J. (1970). *Aspects of educational technology* (vol. 4). Association for Programmed Learning and Educational Technology.

Berking, P., & Gallagher, S. (2016). *Choosing a learning management system*. Advanced Distributed Learning. https://adlnet.gov/assets/uploads/ChoosingAnLMS.pdf

Buckley, P., & Doyle, E. (2014). Gamification and student motivation. *Interactive Learning Environments*, *24*(6), 1162–1175. https://doi.org/10.1080/10494820.2014.964263

Cassidy, C. (2023). Australian universities split on using new tool to detect AI plagiarism. *The Guardian*. https://www.theguardian.com/australia-news/2023/apr/16/australian-universities-split-on-using-new-tool-to-detect-ai-plagiarism

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, *8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Dahlstrom, E., Brooks, D. C., & Bichsel, J. (2014). The current ecosystem of learning management systems in higher education: Student, faculty, and IT perspectives. *Educause*. https://doi.org/10.13140/RG.2.1.3751.6005

Flavin, M. (2012). Disruptive technologies in higher education. *Research in Learning Technology, ALT-C 2012 Conference Proceedings*. https://doi.org/10.3402/rlt.v20i0.19184

Forehand, M. (2005). Bloom's Taxonomy. In M. Orey (Ed.), *Emerging perspectives on learning, teaching and technology* (pp. 41–47). Global Text. https://textbookequity.org/Textbooks/Orey_Emergin_Perspectives_Learning.pdf

Gejendhiran, S., Arokia Anicia, S., Selvam, V., & Kalaimani, M. (2020). Disruptive technologies – A promising key for sustainable future education. *9th World Engineering Education Forum, Elsevier, Procedia Computer Science*, *172*, 843–847. https://doi.org/10.1016/j.procs.2020.05.121

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. *IEEE*, 3025–3034. https://ieeexplore.ieee.org/document/6758978

Hew, K. F. (2015). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS. *British Journal of Educational Technology*, *46*(2), 288–301. https://doi.org/10.1111/bjet.12235

Hrastinski, S. (2008). What is online learner participation? A literature review. *Computers & Education*, *51*(4), 1755–1765. https://doi.org/10.1016/j.compedu.2008.05.005

Kuppusamy, P. (2019). Smart education using internet of things technology. In L. Y. Zongmin Ma (Ed.), *Emerging technologies and applications in data processing and management* (pp. 385–412). IGI Global. https://doi.org/10.4018/978-1-5225-8446-9.ch017

Massey University. (2023). *Use of artificial intelligence in assessment policy*. Retrieved April 3, 2024, from https://www.massey.ac.nz/documents/1445/Use_of_Artificial_Intelligence_in_Assessment_Policy.pdf

NZQA. (n.d.). *Assessment tools and approaches*. Retrieved July 24, 2023, from https://www.nzqa.govt.nz/providers-partners/assessment-and-moderation-of-standards/assessment-of-standards/generic-resources/assessment-tools-and-approaches/#:~:text=There%20are%20two%20different%20types,task%20assessment%20and%20evidence%20assessment

Oliveira, P. C. D., Cunha, C. J. C. D. A., & Nakayama, M. K. (2016). Learning management systems (LMS) and e-learning management: An integrative review and research agenda. *Journal of Information Systems and Technology Management*, *13*, 157–180. https://www.scielo.br/j/jistm/a/TT3Pk4mwkp5Cmmbf4NbfPvw/

Salhab, R., & Daher, W. (2023). The impact of mobile learning on students' attitudes towards learning in an educational technology course. *Multimodal Technologies and Interaction*, *7*(7), 74. https://doi.org/10.3390/mti7070074

Saunders, S. (2023, February). Rather than ban generative AI, universities must learn from the past. *University World News*. https://www.universityworldnews.com/post.php?story=20230221075136410

Sharples, M., Taylor, J., & Vavoula, G. (2010). A theory of learning for the mobile age. In B. Bachmair (Ed.), *Medienbildung in neuen Kulturräumen* (pp. 87–99). VS Verlag für Sozialwissenschaften. https://www.researchgate.net/publication/226417865_A_Theory_of_Learning_for_the_Mobile_Age

Sheng, H., Siau, K., & Nah, F. F. H. (2010). Understanding the values of mobile technology in education: A value-focused thinking approach. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *41*(2), 25–44. https://doi.org/10.1145/1795377.1795380

University of Waikato. (n.d.). *Academic integrity – Student assessment guide*. Retrieved April 3, 2024, from https://www.waikato.ac.nz/students/student-assessment/assessment-at-university/academic-integrity

Zabolotniaia, M., Cheng, Z., Dorozhkin, E., & Lyzhin, A. (2020). Use of the LMS Moodle for an effective implementation of an innovative policy in higher educational institutions. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(13), 172–189. https://www.learntechlib.org/p/217602/

AUTHORS

Dr Eltahir Kabbar is a Senior Lecturer at the School of Computing Electrical and Applied Technology at Unitec. His research interests include diffusion of innovation (DOI), technology adoption and e-government systems. He has several publications, including a book chapter, journal articles and international conferences. ekabbar@unitec.ac.nz

Dr Bashar Barmada is a Senior Lecturer and Academic Programme Manager in the School of Computing Electrical and Applied Technology at Unitec. His research areas are mainly in IoT applications and network security. bbarmada@unitec.ac.nz