Education Research and Perspectives
An international journal

Volume 51, 2024, Pages 37-54

AI and Assessment Design in Engineering Disciplines

Pejman Ghasemi Poor Sabet[1†]

Shen (Jason) Zhan [2]

Milad Baghalzadeh Shishehgarkhaneh[3]

1. Civil Construction Department, Acknowledge Education

2. Faculty of Engineering and Information Technology, University of Melbourne

3. Faculty of Engineering, Monash University

Artificial intelligence (AI) is transforming multiple facets of the competitive business world and education. Despite this, the full potential of AI applications within education remains unclear because of the lack of a comprehensive framework on how to use AI in developing assessments across various academic disciplines. While incorporating AI into assessment design can streamline the integration of diverse learning components, it is essential to establish clear performance criteria to ensure the validity of these assessments. In engineering education, the integration of AI in assessment design poses ethical concerns, accountability issues, and limitations in capturing diverse learning forms, and therefore careful consideration is required to ensure fairness and pedagogical value are maintained. This paper aims to conceptualise effective assessment processes using AI in engineering education and demonstrate the potential of various validity techniques, providing a comprehensive framework for leveraging AI to enhance assessment accuracy and effectiveness in engineering education. The discussion elucidates the key components necessary for developing effective assessments through the application of AI capabilities.

[†]Address for correspondence: Acknowledge Education, 168 Exhibition Street, Melbourne, VIC 3000, Australia Email: pejman.sabet@ae.edu.au

Introduction

Artificial intelligence (AI) is driving transformative changes across many sectors worldwide (Gruetzemacher & Whittlestone, 2022). Higher education is no exception, and the debate surrounding the benefits of AI seems far from reaching a conclusion (Moorhouse et al., 2023). Lodge et al. (2023) argued that the application of AI technologies in educational practices promises to enhance learning outcomes, streamline administrative processes and offer innovative approaches to assessment design.

Recent studies such as Chen et al. (2020) have emphasised that AI-driven education platforms can offer personalised learning experiences by leveraging machine learning algorithms and intelligent tutoring systems, highlighting the shift from traditional computer-based education to intelligent, adaptive systems that cater to individual student needs. Similarly, Ciolacu et al. (2018) discussed how early recognition systems, incorporating AI, can predict student performance, further supporting the potential for personalised learning paths to enhance student outcomes.

However, the potential benefits of AI in education have not been fully realised because of the absence of a unified framework spanning diverse academic disciplines (Rudolph et al., 2023). The absence of such a framework is particularly evident in interdisciplinary approaches, where AI technologies such as those explored in multiple intelligence-based teaching systems could be integrated into curricula to foster creativity and problem-solving across fields (Lo et al., 2021). As expressed by Kizilcec et al. (2024), a further consideration is doubts about whether AI-based assessment is authentic. As highlighted by Ting et al. (2023), concerns over the authenticity and validity of AI-driven assessments continue to grow. Although these authors proposed a framework for automating assessments through natural language processing and machine learning, they noted the challenges in ensuring these assessments meet educational standards. Therefore, this study is based on the research question: How can the application of AI extend the design creativity in curricula and contribute to a valid assessment?

As educational institutions and educators explore the use of AI in education, particularly in the realm of assessment, they encounter the challenge of establishing robust performance criteria to validate AI-driven assessments (Swiecki et al., 2022). The lack of a comprehensive framework complicates efforts to leverage AI effectively and ensure the validity of assessments generated using AI applications (Memarian & Doleck, 2024). However, as demonstrated by researchers such as Fayoumi and Hajjar (2020), who applied artificial neural networks to forecast academic performance and improve decision-making processes, higher education frameworks that incorporate advanced learning analytics and AI technologies can provide deeper insights into student performance and feedback loops.

To address the lack of a clear framework for how to use AI assessments in education, this paper conceptualises the process of ensuring effective assessment using AI. The study explains several validity techniques that can be applied through AI and proposes an AI-based framework to ensure assessment accuracy and effectiveness. By providing insights into the key components necessary for developing AI-driven assessments, this research offers a structured approach to harnessing AI's full potential in education. Through a detailed discussion, the paper illustrates how AI capabilities can be strategically applied to enhance assessment practices, ultimately contributing to more effective and reliable educational outcomes.

Literature Review

Effective Assessment and AI Applications in Engineering

AI refers to computational algorithms that imitate biological mental processes and can be employed for 'tasks like learning, understanding, estimating, problem-solving, suggesting, and decision-making across various fields, including engineering design' (Yüksel et al., 2023, p.1). Valid assessment is an essential part of evaluating students' learning, particularly in engineering education contexts, where assessments often measure complex problem-solving skills and practical applications (Almond et al., 2002). The traditional process of creating a well-designed assessment product can be time-consuming. This process requires carefully crafting tasks that map learning items across the learning context. In this domain, AI technologies provide innovative solutions by supporting assessment design and automating repetitive tasks, making the process more efficient (Menekse, 2023).

In the field of engineering, project-based learning and assessment play a vital role in high-quality education (Palmer & Hall, 2011), and AI applications can enhance these processes by automating assessment design tasks. This automation could be partial or full, depending on the criteria specified for the AI tool (Swiecki et al., 2022). Swiecki et al. (2022) discussed that some AI applications can generate different types of assessment tasks (e.g. multiple-choice tests and open-ended question tests); however, limited research has discussed the validity of such AI-generated assessments.

Potential Pitfalls in AI Application for Designing Assessments in Engineering

The application of AI in assessment design within engineering education presents several potential challenges, including ethical concerns, accountability issues, and limitations in assessing diverse forms of learning. One major ethical issue is data privacy, which arises because AI's reliance on large-scale data collection can compromise student confidentiality. Additionally, AI tools, such as learning analytics and generative AI, may introduce algorithmic biases that could perpetuate existing inequalities in education and reinforce harmful social stereotypes (Akgun & Greenhow, 2022).

Another significant challenge is the 'black box' nature of AI decision-making, where the lack of transparency in how AI algorithms generate outcomes makes it difficult for educators and students to understand or question assessment results. This could lead to biased assessments because AI algorithms may embed the biases of the programmers and designers who develop the algorithms (Hanesworth et al., 2019). This highlights the need for oversight and accountability mechanisms to ensure that AI-driven assessments in education remain fair and transparent.

Additionally, while AI offers efficiency, it may limit the pedagogical value of assessments. Traditional assessments are often used by educators to offer personalised feedback, inspire students and adapt teaching methods to individual needs. However, AI-driven assessments tend to standardise evaluations, which can undermine creative, ethical and innovative aspects of learning—qualities that are essential in engineering education. The risk is that AI will reinforce a narrow focus on technical skills and problem-solving, neglecting broader educational goals such as collaboration, creativity and critical thinking (Swiecki et al., 2022).

Validity Techniques via AI for Designing Assessment

To ensure the validity of assessment design using AI, assessment designers and educators should carefully consider the key factors outlined below.

Item analysis

Item analysis plays a crucial role in refining questions intended for future tests and eliminating misleading items from current assessments. Therefore, it is essential for educators to be proficient in testing techniques to ensure they can accurately and effectively evaluate student progress (Quaigrain & Arhin, 2017).

Content alignment

Ajjawi et al. (2020) discussed that successful course design requires a clear definition of meaningful course objectives; the alignment of relevant learning activities and tasks; and the establishment of measurable assessments. These authors defined alignment as the extent to which there is conceptual, procedural and methodological consistency among the various components of a curriculum system (Barthakur et al., 2022). Assessment rubrics can effectively contribute to aligning assessment content with the educational content (Ajjawi et al., 2020).

Construct validity

Construct validity refers to the degree to which an assessment accurately measures the theoretical construct it claims to assess. In the context of AI-based assessments, construct validity involves verifying that the AI system is evaluating the intended skills or knowledge rather than extraneous factors. This validation process must integrate multiple sources of evidence, including content relevance, response processes and internal consistency. Downing (2003) emphasised that construct validity encompasses various facets such as the relationship to other variables and the consequences of the assessment's use, ensuring that the AI is truly capturing the intended educational outcomes. Miller and Linn (2000) highlighted the need for structural and external validation to assess how well AI-driven assessments generalise across different contexts and student groups. Ensuring construct validity not only supports the credibility of AI assessments but also enhances their utility in educational decision making.

Reliability check

Reliability in assessment design is vital to ensure that student evaluations are consistent, accurate, and meaningful over time. Reliability focuses on the consistency of assessment results when applied under similar conditions across different cohorts or contexts. Effective assessment frameworks in engineering courses often rely on capstone projects or senior design experiences, which provide students with opportunities to demonstrate the skills they have developed throughout their degree (Damaj & Yousafzai, 2019). These frameworks must be carefully designed to align with learning outcomes and ensure minimal variability in evaluation across diverse assessment components. Machine learning models, such as extreme learning machines, have been introduced to improve the reliability of assessments by analysing patterns in student performance data. These models have demonstrated higher predictive accuracy in engineering courses compared to traditional assessment methods (Deo et al., 2020).

Bias detection

Bias detection in AI-based educational assessments is essential for ensuring fairness and accuracy in evaluations. Algorithmic bias can arise from the data used to train the model or from the model itself, leading to unequal treatment of certain groups, such as those defined by race, gender, or socioeconomic status (Baker & Hawn, 2022).

One key method for detecting bias is statistical analysis. For example, fairness metrics such as the N-Sigma method enable systematic identification of biased outputs by comparing AI-generated decisions across different demographic groups (DeAlcala et al., 2023). Bias in educational contexts can also result from an algorithm's reliance on historical data that embeds societal biases, further complicating AI's role in equitable decision-making (Srinivasan & Chander, 2021).

To address these challenges, developing frameworks for the regular assessment and mitigation of biases is critical. For instance, frameworks applied to AI decision-making processes for scholarships can serve as a model to ensure fairness and inclusivity in AI-based assessments (Austin et al., 2023).

Feedback analysis

Feedback analysis in AI-driven educational assessments plays a vital role in enhancing student learning by providing timely and personalised feedback. AI technologies, such as natural language processing and educational data mining, enable sophisticated feedback systems that surpass the capabilities of traditional methods (Shishehgarkhaneh et al., 2024). For example, AI-assisted systems can generate both written and non-verbal feedback based on student interaction data, improving the feedback process (Bulut & Wongvorachan, 2022). In higher education, the integration of AI models enables immediate feedback that can be tailored to each student's learning experience, enhancing motivation and engagement (Hooda et al., 2022). The use of AI to classify and analyse student feedback via sentiment analysis can help educational institutions improve teaching practices and identify areas for improvement (Shaik et al., 2022).

Predictive validity

Predictive validity refers to the extent to which an assessment accurately predicts future outcomes, such as academic success or job performance. In AI-driven educational assessments, ensuring predictive validity is essential to confirm that AI tools can reliably forecast student performance. Recent research indicates that machine learning models, such as logistic regression and neural networks, demonstrate high predictive accuracy in determining future academic achievement based on student performance data (Fokkema et al., 2022). However, these models require careful validation to ensure they perform effectively across diverse populations and within different educational contexts (Lau & Yuen, 2009).

Iterative improvement

Iterative improvement is a process of continuously refining educational assessments by using data from previous iterations to enhance accuracy and outcomes. In the context of AI-driven educational assessments, iterative improvement helps to incrementally optimise models and tools, improving both predictive validity and student engagement over time. AI techniques, such as iterative learning control, can significantly improve assessment performance by fine-tuning algorithms based on feedback from prior assessments, leading to better tracking of student progress and more accurate predictions (Altın et al., 2017). Moreover, in e-learning platforms, iterative improvement allows AI to personalise learning recommendations by continuously adjusting models based on student performance data, thereby enhancing learning outcomes (Bagunaid et al., 2022).

Validation report

A validation report is essential to ensure the accuracy, reliability and credibility of educational assessments, particularly those involving AI-driven tools. In the realm of AI-based educational assessments, validation involves collecting evidence to support the interpretations and decisions made based on assessment outcomes. Key frameworks, such as Messick's six facets of construct validity, emphasise the importance of content, structural, and generalisability aspects in ensuring that educational assessments are robust and defendable in practice (Miller & Linn, 2000). Additionally, systematic validation methods, including model-centred validation, expert opinion and trial-based assessments, help in continuously refining AI systems by identifying errors and improving the systems' predictive capabilities (Myllyaho et al., 2021).

Discussion and Conclusions

This conceptual paper presents some professional insights from a civil construction department from an Australian educational provider delivering vocational education training (VET) and higher education (HE) courses. Based on Australian guidelines for VET and HE courses, the practice of validity techniques is mandated for assessment developers; rather, all the validity schemes discussed in the literature review and listed in Table 1 must be implemented to ensure an assessment is valid and effectively measures students' learning.

The findings of this paper highlight a structured approach to developing valid assessments via AI applications, addressing the critical gap in comprehensive guidelines for applying AI in the context of engineering assessment. AI-based assessments leverage various inputs, including curriculum objectives, student data and industry experiences, to ensure valid and reliable outcomes. These assessments use multiple validity techniques such as content alignment, construct validity, reliability checks, bias detection and feedback analysis to refine the version and improve it over time. The iterative application of these techniques ensures that each subsequent assessment version becomes more robust, addressing overlaps and integrating complementary aspects of the techniques. The result is a continuously validated assessment tool that adapts based on predictive validity and iterative improvement.

In the literature review, the rationale for selecting specific validity techniques has been clearly established by linking them to essential inputs and expected outputs. These techniques can be automatically applied in developing a valid assessment, provided they are supported by the required inputs. Table 1 illustrates the conceptual function of both a human and AI-based assessments, detailing how specific inputs and validity techniques contribute to producing a valid and robust assessment process. This comparative analysis demonstrates that while human-based assessment processes are often time-intensive and prone to inaccuracies, AI-based assessments significantly enhance efficiency and precision in measuring student learning outcomes.

Table 1. The conceptual function of AI-based assessment

AI-based assessment		
Input	***Validity techniques***	***Output***
• Curriculum and learning objectives • Student data • Assessment design parameters • Content sources • Industry experience input	• Content alignment • Construct validity • Reliability checks • Bias detection • Feedback analysis • Predictive validity • Iterative improvement • Validation reports	Valid assessment

When integrating AI into assessment design in engineering education, several important challenges must be addressed. One is the potential overreliance on AI, which may narrow the scope of assessments. That is, AI systems are often optimised to detect patterns within predefined norms, reinforcing the precedence of traditional assessment criteria that emphasise technical skills but overlook skills such as creativity, innovation and teamwork, which are essential in engineering. Another challenge is algorithmic bias. That is, AI trained on historical data can perpetuate existing biases in educational settings, resulting in assessments that favour certain student groups or reinforce outdated norms, limiting fairness and inclusivity.

Additionally, the black-box nature of AI decision-making raises issues relating to transparency, making it difficult for educators and students to understand how specific outcomes are derived. Finally, AI's limitations in recognising complex reasoning, ethical considerations or innovative problem-solving further complicate its role in designing engineering assessments. While AI can improve efficiency, these challenges underscore the need for careful, balanced integration of AI in engineering education assessment design to ensure that it supports, rather than restricts, the broader educational objectives of engineering educational courses.

Designing a valid assessment can be partially or fully automated, depending on the criteria specified for the AI tool. In the field of engineering, automating analytical questions that originate from industry-based scenarios are the most important elements that measure students' learning. Generating such tasks requires sufficient input experience from the designer in determining the criteria of the assessment and ensuring the AI can include these criteria in the assessment it creates. Therefore, it is crucial that the selected validity techniques are aligned with the unique demands of engineering assessments, and the application of these techniques is explicitly justified to meet both academic and industry requirements. Also, the success of any AI tool used to create assessments in engineering education will depend on the level of academic ability and industry experience of the designer.

Authors

Dr. Pejman Sabet embarked on his professional journey by earning a bachelor's degree in civil engineering in 2004. He further expanded his expertise with a master's degree in 2013 and a PhD in 2020, both in Construction Management, completed through distinguished institutions in Malaysia and Australia. Boasting over a decade of industry experience, Pejman has held key roles including Site Engineer, Project Engineer, and Construction Supervisor on various international and Australian projects. He is a proud professional member of Engineers Australia and an accredited NABERS assessor specializing in energy efficiency. His hands-on industry experience has been complemented by an academic career, where he contributed to teaching and research in Civil Engineering and Construction Management at esteemed Australian institutions such as Curtin University and the Engineering Institute of Technology until 2021. Pejman has published extensively on the adoption of advanced techniques in construction projects, focusing on innovative approaches and emerging technologies that enhance efficiency, sustainability, and project outcomes within the industry. In his current role at Acknowledge Education, Pejman initially served as the National Course Coordinator for both the vocational and higher education sectors. He now serves as the Head of Department, where he continues to drive growth and innovation within both industry and academia.

Shen (Jason) Zhan holds an Honours degree in Engineering, majoring in Civil and Environmental Engineering, from the University of Auckland, New Zealand. He has several years of experience as a structural design engineer and has contributed to academia as a lecturer and tutor at the University of Auckland, specializing in construction management and engineering courses. Jason is currently a PhD student and graduate researcher at the Teaching and Learning Lab, University of Melbourne, Australia. His research focuses on exploring how universities can better support engineering students and graduates in unlocking their full employability potential.

Milad Baghalzadeh Shishehgarkhaneh is currently pursuing his PhD in Civil Engineering at Monash University, Melbourne, Australia. Concurrently, he holds lecturer positions at Acknowledge Education college's Melbourne campus. In 2024, he was recognized as a Rising Star by The Australian for his contributions to engineering and computer science. His research focuses on leveraging advanced Artificial Intelligence (AI) methodologies, notably Transformer architectures, to address challenges in construction supply chain risk management. Additionally, Milad is exploring the integration of AI with Blockchain technology to enhance resilience in construction projects. With over 30 published works, including journal articles and book chapters, he has made significant contributions to the field. His diverse research interests encompass construction supply chain management (CSCM), machine learning, natural language processing (NLP), blockchain technology, and Building Information Modelling (BIM). Through his academic endeavours, Milad aims to drive innovation in construction management practices, ensuring projects are executed efficiently and resiliently amidst evolving industry challenges. His research has been published in prestigious international journals, including Automation in Construction, IEEE Access, and Scientific Reports (Nature).

References

Ajjawi, R., Tai, J., Huu Nghia, T. L., Boud, D., Johnson, L., & Patrick, C.-J. (2020). Aligning assessment with the needs of work-integrated learning: The challenges of authentic assessment in a complex context. *Assessment & Evaluation in Higher Education*, *45*(2), 304–316. https://doi.org/10.1080/02602938.2019.1639613

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, *2*(3), 431–440. https://doi.org/10.1007/s43681-021-00096-7

Almond, R., Steinberg, L., & Mislevy, R. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *The Journal of Technology, Learning and Assessment*, *1*(5).

Altın, B., Willems, J., Oomen, T., & Barton, K. (2017). Iterative learning control of iteration-varying systems via robust update laws with experimental implementation. *Control Engineering Practice*, *62*, 36–45. https://doi.org/10.1016/j.conengprac.2017.02.005

Austin, T., Rawal, B. S., Diehl, A., & Cosme, J. (2023). AI for equity: Unpacking potential human bias in decision making in higher education. *AI, Computer Science and Robotics Technology*, *2023*(2), 1–17. https://doi.org/10.5772/acrt.20

Bagunaid, W., Chilamkurti, N., & Veeraraghavan, P. (2022). AISAR: Artificial intelligence-based student assessment and recommendation system for e-learning in big data. *Sustainability*, *14*(17), 10551. https://doi.org/10.3390/su141710551

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 1–41. https://doi.org/10.1007/s40593-021-00285-9

Barthakur, A., Joksimovic, S., Kovanovic, V., Richey, M., & Pardo, A. (2022). Aligning objectives with assessment in online courses: Integrating learning analytics and measurement theory. *Computers & Education*, *190*, 104603. https://doi.org/10.1016/j.compedu.2022.104603

Bulut, O., & Wongvorachan, T. (2022). *Feedback generation through artificial intelligence* [Paper presentation]. The Open/Technology in Education, Society, and Scholarship Association Conference. https://doi.org/10.18357/otessac.2022.2.1.125

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

Ciolacu, M., Tehrani, A. F., Binder, L., & Svasta, P. M. (2018). *Education 4.0—Artificial intelligence assisted higher education: Early recognition system with machine learning to support students' success* [Paper presentation]. 2018 IEEE 24th International Symposium for Design and Technology in Electronic Packaging (SIITME). https://doi.org/10.1109/SIITME.2018.8599203

Damaj, I., & Yousafzai, J. (2019). Effective assessment of student outcomes in computer engineering programs using a minimalistic framework. *International Journal of Engineering Education*, *35*(1), 59–75.

DeAlcala, D., Serna, I., Morales, A., Fierrez, J., & Ortega-Garcia, J. (2023). *Measuring bias in AI models: An statistical approach introducing N-Sigma* [Paper presentation]. 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC). https://doi.org/10.1109/COMPSAC57700.2023.00176

Deo, R. C., Yaseen, Z. M., Al-Ansari, N., Nguyen-Huy, T., Langlands, T. A. M., & Galligan, L. (2020). Modern artificial intelligence model development for undergraduate student performance prediction: An investigation on engineering mathematics courses. *IEEE Access*, *8*, 136697–136724. https://doi.org/10.1109/ACCESS.2020.3010938

Downing, S. M. (2003). Validity: On the meaningful interpretation of assessment data. *Medical Education*, *37*(9), 830–837. https://doi.org/10.1046/j.1365-2923.2003.01594.x

Fayoumi, A. G., & Hajjar, A. F. (2020). Advanced learning analytics in academic education: Academic performance forecasting based on an artificial neural network. *International Journal on Semantic Web and Information Systems*, *16*(3), 70–87. https://doi.org/10.4018/IJSWIS.2020070105

Fokkema, M., Iliescu, D., Greiff, S., & Ziegler, M. (2022). Machine learning and prediction in psychological assessment. *European Journal of Psychological Assessment*, 38(3), 165–175. https://doi.org/10.1027/1015-5759/a000714

Gruetzemacher, R., & Whittlestone, J. (2022). The transformative potential of artificial intelligence. *Futures*, *135*, 102884. https://doi.org/https://doi.org/10.1016/j.futures.2021.102884

Hanesworth, P., Bracken, S., & Elkington, S. (2019). A typology for a social justice approach to assessment: Learning from universal design and culturally sustaining pedagogy. *Teaching in Higher Education*, *24*(1), 98–114. https://doi.org/10.1080/13562517.2018.1465405

Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*, *2022*(1), 5215722. https://doi.org/10.1155/2022/5215722

Kizilcec, R. F., Huber, E., Papanastasiou, E. C., Cram, A., Makridis, C. A., Smolansky, A., Zeivots, S., & Raduescu, C. (2024). Perceived impact of generative AI on assessments: Comparing educator and student perspectives in Australia, Cyprus, and the United States. *Computers and Education: Artificial Intelligence*, *7*, 100269. https://doi.org/https://doi.org/10.1016/j.caeai.2024.100269

Lau, W. W., & Yuen, A. H. (2009). Predictive validity of measures of the pathfinder scaling algorithm on programming performance: Alternative assessment strategy for programming education. *Journal of Educational Computing Research*, *41*(2), 227–250. https://doi.org/10.2190/EC.41.2.e

Lo, F., Su, F., Chen, S., Qiu, J., & Du, J. (2021). *Artificial intelligence aided innovation education based on multiple intelligence* [Paper presentation]. 2021 IEEE International Conference on Artificial Intelligence, Robotics, and Communication (ICAIRC). https://doi.org/10.1109/ICAIRC52191.2021.9544874

Lodge, J., Howard, S., & Bearman, M. (2023). *Assessment reform for the age of artificial intelligence*. Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf

Memarian, B., & Doleck, T. (2024). A review of assessment for learning with artificial intelligence. *Computers in Human Behavior: Artificial Humans*, *2*(1), 100040. https://doi.org/https://doi.org/10.1016/j.chbah.2023.100040

Menekse, M. (2023). Envisioning the future of learning and teaching engineering in the artificial intelligence era: Opportunities and challenges. *Journal of Engineering Education*, *112*(3), 578–582. https://doi.org/10.1002/jee.20539

Miller, D. M., & Linn, R. L. (2000). Validation of performance-based assessments. *Applied Psychological Measurement*, *24*(4), 367–378. https://doi.org/10.1177/01466210022031813

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, *5*, 100151. https://doi.org/https://doi.org/10.1016/j.caeo.2023.100151

Myllyaho, L., Raatikainen, M., Männistö, T., Mikkonen, T., & Nurminen, J. K. (2021). Systematic literature review of validation methods for AI systems. *Journal of Systems and Software*, *181*, 111050. https://doi.org/10.1016/j.jss.2021.111050

Palmer, S., & Hall, W. (2011). An evaluation of a project-based learning initiative in engineering education. *European Journal of Engineering Education*, *36*(4), 357–365. https://doi.org/10.1080/03043797.2011.593095

Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, *4*(1), 1301013. https://doi.org/10.1080/2331186X.2017.1301013

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

Shaik, T., Tao, X., Li, Y., Dann, C., McDonald, J., Redmond, P., & Galligan, L. (2022). A review of the trends and challenges in adopting natural language processing methods for education feedback analysis. *IEEE Access*, *10*, 56720–56739. https://doi.org/10.1109/ACCESS.2022.3177752

Shishehgarkhaneh, M. B., Moehler, R. C., Fang, Y., Hijazi, A. A., & Aboutorab, H. (2024). Transformer-based named entity recognition in construction supply chain risk management in Australia. *IEEE Access*, *12*, 41829–41851. https://doi.org/10.1109/ACCESS.2024.3377232

Srinivasan, R., & Chander, A. (2021). Biases in AI systems. *Communications of the ACM*, *64*(8), 44–49. https://doi.org/10.1145/3464903

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, *3*, 100075. https://doi.org/https://doi.org/10.1016/j.caeai.2022.100075

Ting, L., Xingqiang, W., Chunhua, H., Yumin, F., Manta, O., & Yue, G. X.-G. (2023). Algorithmic framework for automated assessment and feedback of artificial intelligence (AI) technology in english intelligent teaching. *Proceedings of the 2023 8th International Conference on Intelligent Information Processing* (167–170). Association for Computing Machinery. https://doi.org/10.1145/3635175.3635205

Yeager, D. S., Fong, C. J., Lee, H. Y., & Espelage, D. L. (2015). Declines in efficacy of anti-bullying programs among older adolescents: Theory and a three-level meta-analysis. *Journal of*

Applied Developmental Psychology, *37*, 36–51. https://doi.org/10.1016/j.appdev.2014.11.005

Yüksel, N., Börklü, H. R., Sezer, H. K., & Canyurt, O. E. (2023). Review of artificial intelligence applications in engineering design perspective. *Engineering Applications of Artificial Intelligence*, *118*, 105697. https://doi.org/https://doi.org/10.1016/j.engappai.2022.105697