ChatGPT and Assessment in Higher Education: A Magic Wand or a Disruptor?

Maira Klyshbekova and Pamela Abbott
[1]Department of Digital Humanities, King's College London, UK
[2]Information School, The University of Sheffield, UK

maira.klyshbekova@kcl.ac.uk (corresponding author)

https://doi.org/10.34190/ejel.21.5.3114

Abstract: There is a current debate about the extent to which ChatGPT, a natural language AI chatbot, can disrupt processes in higher education settings. The chatbot is capable of not only answering queries in a human-like way within seconds but can also provide long tracts of texts which can be in the form of essays, emails, and coding. In this study, in the context of higher education settings, by adopting an experimental design approach, we applied ChatGPT-3 to a traditional form of assessment to determine its capabilities and limitations. Specifically, we tested its ability to produce an essay on a topic of our choice, created a rubric, and assessed the produced work in accordance with the designed rubric. We then evaluated the chatbot's work by assessing ChatGPT's application of its rubric according to a modified version of Paul's (2005) Intellectual Standards rubric. Using Christensen et al.'s (2015) framework on disruptive innovations, our study found that ChatGPT was capable of completing the set tasks competently, quickly, and easily, like a "magic wand". However, our findings also challenge the extent to which all of the ChatGPT's demonstrated capabilities can disrupt this traditional form of assessment, given that there are aspects of its construction and evaluation that the technology is not yet able to replicate as a human expert would. These limitations of the chatbot can provide us with an opportunity for addressing vulnerabilities in traditional forms of assessment in higher education that are subject to academic integrity issues posed by this form of AI. We conclude the article with implications for teachers and higher education institutions by urging them to reconsider and revisit their practices when it comes to assessment.

Keywords: ChatGPT, Higher education, Artificial intelligence (AI), Assessment, Academic integrity

1. Introduction

Given today's realia, the role of technology has become indispensable in every sector of our lives, including education. In educational settings, *Information and Communication Technologies* (*ICTs),* i.e., digital technologies that process and disseminate information for purposeful decision-making and collaboration, have proven to be efficient for both teaching and learning practices (Fernández-Gutiérrez, Gimenez and Calero, 2020). Additionally, digital technologies and software played a significant role in supporting remote learning which was accelerated by the global pandemic (Susnjak, 2022). As a tool, technology is generally seen as a supportive and assistive aid (Köprülü, 2021), especially in higher education settings.

However, a recently launched *Artificial Intelligence* (*AI*) technology called *ChatGPT* has not only captured people's attention in a short space of time but has also been able to spark heated scholarly debates. In this paper, AI refers to digital technologies that engage in reasoning and self-correction based on learning obtained through training on large knowledge bases (Kok et al., 2009). ChatGPT, launched in November 2022 by Open AI (OpenAI, 2023), has created controversy around its use and application in educational settings. For example, Zhai (2022) explored the capabilities of ChatGPT to write an academic paper and concluded that since the output was of good quality, students could now outsource their writing tasks. In another study conducted by Susnjak (2022), ChatGPT was able to produce reasonable answers during assessment with a level of accuracy that could potentially be used to disrupt academic integrity in online exams. Strikingly, GPT-3, Thunström and Steingrimsson's (2022) study queried the chatbot to write an entire academic article about itself which was then published in a journal. This study alone confirmed how little human input and involvement this AI requires which makes us question its long-term effect in disrupting academic integrity and jeopardising the roles of educators.

In order to address these concerns, it is important to investigate not only the capabilities of ChatGPT but to also understand its limitations. To date, many studies have focused on the chatbot's potential to disrupt educational processes (Zhai, 2022), but there is still a dearth of research in identifying its gaps specifically in educational settings. Since the chatbot may be used to replicate an act of assessment between students and educators, we believe it is important to understand to what extent this replication impacts traditional processes in Higher Education settings. Considering these points in this study, we evaluated ChatGPT's performance by tasking it with academic essay writing in order to identify its capabilities and limitations specifically in higher education

Cite this article: Klyshbekova, M and Pamela Abbott, P. 2024, "ChatGPT and Assessment in Higher Education: A Magic Wand or a Disruptor?", *Electronic Journal of e-Learning*, 22(2), pp 30-45, https://doi.org/10.34190/ejel.21.5.3114

settings. The essay as a traditional form of assessment is already critiqued as approaching obsolescence due to the availability of much openly accessible knowledge that can be easily copied and the difficulty in achieving consistency and reliability in marking it (Race, 2018).

Taking into consideration all the above concerns, the study was thus guided by the following three research questions:

Q1: To what extent can ChatGPT replicate the process of delivering an academic essay on a topic created at random?

Q2: To what extent can ChatGPT effectively assess the essay created by itself?

Q3: With reference to the findings from Q1 and Q2, to what extent can ChatGPT be considered a disruptor?

The paper is expected to contribute to the sparse literature on the implications of the use of ChatGPT on higher education assessment, an issue which is receiving increasing attention in academia (Kasneci et al., 2023) and in educational practice (Stokel-Walker, 2022; Rudolph, Tan and Tan, 2023). It brings a unique perspective on the AI's capabilities and limitations as a disruptor in higher education assessment.

2. Literature Review

Studies about the role of technology in higher education have reported that it can enhance the learning process and impact student engagement (Bond et al., 2020) and improve communication between educators and students (Akour and Alenezi, 2022). Menkhoff and colleagues (2015), for example, reported that students found writing entries on Twitter a more interactive experience than traditional lectures. In a different study, Oliva-Córdova, Garcia-Cabot and Amado-Salvatierra (2021) highlighted that the use of technology as a pedagogical tool can improve the learner experience. Other studies highlighted the capacity of technological tools to create dynamic, creative, and learner-centered spaces (Fojtik, 2014).

Recently, rapid changes in the role of technology in higher education occurred during the *emergency remote teaching* (*ERT*) period, defined as a "putatively ephemeral shift to remote teaching to continue teaching and learning during emergencies" (Sum and Oancea, 2022, p. 1). The pandemic was a turning point for elevating the role of technology in schools and in higher education settings, necessitating educators to shift to more technology use. Technology became an integral part of teaching and a mediator between educators and students who connected through online learning platforms.

Notwithstanding the benefits of remote teaching (Susnjak, 2022), technological challenges were also evident such as bad Internet connection (Sum and Oancea, 2022), lack of resources and institutional support (Joshi, Vinay and Bhaskar, 2020), and the issue of sustaining academic integrity in online exams (Susnjak, 2022). Online examinations posed a serious challenge due to the lack of direct supervision and "the ease with which students may be able to access and share resources during the exam" (Susnjak, 2022, p. 2). It seems that this online mode of assessment became particularly prone to cheating (Arnold, 2016) even more so than traditional exams. In order to prevent the growth of cheating in online exams, institutions began to utilise relevant technological tools including software such as proctoring and plagiarism detection (Susnjak, 2022). Educators were encouraged to move away from multiple-choice questions and use tasks that involved critical thinking and creativity (Whisenhunt et al, 2022). Due to these issues, emergency remote teaching was seen as a *disruptor* by many (Sum and Oancea, 2022), a phenomenon further discussed in the following section.

2.1 Disruptive Innovation

Regardless of the mentioned advantages and benefits, technology in education can also be labeled as disruptive, especially non-institutional technologies in higher education (Flavin, 2012). The term disruptive technology was coined by Christensen (1997) to identify easy-to-use technologies with the potential to displace more incumbent technologies. Disruptive technology can be described as changing

> *"traditional practices usually starting with a small number of users and then growing over time in such a way that it displaces a well-established and prominent practice" (Siddhpura and Siddhpura, 2020, p. 494).*

The term "disruptive technology" was later changed to "disruptive innovation" as a way of extending it and including other sectors such as service sectors (Christensen and Raynor, 2003). The theory of disruptive innovation can be applied to many sectors including higher education (Flavin, 2021). Christensen and Raynor (2003) propose two types of disruption which are "new market disruption" and "low-end disruption". New

market disruption can be understood as establishing a new market that meets the needs of the customers which previously were unmet (Christensen and Raynor, 2003). A low-end disruption can be understood as the use of new technologies to exceed the performance of the existing market (Christensen and Raynor, 2003). Disruptive innovations can be contrasted with sustaining and efficiency innovations that merely improve existing goods and enhance processes respectively, rather than bring about transformative change (Christensen, Raynor and Mcdonald, 2015).

In higher education contexts, Al-Imarah and Shields (2016) propose three characteristics of disruptive innovation, performance, benefits, and market. New disruptive innovations initially do not perform well but they have a tendency to improve over time resulting in the attraction of mainstream customers (Al-Imarah and Shields, 2016). The benefits of disruptive innovations consist of low prices, convenience, and ease of use. Regarding the market, disruptive innovations initially have limited customers, but once they develop their performance, they can create new markets and become competitors for existing mainstream goods and services (Christensen, Raynor and Mcdonald, 2015). In Al-Imarah and Shields's (2016) study, they explored Massive Open Online Courses' (MOOCs) potential to disrupt existing models of higher education. Their analyses, using these three characteristics, revealed that MOOCs only supported one characteristic of disruptive innovation, i.e., they have the possibility of creating a new market by targeting part-time students, distance learners, and self-directed students (Al-Imarah and Shields, 2016). They concluded that MOOCs can therefore only be characterised as sustainable innovation.

Although disruptive innovation theory has been critiqued for its inability to predict an innovation's disruptive potential (Flavin, 2016), it still remains a useful framework for analysing new technologies in higher education settings, particularly to learn "why some succeed and some fail" (Flavin, 2021, p. 3). Flavin (2021) argued that disruptive innovations offer cheap and easy-to-use technologies that first attract a particular market and later target mainstream ones as well. This can be seen, for example, in the case of budget airlines disrupting incumbent airlines (Kumar, 2006) or car rentals disrupting dominant car brands (Markides and Sosa, 2013). Similarly, in higher education, the emergence of ChatGPT is attracting heated debates about its disruptive potential. Some are identifying it as a revolutionary technology that can improve higher education processes (Fauzi et al., 2023) whereas others seem to be thinking differently (Kasneci et al., 2023), a debate which is discussed further below.

2.2 ChatGPT

ChatGPT is a *large language model* (*LLM*) developed by OpenAI to engage with and respond to users' prompts and questions. LLMs are specialised AI technologies that use natural language processing (NLP) to generate human-like text and complete a variety of language-related tasks (Kasneci et al., 2023). ChatGPT is built upon the LLM technology, Generative Pretrained Transformer. ChatGPT, also known as a chatbot, an application that mimics conversations with humans, responds to queries within seconds (Rudolph, Tan and Tan, 2023) and was trained on a wide variety of texts including academic articles, books, and websites covering themes and topics from science, fiction, and news reports (Shen et al., 2023). It is also trained through human feedback (Zhai, 2022), enabling it to expand its corpus and areas of knowledge over time.

ChatGPT-3 was released 30th of November, 2022 (Tate et al., 2022), and quickly attracted attention due to several reasons. Firstly, the chatbot was made available to the public through setting up a free account. Secondly, the interface of the chatbot is user-friendly, in the form of a conversation (Tate et al., 2022). Thirdly, the chatbot creates a sense of interaction, taking into account previous prompts and questions. However, it is worth noting that in the process of preparing this paper, OpenAI released a new version of the chatbot on 1st of February 2023 which operates on a paid subscription model and offers access to new updates, faster response times, and quicker access even at peak times (OpenAI, 2023).

Several studies investigating ChatGPT's capabilities have highlighted its remarkable ability to produce a fluent piece of writing on a wide range of topics (Shen et al., 2023; Tate et al., 2022). For instance, Tate et al. (2022) noted that besides giving answers to factual questions the AI is also capable of writing essays, poems, reports, plays, and stories on "almost any subject described; writing a critique of that same text from the point of view of a teacher, professor or literary scholar" (pp. 4-5). Pavlik's (2023) study also confirmed this after tasking the chatbot with factual questions, demonstrating that ChatGPT was not only capable of generating high-quality written texts but also of presenting these texts in a manner similar to that of a human. Zhai (2022) reported that it took only a few hours to write and finish a study using ChatGPT and required little human input. Terwiesch (2023) mentioned that exam development normally takes around 20 hours and then another 10 for the Teaching Assistant to test it and prepare solutions. But with the help of ChatGPT, it was possible to develop the exam in

10 hours and then cut the Teaching Assistant's time from 10 to 5 hours (Terwiesch, 2023). Although in Zhai's (2022) and Terwiesch's (2023) studies, it is evident that the chatbot increases productivity and efficiency, it also has the potential to put roles at risk such as that of Teaching Assistants.

In this vein, the chatbot has the potential to emulate human endeavour by, for example, co-authoring a journal article. For example, (O'Connor and ChatGPT, 2023) is "co-authored" by ChatGPT, thus raising ethical issues around authorship of academic outputs and academic integrity.

In defence of academic integrity, O'Connor and ChatGPT (2023) argued that the chatbot could be used to combat plagiarism by "providing students with tools and resources that can help them properly cite" (p. 1) the needed sources. However, Qadir's (2022) study reported that the chatbot generated references with non-functional links and even ones that did not exist. Aydin and Karaarslan's (2022) study also explored ChatGPT's ability to avoid plagiarism by asking the chatbot to paraphrase academic articles' abstracts and then check the paraphrased abstracts for plagiarism. They found that the match rates of ChatGPT's paraphrased abstracts were high which made them conclude that the chatbot may not be able to successfully elude plagiarism detection.

Both Pavlik's (2023) and Aydin and Karaarslan's (2022) studies urged educators to consider the role of ChatGPT and its impacts on academic integrity in higher education. ChatGPT is likely a forerunner of many similar LLMs (Tate et al., 2022), and the number of users who will "consult" or "use" these AIs will eventually grow making it even more important to identify how this technology will influence higher education settings.

3. Methodology

This study adopts an experimental design approach where we tested ChatGPT-3's performance and its ability to respond to a set of fixed queries. To investigate the extent to which ChatGPT can influence the processes of assessment in higher education, we selected the essay, a traditional form of assessment, which nominally consists of these processes: design, delivery, and marking.

Building upon prior studies (Susnjak, 2022; Zhai, 2022), we designed *five specific queries* which tested ChatGPT's ability to: write an essay on a specific topic, using a specific referencing style (*design the essay*); write an introduction, body, and conclusion of the essay (*deliver the essay*); and also design a rubric and assess its own work in accordance with this rubric (*mark the essay*). The queries were conducted on ChatGPT's main website, chat.openai.com. The entire period of testing and analysis lasted from February until mid-March 2023. We documented our observations of ChatGPT's responses to the queries related to design and delivery and conducted our independent assessment of its marking by using Paul's (2005) Intellectual Standards rubric. To analyse ChatGPT's disruptive capacity, we applied the disruptive innovation lenses from the literature (Christensen, Raynor and Mcdonald, 2015; Al-Imarah and Shields, 2019; Flavin, 2021).

The following steps constitute the experimental research design:

Step 1. Firstly, we set up an account at Open AI which granted free-to-use public access to ChatGPT. The ChatGPT version that we used was released on January 30th, 2023, which had improved features over the previous version of January 9th, 2023, improving the chatbot's response on a wide range of topics, including an update in factuality and mathematical problem-solving. Once the account was set up, the main interface of the website provided examples, capabilities, and limitations of the chatbot (See Fig. 1).

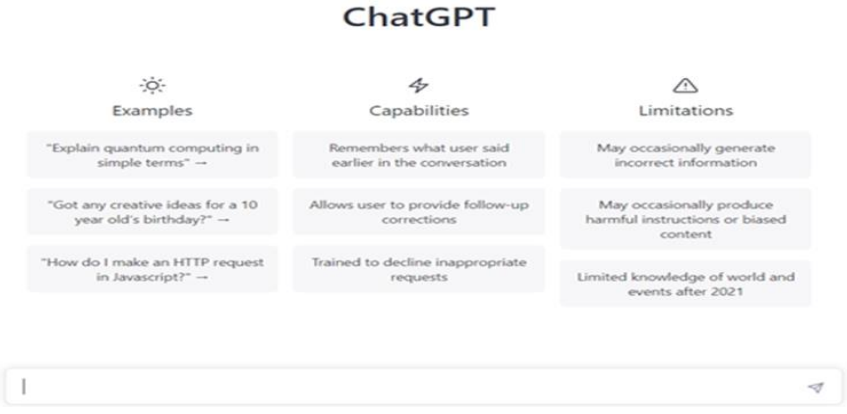


Figure 1: ChatGPT interface

Step 2. Secondly, by opening the new chat section we tasked the chatbot to write an academic essay on Technology in Education using three queries. The **first query** was: *Write an introduction on the topic of Technology in Education with the context and aims of the article. State your argument and also indicate your answer to the argument. Write with 5 references in APA style*.

After the chatbot produced an answer, we posed the **second query**: *Try to develop your argument point by point. Mention the reasons that support your argument that were stated in the introduction. Use 5 different reference sources and write the essay with 500 words.* The chatbot produced the body of the essay containing five paragraphs. We posed the **third query** which was on the last part of the essay, the conclusion: *Summarise the reasons that support the mentioned argument without including any new information. Remind the reader of the points you mentioned and how you addressed the posed question.*

Step 3. After the chatbot produced all three parts of the essay, we tasked it with two further queries related to marking. The **fourth query** was: *Design a rubric to rate this essay*. The chatbot produced a points-based rubric with five criteria. These criteria were introduction, development of argument, summary and conclusion, presentation, creativity and originality. Once the rubric was produced, we asked the chatbot to rate the essay in accordance with its rubric, by using the following **fifth query**: *Rate this essay in accordance with the rubric you have designed.*

Step 4. After ChatGPT rated its own work, we independently evaluated the chatbot's essay using a modified version of Paul's (2005) Intellectual Standards rubric. Paul's (2005) rubric includes criteria such as clarity, accuracy, precision, relevance, depth, breadth, logic, significance, fairness. We further adapted the rubric with the additional criterion of "originality and creativity". Thus, the modified rubric contains the following criteria:

1. Clarity (the text is clear and easy to understand)
2. Accuracy (the information in the text is provided accurately)
3. Precision (the text is precise and provides enough details)
4. Relevance (the text is relevant to the topic and/or identified issue)
5. Depth (the text provides an in-depth analysis)
6. Breadth (the text considers other perspectives and/or viewpoints)
7. Logic (the text is presented in a logical manner)
8. Significance (the text reminds the reader of the significance of the topic)
9. Fairness (the text represents the viewpoints in a fair manner)
10. Originality and Creativity (the text demonstrates creative and original ideas).

Step 5. We analysed ChatGPT's disruptive capacity by applying the disruptive innovation lenses from the literature, which state that technologies are disruptive to the extent that they satisfy two main criteria, (a) provide an easy-to-use, cost-efficient, convenient, and simple alternative to the status quo and (b) when put to use significantly alter existing processes (Christensen, Raynor and Mcdonald, 2015; Al-Imarah and Shields, 2019; Flavin, 2021).

Step 6. We then decided to task ChatGPT to assess its essay again but this time using the rubric that we have designed in Step 4 to avoid the potential bias of the chatbot awarding itself overly favourable marks through the use of its own rubric. We could thus compare its use of a rubric independent from its own and contrast it with our assessment.

In the following sections, the results of the queries are presented and discussed.

4. Findings

In this section, we demonstrate how we applied the experimental research design to address the three guiding research questions, which serve as subheadings for each subsection below.

4.1 To What Extent can ChatGPT Replicate the Process of Delivering an Academic Essay on a Topic Created at Random?

From the responses to our queries 1, 2, and 3, it is evident that ChatGPT is capable of producing an essay on a specific topic, created at random, however, as can be seen from the query response below, the content of the essay that the chatbot generated was rather disappointing (Figs. 2-4). The chatbot created a very general and descriptive essay and in some cases, it was quite repetitive. For example, the generated text was quite restricted and did not incorporate wide-ranging perspectives as only five sources were consulted. Interestingly, all of these sources were from the years 2001, 2002, 2005, 2010, and 2011, with no recent references included (See Fig.4).

Even though the chatbot was able to produce the essay within one minute, the content itself was not remarkable and lacked creativity. For example, the generated text did not incorporate any innovative ideas or perspectives nor did it produce any original thinking. Moreover, there were inconsistencies in the chatbot's reply to the queries since the introduction and main body, for which the chatbot was tasked to provide 500 words each, only generated 258 and 287 words (See Figs 2-3).

In both the introduction and main body sections, we asked the chatbot to provide five different references in APA style. Although the chatbot managed to provide the requested style correctly, it was still not capable of providing five different references (see Fig 2 below). Both the introduction and main body sections had identical reference sources all of which were given only as in-text citations. Upon looking into the generated reference list (Fig 4) which contained only five sources we realised that none of the provided links were functional. In fact, two of the sources turned out to be bogus references, one reference's title was not correct and another reference's author name was given as Turkly (2011) instead of Turkle (2011).

Technology has revolutionized the way we live, work, and learn. In the field of education, technology has brought about tremendous opportunities and benefits, but also some challenges. The aim of this essay is to explore the impact of technology on education and to analyze its potential benefits and drawbacks.

The argument of this essay is that technology can greatly enhance the learning experience and promote student success, but it must be approached in a strategic and responsible manner, taking into consideration the needs and goals of the education system and its stakeholders. While technology has the potential to increase student engagement, motivation, and academic performance (Warschauer & Matuchniak, 2010), it can also lead to reduced face-to-face interaction, decreased social skills, and increased anxiety, depression, and other mental health problems (Turkle, 2011).

To answer this question, this essay will first provide a general overview of the integration of technology in education, including its benefits and challenges. Next, it will examine the role of technology in student-centered learning and collaboration (Roschelle & Pea, 2002) and in teacher professional development (Swan & Shen, 2005). Finally, the essay will consider the importance of ensuring equitable access to technology for all students (Cuban, 2001) and of approaching technology integration in a thoughtful and strategic manner.

In conclusion, technology in education can greatly enhance the learning experience and promote student success, but it must be approached in a responsible and strategic manner. By considering the benefits and challenges of technology integration, education systems can work to ensure that technology is used to support student learning and well-being.

Figure 2: Outcome related to query 1. Word count = 258

In support of the argument that technology can greatly enhance the learning experience and promote student success, there are several reasons to consider. First, technology can increase student engagement and motivation. The use of technology in the classroom can provide students with access to a wide range of interactive and multimedia resources that can help to increase their interest and understanding of the subject matter (Warschauer & Matuchniak, 2010). Furthermore, technology can allow for personalized and differentiated instruction, meeting the diverse learning needs of individual students (Cuban, 2001). Second, technology can facilitate student collaboration and teamwork. The use of communication and collaboration tools such as online forums, video conferencing, and social media can provide students with opportunities to work together on projects, share ideas, and receive feedback from peers and teachers (Roschelle & Pea, 2002). This can help to build strong relationships among students and increase their sense of community and belonging. Third, technology can support teacher professional development. The integration of technology in teacher education programs can provide teachers with opportunities to learn about the latest instructional strategies and tools, as well as to collaborate with peers and engage in ongoing professional learning (Swan & Shen, 2005). This can help to ensure that teachers are well-equipped to use technology effectively in their classrooms. However, while technology can provide numerous benefits for students and teachers, it is important to approach its integration in a responsible and strategic manner. For example, it is crucial to ensure equitable access to technology for all students, regardless of their socio-economic background or geographic location (Cuban, 2001). Additionally, it is important to be aware of the negative effects of technology on student well-being, such as decreased face-to-face interaction and increased anxiety and depression (Turkle, 2011).

Figure 3: Outcome related to query 2. Word count = 287

In summary, the integration of technology in education has the potential to greatly enhance the learning experience and promote student success by increasing student engagement and motivation, facilitating collaboration and teamwork, and supporting teacher professional development. However, it is crucial to approach technology integration in a responsible and strategic manner, taking into account the need for equitable access to technology and the potential negative effects on student well-being. The argument in this essay supports the idea that technology can have a positive impact on education when used in a thoughtful and strategic manner, and highlights the importance of considering both the benefits and challenges of technology integration.

References:

Cuban, L. (2001). Oversold and underused: Computers in the classroom. Cambridge, MA: Harvard University Press.

Roschelle, J., & Pea, R. D. (2002). Making sense of argumentation and communication technologies in school mathematics. Journal of Research in Mathematics Education, 33(2), 137-163. https://doi.org/10.2307/749763

Swan, K., & Shen, J. (2005). The integration of technology into teacher education programs: A review of the literature. Journal of Technology and Teacher Education, 13(3), 571-596. https://doi.org/10.1080/10705630500264956

Turkly, S. (2011). Alone together: Why we expect more from technology and less from each other. New York, NY: Basic Books.

Warschauer, M., & Matuchniak, T. (2010). New technologies and learning. Annual Review of Applied Linguistics, 30, 276-292. https://doi.org/10.1017/S0267190510000070

Figure 4: Outcome related to query 3

4.2 To What Extent can ChatGPT Effectively Assess the Essay Created by Itself?

From the responses to our queries 4 and 5, it is evident that ChatGPT is capable of designing a bespoke rubric to mark the essay it produced and then rate its work in accordance with this rubric. However, with respect to the query responses the following issues were revealed. Even though the chatbot generated a generic rubric to assess the essay, we found some issues with the assessment itself. In our first attempt at asking the chatbot to rate the essay (see Fig 6), it marked the essay at 91 out of 100 points. We thought that the score was too high given the quality issues already noted, thus, we decided to ask the chatbot to rate the work again (see Fig 7) and surprisingly the second time the score was different. This time the chatbot rated the work at 93 out of 100 points, higher than in the previous attempt. Finally, on a third attempt at rating the essay (see Fig 8), the chatbot's score this time was 88 out of 100. One of the chatbot's comments on this score was "*The essay could benefit from further innovative ideas and perspectives to increase its creativity and originality*" with which we

surely agreed. However, the three attempts generated three different scores on the same essay which made us yet again question the chatbot's credibility and reliability.

In addition to the rating problems, we also found some issues with the designed rubric and calculations. ChatGPT's rubric was simplistic and generic, with only five rating criteria totalling 100 points (see Fig 5). Each of the criteria was weighted differently, with no clear logic evident for the assignment of weightings. Development of Argument (40 points) and Summary and Conclusion (20 points) were weighted the most even though the latter more or less summarised points made in the argument of the essay, introducing no new information, but achieving half the score of the former. The Introduction (10 points) and Presentation (10 points) categories were equally weighted although one category is about content while the other is about the look and feel of the work; it is unclear why they would attract the same weighting. Interestingly, Creativity and Originality were also given only 10 points, but this aspect is deemed of critical importance in constructing an essay (Bekurs and Santoli, 2004). Additionally, the total of the rubric's weightings was 90 not 100 points. However, ChatGPT presents it as a 100-point marking scale and even assessed the essay using these false weightings. It seems that the chatbot may have misapplied the rubric by allocating double the amount of marks for some categories while giving less for others. This demonstrates the chatbot's incompetence both in creating the rubric and fairness in allocating the weightings.

From our own independent assessment of the chatbot's generated essay, we gave it a total score of 41 out of 100 (See Fig. 9). Although we thought the work was written in a logically relevant, understandable, and easy-to-read manner we still found many issues. The content was very generic, and simple and did not demonstrate any in-depth analysis. For example, the essay did not conduct a thorough examination of the impact of technology on education and did not provide any detailed examples. The arguments were not backed up by evidence and the provided sources were inaccurate. For instance, there were five sources in total and none of the provided links were functional. The text failed to illustrate any original and creative thinking and did not integrate any innovative ideas or perspectives. For example, the text highlights the importance of considering both the benefits and challenges of technology integration but does not engage in discussing the ways in which teachers can creatively integrate technology into their practices. Notably, ChatGPT's assessment using our rubric resulted once more in a quite generous score of 88/100 (see Fig 10).

1. Introduction (10 points)

- *Clarity of the context and aim of the essay (3 points)*
- *Clearly stated argument (3 points)*
- *Explanation of how the question will be answered (2 points)*
- *Relevance to the topic of technology in education (2 points)*

2. Development of argument (40 points)

- *Coherent and logical flow of ideas (10 points)*
- *Evidence-based support for argument (10 points)*
- *Use of relevant and credible sources (5 points)*
- *Integration of personal insights and perspectives (5 points)*
- *Consideration of alternative viewpoints (5 points)*
- *Clear and concise language (5 points)*

3. Summary and conclusion (20 points)

- *Recap of the main points (5 points)*
- *Restatement of the argument (5 points)*
- *Clear and concise language (5 points)*
- *Reflection on the implications of the argument (5 points)*

4. Presentation (10 points)

- *Clarity of organization and structure (3 points)*
- *Correct use of grammar, spelling, and punctuation (3 points)*
- *Effective use of visuals and formatting (2 points)*
- *Adherence to the required citation style (2 points)*

5. Creativity and originality (10 points)

- *Use of innovative ideas and perspectives (5 points)*
- *Integration of personal experiences and insights (2 points)*
- *Reflection on the broader implications and application of the argument (3 points)*

Total: 100 points

Figure 5: Outcome related to query 4 – ChatGPT's generated rubric

Introduction (10 points) - 9/10

The introduction provides a clear context for the essay and states the argument well. It also explains the aim and approach of the essay.

Development of Argument (40 points) - 36/40

The essay provides a comprehensive examination of the impact of technology on education, including its benefits and challenges. The argument is well-supported with evidence and credible sources. The personal insights and perspectives are integrated effectively into the argument. The essay also considers alternative viewpoints. The language is clear and concise.

Summary and Conclusion (20 points) - 18/20

The summary and conclusion provide a clear recap of the main points and restate the argument effectively. The language is clear and concise. The reflection on the implications of the argument is well done.

Presentation (10 points) - 9/10

The essay is well-organized and structured. The use of grammar, spelling, and punctuation is correct. The visuals and formatting are effective. The citation style is adhered to correctly.

Creativity and Originality (10 points) - 9/10

The essay provides innovative ideas and perspectives, integrates personal experiences and insights effectively, and reflects on the broader implications and applications of the argument well.

Overall, the essay is well written and explores the impact of technology on education in a comprehensive and thoughtful manner. The argument is well-supported and the presentation is effective. A total score of 91/100 is earned.

Figure 6: Outcome related to query 5 – ChatGPT's rating of the essay. (Iteration 1). A total score of 91/100 (note. ChatGPT's incorrect calculation)

Introduction (10 points) - 9/10

The introduction provides clarity of the context and aim of the essay. It also explains how the question will be answered.

Development of Argument (40 points) - 37/40

The essay provides a coherent and logical flow of ideas. The argument is supported with relevant and credible sources. The essay integrates personal insights and perspectives, but at the same time considers alternative viewpoints.

Summary and Conclusion (20 points) - 18/20

The conclusion provides a recap of the main points and restates the argument. The presented language is clear and concise. The implications of the argument are reflected effectively.

Presentation (10 points) - 9/10

The structure and organization of the essay is efficient. The use of grammar, spelling, and punctuation is effective. Both the visuals and formatting are effective too. The required citation style is presented correctly.

Creativity and Originality (10 points) - 10/10

The essay is innovative and provides different perspectives. It not only integrates insights and experiences, but also provides a reflection on the argument application and implications.

Total: 93/100 points

Overall, the essay provides a clear and well-supported argument about the benefits and challenges of technology integration in education. The essay effectively integrates evidence from relevant and credible sources and provides personal insights and perspectives on the topic. The organization and structure of the essay are clear and effective, and the language is generally clear and concise. The essay shows creativity and originality by reflecting on the broader implications and application of the argument. However, there are minor issues with the clarity of the argument in the introduction, and some minor errors in grammar, spelling, and punctuation throughout the essay.

Figure 7: Outcome related to query 5 – ChatGPT's rating of the essay. Iteration 2. A total score of 93/100 (note. ChatGPT's incorrect calculation)

Introduction (10 points) - 9/10

The context for the essay and the argument is clear and effective. The essay explains the relevance of the topic its aim and its approach.

Development of Argument (40 points) - 40/40

The essay is coherent and provides a logical flow of ideas. The essay uses credible sources and integrates alternative viewpoints. The language of the essay is clear and concise. The arguments are supported by the evidence.

Summary and Conclusion (20 points) - 20/20

The summary and conclusion provide a strong recap of the main points reflection on the implications of the argument.

Presentation (10 points) - 10/10

The presentation is well-organized and adheres to the required citation style. The use of grammar, spelling, and punctuation is correct. The essay uses effective visuals and formatting.

Creativity and Originality (10 points) - 8/10

The essay does not provide innovative ideas and perspectives. The integration of personal experiences and insights is not effective enough. The reflection on the broader implications and applications of the argument is not well enough.

Overall, I would rate this essay as 88/100. The essay effectively presents a clear and well-supported argument, with a coherent and logical flow of ideas and use of relevant and credible sources. The language is clear and concise, and the essay effectively considers alternative viewpoints. The summary and conclusion provide a strong recap and reflection on the implications of the argument. The presentation is well-organized and adheres to the required citation style. However, the essay could benefit from further innovative ideas and perspectives to increase its creativity and originality.

Figure 8: Outcome related to query 5 – ChatGPT's rating of the essay. Iteration 3. A total score of 88/100 (note. ChatGPT's incorrect calculation)

Clarity (10 points) - 7/10. The text was very simple and easy to read. The meaning of the text is understandable and easy to grasp.

Accuracy (10 points) - 3/10. The sources given in the text were made up and not accurate. Thus making the alleged facts unreliable.

Precision (10 points) - 5/10. The text provides generic information without giving enough details. Alleged facts could have been more specific.

Relevance (10 points) - 6/10. The text was relevant to the topic but it did not identify and cite the key authors' works. This makes us assume that the knowledge in specific areas is still weak and the bot is unable to demonstrate broad reading.

Depth (10 points) - 3/10. The text did not provide an in-depth analysis. The generated answers were constrained only to 258 and 287 even though we requested to write 500 words for each of the two sections.

Breadth (10 points) - 3/10. The text was rather narrow-minded and did not incorporate wide-ranging perspectives and multiple viewpoints. In all sections, the same sources were used repeatedly and new ones were not incorporated.

Logic (10 points) - 6/10. The text is presented in a simplistic but logically relevant way.

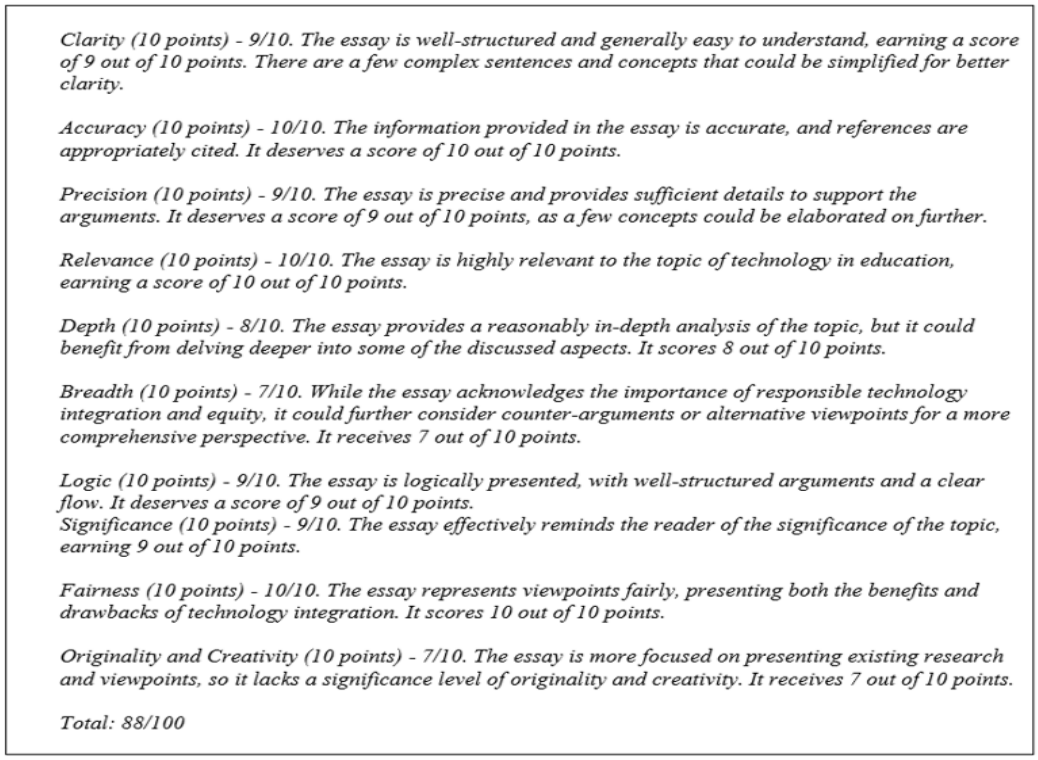
Significance (10 points) - 6/10. The text reminds the reader about the significance of the topic but in a rather simplistic manner.

Fairness (10 points) - 1/10. The text provided made-up references. The bot uses the data to which only it has access and which is not independently verifiable, which potentially puts fairness into question.

Originality and creativity (10 points) - 1/10. The text failed to demonstrate originality and creativity. It was quite repetitive and generic and did not contain any novel or original ideas.

Total: 41/100

Figure 9: Researchers' independent assessment of ChatGPT's essay. *A total score of 41/100.*



Clarity (10 points) - 9/10. The essay is well-structured and generally easy to understand, earning a score of 9 out of 10 points. There are a few complex sentences and concepts that could be simplified for better clarity.

Accuracy (10 points) - 10/10. The information provided in the essay is accurate, and references are appropriately cited. It deserves a score of 10 out of 10 points.

Precision (10 points) - 9/10. The essay is precise and provides sufficient details to support the arguments. It deserves a score of 9 out of 10 points, as a few concepts could be elaborated on further.

Relevance (10 points) - 10/10. The essay is highly relevant to the topic of technology in education, earning a score of 10 out of 10 points.

Depth (10 points) - 8/10. The essay provides a reasonably in-depth analysis of the topic, but it could benefit from delving deeper into some of the discussed aspects. It scores 8 out of 10 points.

Breadth (10 points) - 7/10. While the essay acknowledges the importance of responsible technology integration and equity, it could further consider counter-arguments or alternative viewpoints for a more comprehensive perspective. It receives 7 out of 10 points.

Logic (10 points) - 9/10. The essay is logically presented, with well-structured arguments and a clear flow. It deserves a score of 9 out of 10 points.
Significance (10 points) - 9/10. The essay effectively reminds the reader of the significance of the topic, earning 9 out of 10 points.

Fairness (10 points) - 10/10. The essay represents viewpoints fairly, presenting both the benefits and drawbacks of technology integration. It scores 10 out of 10 points.

Originality and Creativity (10 points) - 7/10. The essay is more focused on presenting existing research and viewpoints, so it lacks a significance level of originality and creativity. It receives 7 out of 10 points.

Total: 88/100

Figure 10: ChatGPT's assessment in accordance with our designed rubric. A total score of 88/100.

4.3 To What Extent can ChatGPT be considered a Disruptor?

With respect to the processes of producing an essay, all of the queries submitted to the chatbot were completed within a short period of time. The full essay, rubric, and marking were completed by the chatbot in approximately one minute. The application is easy to use with a user-friendly, easy-to-navigate interface (see Fig 1) and can be accessed through a simple URL on the Web (openai.com). No costs were associated with its use. With respect to each of the processes of design, delivery, and marking of the essay assessment, we made the following observations. Regarding the design, although the chatbot executed the essay design, it was flawed as noted in the findings above. Regarding the delivery, our own independent assessment of the outcome resulted in a very mediocre grade due mainly to lapses in accuracy, depth, breadth, fairness, originality, and creativity. With regard to the marking, our observations and independent assessment revealed serious flaws as outlined in the previous section, i.e., a flawed rubric and inconsistent marking. Thus, when applying the disruptive innovation lens, we find that ChatGPT satisfied the first criterion, i.e., that it was fast, convenient, easy to use, and cost-efficient, however, with respect to the second criterion, its effect on the existing processes was inconclusive. While it produced plausible outcomes in response to the queries, those outcomes, upon closer examination were questionable. The implications of this finding for ChatGPT's potential to disrupt markets in higher education will be further discussed in the following section.

5. Discussion

Drawing on the findings, we believe that ChatGPT has not reached its exceptional levels yet. Although we agree with Shen et al.'s (2023) findings that it demonstrated a number of impressive capabilities, we still believe that the chatbot is far from being "exceptional" let alone be trusted with work for academic purposes. The discussion of the findings is structured according to the research questions in the sections that follow.

5.1 To What Extent can ChatGPT Replicate the Process of Delivering an Academic Essay on a Topic Created at Random?

Although the results of this study indicated that ChatGPT was capable of delivering an academic essay chosen at random it still has not succeeded in producing efficient work. On one hand, the chatbot was quick in responding to the queries and in fact, was able to produce written work in less than one minute's time, which may be favored

by many, especially by those users who are looking for a faster way to accomplish their tasks. On the other hand, it is important to note that as per our independent assessment of ChatGPT's produced work, we found it to be of a quite mediocre level. There were many inconsistencies in the chatbot's replies to our queries and these are discussed below.

First, as noted in the findings, it became clear that ChatGPT was not capable of producing a written piece of work with the requested number of words. Similar findings have also been reported by Rudolph, Tan and Tan's (2023) study. In their study, they tasked the chatbot to write a 2000-word essay but the chatbot was only capable of producing 500 words regardless of repeated attempts.

Second, the chatbot could not produce the needed results for referencing and identifying key scholars' works. The chatbot kept using the same five sources across different sections and did not incorporate any recent up-to-date sources. Upon examination, it became clear that none of the source links were functional and in fact, two of them turned out to be bogus. Additionally, one of the provided reference's titles was incorrect and another reference contained a mistake in the author's name. These findings resonate with Farrokhnia et al. (2023), Rudolph, Tan and Tan (2023), and Qadir (2022) who also reported that the chatbot generated made-up reference lists with non-functional links. In the cases when the users prompted the chatbot to provide an up-to-date reference list, it fabricated non-existent sources (Farrokhnia et al., 2023). Our findings in this study also corroborate that the chatbot is still not capable of producing reliable and up-to-date references.

Third, the content produced by ChatGPT was rather disappointing when we independently assessed it. The essay was primarily descriptive and did not provide any evidence to back up its arguments and statements. It lacked original thinking and was quite generic. This finding is in line with Pavlik (2023), Rudolph et al. (2023), and Tlili et al. (2023).

5.2 To What Extent can ChatGPT Effectively Assess the Essay Created by Itself?

Although it took less than one minute for the chatbot to rate the work, the results showed that it is still not capable of critical thinking and assessment, a finding opposite to Susnjak's (2022) study, which claimed that ChatGPT demonstrated critical thinking skills and was capable of generating high-quality text that was hard to distinguish from humans' work. Additionally, although the essay is seen by some as problematic, it is still one of the few vehicles for assessment that allows for creative and critical thinking (Bekurs and Santoli, 2004). We have determined that ChatGPT, while able to perform well on many of the criteria used to examine intellectual ability (Paul, 2005), would struggle to master critical thinking dimensions such as significance and fairness and more intangible ones such as originality and creativity. AI learning modules are only as good as the data on which they are trained and creativity requires divergence rather than convergence of ideas (Rudolph, Tan and Tan, 2023). As for the marking aspect of the chatbot-generated assessment exercise, the errors noted with the rubric and its application suggest that this part of the process is underdeveloped and subject to spurious outcomes, hence not reliable (Kabir et al., 2023). As illustrated in the findings, the chatbot was unable both to create a rubric and fairly assess the work in accordance with the rubric Additionally, the chatbot demonstrated inconsistency in its evaluation of the work. While it is acknowledged that assessment by humans can also be inconsistent and biased (Hanesworth, Bracken and Elkington, 2019), results based on AI assessment are quite often "black-boxed" and neither educators nor learners have access to the logic and reasoning deployed by the chatbot in arriving at its judgements (Swiecki et al., 2022). This is different to human-based assessment where familiarity with the learners and the learning context provide a background to, and possible explanations for, human-based judgements. Human markers are also available to query in case of moderation of marks, for example, unlike chatbots. ChatGPT's generous assessment using our independent rubric also casts doubt on the chatbot's capacity for fair and critical assessment as it appears to provide random scores, echoing Farrokhnia et al. (2023) who argued that the chatbot lacks higher-order thinking skills and is unable to evaluate the quality of responses. It has also been argued that to develop the credibility of AI-based assessment, large investment in training on appropriate and targeted datasets is needed, however, this would add further complexity to the previously mentioned problem of blackboxed AI judgement in assessment (Mizumoto and Eguchi, 2023).

5.3 With Reference to the Findings From Q1 and Q2, to What Extent can ChatGPT be Considered a Disruptor?

Vis-à-vis the third research question, the Findings section has established limited evidence of ChatGPT being a disruptive innovation with respect to its ability to alter existing processes related to essay assessment. According to Christensen et al. (2015) and Al-Imarah and Shields (2019), disruptive innovations either create new markets or feed into the low-end of existing markets, i.e., those individuals who are less discerning and demanding. Additionally, the innovation, rather than being transformative of that market, could instead be either sustaining

or efficiency-enhancing. To determine how ChatGPT would affect a market related to essay assessment, we considered contract cheating (Clarke and Lancaster, 2006) where there is an existing "market" for commissioning bespoke essays (Ellis, Zucker and Randall, 2018) e.g., using so-called essay mills (Sweeney, 2023), or using pre-written material from essay "banks" (Medway, Roper and Gillooly, 2018). The "buyer", i.e., the learner who wishes to cheat, can choose from different levels of quality, price, turnaround, and subject matter on offer (Wallace and Newton, 2014; Rigby et al., 2015). Higher education institutions acknowledge the difficulty in both detecting and deterring the acquisition of essay assignments in this manner (Ellis, Zucker and Randall, 2018). Some learners already use essay mills to cheat the system and produce work that they believe will attain a desirable grade, however, this level of quality is not guaranteed (Medway, Roper and Gillooly, 2018). ChatGPT, since it is relatively easy to use, convenient, cheap and quick, is likely to disrupt this contract cheating essay business model if it can produce essays that at least meet a decent grade threshold for the learner wishing to submit non-original work. We have demonstrated, however, in the findings, that the quality of a ChatGPT-generated essay is likely to be a bare pass. Hence, ChatGPT is likely to affect the status quo in contract cheating essay business models at the low end of the existing market. This finding concurs with speculations about ChatGPT's influence on higher education assessments (Stokel-Walker, 2022). It can produce a product desirable for learners who are not discerning "customers", i.e., who just need to pass their courses but are not seeking a particular academic performance level. Since this may not represent a unique market segment, it is unlikely, then, that ChatGPT would be able to create a new market for cheating on essay assignments. We have also already observed in the findings that, related to the design, delivery, and marking processes of the essay assessment, ChatGPT cannot be considered transformative, since its outcomes were deemed to have errors. Thus, it could be a sustaining and/or efficiency-enhancing innovation in this market. The errors observed in the findings negate its ability to enhance efficiency, even though the results were generated far more quickly than would be done by a human learner. However, since it could strengthen and retain "customers" at the lower-end of the market for contract cheating in essays, it can be assumed that ChatGPT will *sustain* technology-enabled ways of cheating, but hardly disrupt or enhance their efficiencies. This resonates with recent studies highlighting that AI utilised in addition to existing essay mills poses an additional threat to academic integrity, but one which is still evolving (Sweeney, 2023).

6. Conclusions and Implications

The paper contributes to the evolving body of research on the influence of ChatGPT on higher education assessment by demonstrating that ChatGPT has not yet reached the level of "disruptor", but it is known that the AI will keep being trained and improved, thus retaining this potential in the future. The chatbot can generate answers to queries within minutes but from our study, it was evident that the results may not always be reliable. Based on our findings and analysis, ChatGPT is more at the "magic wand" than the disruptor end of the spectrum of disruptive innovations. Like a magician, it has achieved the illusion of a completed assessment but not an authentic innovation. It has achieved a quick and superficially effective delivery of the essay assessment and at least sustainable innovation in the market for contract cheating on essays, however, it has not succeeded in transforming the product or in creating new markets. It may have an effect on sustaining the lower end of the market for such cheating but not in disrupting it. These current limitations of the chatbot open up a window of opportunity for addressing current weaknesses of the traditional essay and other forms of assessment that may be vulnerable to cheating from AI-generated content. The following implications discuss this further.

6.1 Implications

Higher education instructors need to be equipped with the knowledge of special AIs that can detect ChatGPT-produced work, e.g. ZeroGPT, or as our study found, to employ a range of techniques to detect it. Our findings suggest that ChatGPT-produced work can be detected by: checking the reference list and determining whether the references are functional; examining the text if the same in-text references are being repeated across different sections; and checking for the lack of inclusion of novel or original ideas.

Studies are reporting that it will become nearly impossible to distinguish the students' own words from the words of this AI (O'Connor and ChatGPT, 2023) due to ChatGPT's capability in generating different responses to the same query (Susnjak, 2022). With the high chances of students outsourcing their tasks, it becomes even more important for educators to reconsider and revisit their assessment practices.

One way forward in addressing these would be the adoption of diverse and innovative tasks that will tap into the students' critical thinking and creativity. Universities could return to oral examinations and also recommend that educators incorporate multimedia into tasks since the chatbot generates answers to text-based prompts

only. Educators could also tap into constructivist ideas that involve students' reflections, collaborations with other students, exploratory discussions, and presentations. Multipart assessment could also be incorporated by including both written and spoken parts. Universities need to educate their students on the ethics of using such AI and establish their regulations in terms of where they stand when it comes to the use of AI. After all, as the chatbot will be trained and improved over time fundamental changes to assessment might need to be in place to avoid the potential disruptions it may bring.

On the flip side, since the chatbot has limitations as far as creativity and originality are concerned it opens up the possibility that students could use it, in essay writing, for the more routine work of finding materials, assembling arguments, and creating a logical flow. This could potentially enable students to then focus on the higher-order capabilities of broader reading, better synthesis of ideas, drawing upon diverse perspectives, i.e., those aspects that increase their creative and original thinking.

6.2 Limitations and Further Work

It is worth mentioning that this study tested ChatGPT-3 which was released on the 30th of November, 2022. Therefore, the results of this study may not apply to the latest ChatGPT version. Despite the mentioned limitations, this study provided interesting findings in relation to ChatGPT's role in higher educational settings, specifically in the area of essay assessment. As ChatGPT will keep actively being developed and improved, future research could look into the conduct of a comprehensive systematic literature review comparing the different releases of this AI and reporting back on what types of features were added and improved as far as they may affect higher education assessment capability. Since new versions of this AI will keep being released future research could also develop special guidelines on approaching ChatGPT ethically and responsibly. We also hope this study will foster more research investigating further academic integrity and cheating concerns when it comes to the use of AIs in academic assessment.

References

Akour, M. and Alenezi, M., 2022. Higher Education Future in the Era of Digital Transformation. *Education Sciences*, 12(11), p.784. Available at: https://doi.org/ 10.3390/educsci12110784.

Al-Imarah, A.A. and Shields, R., 2019. MOOCs, disruptive innovation and the future of higher education: A conceptual analysis. *Innovations in Education and Teaching International,* 56(3), pp.258-269. Available at: https://doi.org/10.1080/14703297.2018.1443828.

Arnold, I.J.M., 2016. Cheating at online formative tests: Does it pay off? *The Internet and Higher Education,* 29, pp.98–106. Available at: https://doi.org/10.1016/j.iheduc.2016.02.001.

Aydin, Ö. and Karaarslan, E., 2022. OpenAI ChatGPT generated literature review: Digital town in healthcare. *Social Science Research Network,* 4308687. Available at: http://dx.doi.org/10.2139/ssrn.4308687.

Bekurs, D. and Santoli, S., 2004. Writing is power: Critical thinking, creative writing, and portfolio assessment. *Essays in Education*, 10(1), p.2. Available at: https://openriver.winona.edu/eie/vol10/iss1/2.

Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O. and Kerres, M., 2020. Mapping research in student engagement and educational technology in higher education: a systematic evidence map. *International journal of educational technology in higher education*, 17(1), pp.1-30. Available at: https://doi.org/10.1186/s41239-019-0176-8.

Christensen, C.M., 1997. *The Innovator's Dilemma: when new technologies cause great firms to fail.* Harvard Business School Press.

Christensen, C.M. and Raynor, M.E., 2003. *The innovator's solution: Creating and sustaining successful growth.* Harvard Business School Press.

Christensen, C. M., Raynor, M.E. and Mcdonald, R., 2015. What Is Disruptive Innovation? *Harvard Business Review*, 93(12), pp.44–53. Available at: https://hbr.org/2015/12/what-is-disruptive-innovation.

Clarke, R. and Lancaster, T., 2006. Eliminating the successor to plagiarism? Identifying the usage of contract cheating sites. In: *2nd international plagiarism conference*. Northumbria Learning Press. pp.1–13.

Ellis, C., Zucker, I.M. and Randall, D., 2018. The infernal business of contract cheating: understanding the business processes and models of academic custom writing sites. *International Journal for Educational Integrity*, 14(1), pp.1–21. Available at: https://doi.org/10.1007/s40979-017-0024-3.

Farrokhnia, M., Banihashem, S.K., Noroozi, O. and Wals, A., 2023. A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International,* pp.1-15. Available at: https://doi.org/10.1080/14703297.2023.2195846.

Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A. and Hatta, H., 2023. Analysing the Role of ChatGPT in Improving Student Productivity in Higher Education. *Journal on Education*, 5(4), pp.14886-14891. Available at: https://doi.org/10.31004/joe.v5i4.2563.

Fernández-Gutiérrez, M., Gimenez, G. and Calero, J., 2020. Is the use of ICT in education leading to higher student outcomes? Analysis from the Spanish Autonomous Communities. *Computers & Education,* 157, 103969. Available at: https://doi.org/https://doi.org/10.1016/j.compedu.2020.103969.

Flavin, M., 2012. Disruptive technologies in higher education. *Research in Learning Technology,* 20*,* pp.102-111. Available at: https://doi.org/10.3402/rlt.v20i0.19184.

Flavin, M., 2016. Technology-enhanced learning and higher education. *Oxford Review of Economic Policy,* 32(4), pp.632-645. Available at: https://doi.org/10.1093/oxrep/grw028.

Flavin M., 2021. A Disruptive Innovation perspective on students' opinions of online assessment. *Research in Learning Technology*, 29. Available at: https://doi.org/10.25304/rlt.v29.2611.

Fojtik, R., 2014. Mobile Technologies Education. *Procedia – Social and Behavioral Sciences,* 143, pp.342-346. Available at: https://doi.org/10.1016/j.sbspro.2014.07.417.

Generative Pretrained Transformer, G., Thunström, A., Osmanovic. and Steingrimsson, S., 2022. Can GPT-3 write an academic paper on itself, with minimal human input?. Available at: https://hal.science/hal-03701250.

Hanesworth, P., Bracken, S. and 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), pp.98–114. Available at: https://doi.org/10.1080/13562517.2018.1465405.

Joshi, A., Vinay, M. and Bhaskar, P., 2020. Impact of coronavirus pandemic on the Indian education sector: Perspectives of teachers on online teaching and assessments. *Interactive Technology and Smart Education,* 18(2), pp.205–226. Available at: https://doi.org/10.1108/ITSE-06-2020-0087.

Kabir, S., Udo-Imeh, D.N., Kou, B. and Zhang, T., 2023. Who Answers It Better? An In-Depth Analysis of ChatGPT and Stack Overflow Answers to Software Engineering Questions. *arXiv preprint.* Available at: https://doi.org/10.48550/arXiv.2308.02312.

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

Kok, J.N., Boers, E.J.W., Kosters, W.A. and Van der Putten, P., 2009. Artificial Intelligence: Definition, Trends, Techniques and Cases. *Artificial Intelligence*, 1, pp.270-299.

Kumar, N., 2006. Strategies to fight low-cost rivals. *Harvard Business Review,* 84(12), pp.104-112. Available at: https://hbr.org/2006/12/strategies-to-fight-low-cost-rivals.

Köprülü, F., 2021. The Effect of Using Technology Supported Material in Teaching English to First-Year Primary School Children: On Their Academic Success During COVID-19. *Frontiers in Psychology,* 12. Available at: https://doi.org/10.3389/fpsyg.2021.756295.

Markides, C. and Sosa, L., 2013. Pioneering and First Mover Advantages: The Importance of Business Models. *Long Range Planning,* 46(4), pp.325-334. Available at: https://doi.org/https://doi.org/10.1016/j.lrp.2013.06.002.

Medway, D., Roper, S. and Gillooly, L., 2018. Contract cheating in UK higher education: A covert investigation of essay mills. *British Educational Research Journal*, 44(3), pp.393–418. Available at: https://doi.org/10.1002/berj.3335.

Menkhoff, T., Chay, Y., Bengtsson, M., Woodard, J. and Gan, B., 2015. Incorporating microblogging ("tweeting") in higher education: Lessons learnt in a knowledge management course. *Computers in Human Behavior,* 51(B), pp.1295-1302. Available at: https://doi.org/10.1016/j.chb.2014.11.063.

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. Available at: https://doi.org/10.1016/j.rmal.2023.100050.

O'Connor, S. and ChatGpt., 2023. Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Education in Practice,* 66, 103537. Available at: https://doi.org/https://doi.org/10.1016/j.nepr.2022.103537.

Oliva-Córdova, L. M., Garcia-Cabot, A. and Amado-Salvatierra, H. R., 2021. Learning analytics to support teaching skills: a systematic literature review. *IEEE Access,* 9, pp.58351–58363. Available at: https://doi.org/10.1109/ACCESS.2021.3070294.

OpenAI., 2023. Introducing ChatGPT Plus. OpenAI. Available at: https://openai.com/blog/chatgpt-plus/

Pavlik, J. V., 2023. Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *Journalism & Mass Communication Educator*, 78(1), pp.84–93. Available at: https://doi.org/10.1177/10776958221149577.

Paul, R., 2005. The state of critical thinking today. *New Directions for Community Colleges,* 2005(130), pp.27-38. Available at: https://doi.org/https://doi.org/10.1002/cc.193.

Race, P., 2018. Is the 'time of the assessed essay' over? Teaching Perspectives from the Business School. Available at: https://blogs.sussex.ac.uk/business-school-teaching/2018/11/14/is-the-time-of-the-assessed-essay-over/

Rigby, D., Burton, M., Balcombe, K., Bateman, I. and Mulatu, A., 2015. Contract cheating & the market in essays. *Journal of Economic Behavior & Organization*, 111, pp.23–37. Available at: https://doi.org/10.1016/j.jebo.2014.12.019.

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

Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G. and Moy, L., 2023. ChatGPT and Other Large Language Models Are Double-edged Swords. *Radiology*, 230163. Available at: https://doi.org/10.1148/radiol.230163.

Siddhpura, A., V, I. and Siddhpura, M., 2020. Current state of research in application of disruptive technologies in engineering education. *Procedia Computer Science,* 172, pp.494-501. Available at: https://doi.org/https://doi.org/10.1016/j.procs.2020.05.163.

Stokel-Walker, C., 2022. AI bot ChatGPT writes smart essays — should professors worry? *Nature*. Available at: https://doi.org/10.1038/d41586-022-04397-7.

Sum, M. and Oancea, A., 2022. The use of technology in higher education teaching by academics during the COVID-19 emergency remote teaching period: a systematic review. *International Journal of Educational Technology in Higher Education,* 19, 59. Available at: https://doi.org/10.1186/s41239-022-00364-4.

Susnjak, T., 2022. ChatGPT: The End of Online Exam Integrity?. *arXiv preprint*. Available at: https://doi.org/10.48550/arXiv.2212.09292.

Sweeney, S., 2023. Who wrote this? Essay mills and assessment – Considerations regarding contract cheating and AI in higher education. *The International Journal of Management Education*, 21(2), p.100818. Available at: https://doi.org/10.1016/j.ijme.2023.100818.

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

Tate, T. P., Doroudi, S., Ritchie, D. and Xu, Y., 2023. Educational Research and AI-Generated Writing: Confronting the Coming Tsunami. Available at: https://doi.org/10.35542/osf.io/4mec3.

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A. and Hickey, D. T., 2023. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments* 10(1), 15. Available at: https://doi.org/10.1186/s40561-023-00237-x.

Wallace, M.J. and Newton, P.M., 2014. Turnaround time and market capacity in contract cheating. *Educational Studies*, 40(2), pp.233–236. Available at: https://doi.org/10.1080/03055698.2014.889597.

Whisenhunt, B. L., Cathey, C. L., Hudson, D. L. and Needy, L. M., 2022. Maximizing learning while minimizing cheating: New evidence and advice for online multiple-choice exams. *Scholarship of Teaching and Learning in Psychology,* 8(2), pp.140–153. Available at: https://doi.org/10.1037/stl0000242.

Zhai, X., 2022. ChatGPT user experience: Implications for education. *arXiv preprint*. Available at: https://doi.org/10.48550/arXiv.2212.09292.