Solving the 2-Sigma Problem through AI-driven Tutoring and AI-driven Marking in ITE

Abstract

Artificial Intelligence (AI) has been a topic of discussion in many sectors and industries because of its capability to mimic human intelligence. The use of AI is growing in the education sector for practical reasons such as task automation, which potentially frees up lecturers to spend more quality time with students. McKinsey reports that 64% of Singapore educators stated not having enough time to cater to personalised learning for their students as much as they want to[1]. Yet, one-to-one human tutoring has long been thought to be the most effective approach to teaching and learning, but is untenable in mass education. Bloom, in 1984, observed that students who receive one-on-one tutoring perform two standard deviations better than students who learn in conventional settings[2], but providing individual tutors for every single student is not realistic. The only solution for now, perhaps, is to leverage on the human-like system of AI to deliver a system that is as effective as human one-on-one tutoring. This paper reports a pilot project undertaken by lecturers and students of the School of Business & Services at the Institute of Technical Education (ITE) College Central, using AI-driven marking and AI-driven tutoring, collectively known as the AI Chatbot, in concert with classroom routine. The AI-driven tutoring affords instant feedback to the students when they are still most engaged in their learning process. The use of AI-driven marking on the other hand, allows continuous formative assessment, where students can attempt the same assessment paper multiple times, without having to wait for their lecturers' feedback. In this pilot study, 24 lecturers across 26 classes and 4 modules used the AI Chatbot for a total of 9 months. The novel nature of AI used in this study sees the technology acceptance model (TAM) being adopted as the methodology to measure the use of the AI Chatbot. There were 18 lecturer ($n_{lecturers}$ = 18) and 480 student ($n_{students}$ = 480) respondents to this survey. Our findings indicate that about 84% of students and lecturers find the AI Chatbot useful and easy to use. About 87% of students and lecturers agree that the AI Chatbot will enhance the effectiveness of teaching and learning. Similarly, about 85% of students and lecturers are agreeable that using AI Chatbot to support learning and assessment is a good idea and that they have positive feelings about using AI-based tools for the assessment for learning. Lastly, a small percentage of about 28% of students and lecturers are fearful of committing errors on the AI Chatbot and using the AI Chatbot generally causes them stress. While this small percentage of respondent experience stress, a larger percentage of lecturer respondents of about 64% of them, experience lower marking fatigue with 50% to 75% of perceived time savings per marking process.

Keywords: *2-sigma problem, Artificial intelligence, Technology Acceptance Model, TAM, Formative assessment, AI-driven marking, AI-driven tutoring, Intelligent grading, Intelligent tutoring*

Principal investigators

Ms. Shalyn TAN
Senior Lecturer - Mentor, Business Services,
School of Business & Services College Central
Institute of Technical Education
shalyn_s_h_tan@ite.edu.sg

Ms. Laura LIAUW
Lecturer, Event Management,
School of Business & Services College Central
Institute of Technical Education
laura_l_c_liauw@ite.edu.sg

Ms. Geraldine HONG
Section Head, LifeSkills,
School of Business & Services College Central
Institute of Technical Education
geraldine_p_p_hong@ite.edu.sg

Mr. Damian TAN
Lecturer, Event Management Services,
School of Business & Services College Central
Institute of Technical Education
damian_t_m_tan@ite.edu.sg

Ms. Wendy LIM
Senior Lecturer, LifeSkills,
School of Business & Services College Central
Institute of Technical Education
wendy_h_p_lim@ite.edu.sg

[1] McKinsey Global Teacher and Student Survey. Average of Canada, Singapore, United Kingdom, and United States in 2017.

[2] Bloom, Benjamin S (June–July 1984). "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring"

Introduction

Understanding the role of timely feedback to students when they are still most engaged in their learning process is an important lever to improve quality of learning. In Singapore, 80% of lecturers routinely assess their students' progress by observing them and providing immediate feedback during class [1]. The most common and effective feedback that lecturers adopt, would be one that enhances active recall, and not merely passive memorization. Active recall involves repeated testing on quiz questions and receiving timely feedback on why certain answers were wrong and specifically which topic to be revised on before attempting the same quiz again. The usual practice of most lecturers in a typical classroom, would be to administer quiz questions, collect student responses, and then spend some time marking and grading them. Lecturers then return to class to share the grading and feedback with the students. This is by far the most practical method that lecturers adopt, as most of these quiz questions involve short or even long answer questions, which requires marking by the lecturers. There needs to be a turnaround time for lecturers to mark before feedback can be given to the students. To achieve optimal learning, timely or immediate feedback is favoured, which drives the creation of a system that can instantly mark short answer questions as effectively as a human marker and immediately return feedback to students.

Bloom, in 1984, found a way to improve student performance, helping the average students perform better than 98% of their peers in a conventional classroom setting [2]. This was attributed to one-on-one human tutoring. Students who received one-on-one tutoring performed two standard deviations better than students who learnt in a conventional class. While one-on-one tutoring improved student performance, providing individual tutors for every single student was not realistic. This led to the conundrum, which Bloom called the 2-sigma problem. Solving the problem required the creation of a system that is as effective as human one-on-one tutoring. Aside from being a personal tutor, the availability of such system allows for personalised learning which contributes to optimal learning. Personalised learning is antithetical to 'one size fits all'. Every student learns at a different pace and different style. McKinsey reported that 64% of Singapore educators stated they did not have enough time to cater to personalised learning for their students as much as they want to [3].

Existing and emerging technologies can potentially free up lecturers to spend more quality time with students. Just as how social media platforms learn our behaviors and preferences to populate personalised content feeds to provide an overall tailored experience, the unique affordance of artificial intelligence (AI) could similarly extend the personalised and tailored experience to our students. AI offers potential benefits through tools that provide real-time feedback to students as they study and complete formative assignments and tasks [4]. These tools might also provide customised and tailored learning experiences and pathways for students. It is increasingly automated and less dependent on a lecturer's time and effort thus freeing up time to be reinvested in individual students in need of additional support and other higher-order tasks and emotional human support. AI can already automate the grading of homework, evaluating essays which allows lecturers to spend more time with students one-on-one [6]. While it may sound onerous to train a machine to grade short answer questions and essays, Moravex's paradox explains it otherwise. High-level reasoning tasks like performing intelligent tests and solving difficult mathematical problems, everything that seems to be hard for human and needs special skills, requires much less computation and is easier to reverse-engineer and programme, and therefore train the machine, in comparison to low level sensory-motor skills like recognizing faces and voices [7].

The need for AI in the education sector is growing in reality and in necessity; in order to provide timely feedback and personalised learning for students, and to assist lecturers in marking and grading short answer or essay type assessments. However, the novel nature of AI remains a source of concern and raises queries in both its effectiveness and efficiency. Despite the evident advantages entailed in the use of AI in the teaching-learning process, there is also a growing concern among society about the implications and risks of the use of this technology in the field of teaching and learning [8].

This paper presents the awareness of the opportunity AI presents; in areas of personalised learning and automated marking of short answer and essay assessments. In this context, it is important to know the position of the lecturers and students regarding the use of AI in education, particularly in the use of AI in tutoring students and in marking short answer and essay type assessments. In order to do so, this paper adopts the Technology Acceptance Model (TAM) [9], to measure the adoption of these AI technologies among lecturers and students of School of Business & Services, College Central at the Institute of Technical Education (ITE), Singapore.

Objectives

The objective of this research was to study the willingness of lecturers and students in ITE to use AI technology for teaching and learning. The AI technology used is the AI-driven tutoring system for the students and AI-driven marking system for the lecturers. These AI technologies are collectively known as AI Chatbot. The measurement of the lecturers' and students' AI Chatbot acceptance leading to their actual use was based mainly on the Technology Acceptance Model (TAM). The research looked at the relationship of lecturers' and students' intention to use AI Chatbot with selected constructs such as their relative advantage, perceived usefulness, perceived ease of use, behavioral intention, anxiety and actual use that would then serve to inform lecturers and school administrators with their overall implementation of AI-driven technologies for teaching and learning.

Literature Review

Technology adoption by lecturers and students

The growing importance of artificial intelligence systems in daily life leads to increasing demands for artificial intelligence (AI) as a topic in schools; and the necessity to consider what students in the 21st century should know about the topic [10]. While we consider the competences and topics that should be covered in school relating to AI, which essentially falls into the area of learning about AI, classroom practices that leverage AI technologies, for instance AI-driven tutoring or AI-driven marking system, which essentially falls into the area of learning with AI, should be concurrently considered. Learning with AI would ease the process of learning about AI, as one would experience AI in action which is always helpful in learning about a subject as one is able to see the subject in practice. Therefore, lecturers and students should use AI technologies as a learning tool, embedded invisibly in the normal routine of the classroom. However, the successful integration of the use of AI technologies in the classroom requires proper preparation of lecturers for this task.

The willingness and decision to adopt technology in the classroom is influenced by lecturers' individual attitudes towards technology use, which were formed from specific underlying personal beliefs about the consequences of adoption [11]. On the basis of AI and its associated technologies being a novel topic, the Technology Acceptance Model (TAM) is used in this research to measure lecturers' and students' willingness to use them. TAM is based on principles from psychology, which specifies how to measure the intrinsic components of beliefs, attitudes and behavior. TAM employs "perceived ease of use" and "perceived usefulness" as the determinants of intention, which in turn determines use [12]. Figure 1 depicts the original TAM (Davis, 1989).

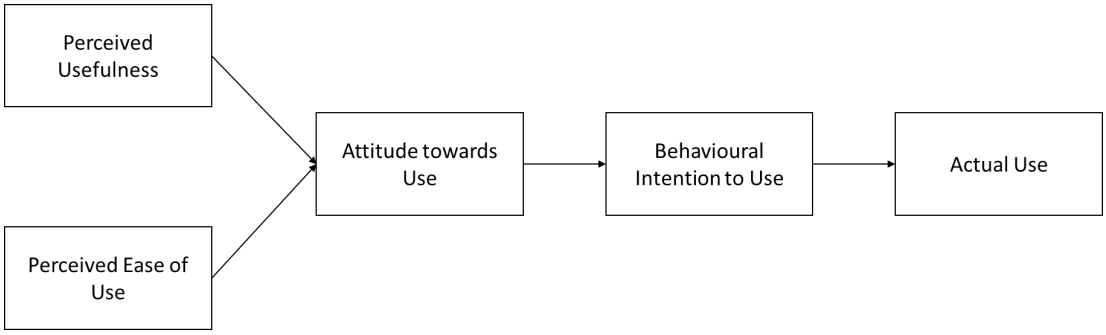


Figure 1: The original technology acceptance model TAM (Davis, 1989)

TAM's main advantage is in its parsimony and flexibility, which allows the application of the model to a wide variety of contexts and technologies. Due to this, there has been many extensions to TAM. One of such, is the addition of Relative advantage as an added construct. Relative advantage [28] is defined here as to the extent an individual perceives AI as being better than the idea it supersedes. Advantages of AI is expressed as effectiveness, time and effort savings, or immediacy of the reward. Based on this, the research proposes the use of TAM with an extended construct called Relative Advantage. Figure 2 shows the extended TAM.

On the other hand, anxiety is a natural reaction of the human body to stress; in other words, it is a kind of feeling of fear or concern about future [30]. In technology acceptance literature, different kinds of anxieties have been defined, such as technology anxiety, computer anxiety, online shopping anxiety, mobile anxiety, and most recently AI anxiety. Strong and negative emotions may be demonstrated in some situations that occur when interacting with technology, so this means technology anxiety is experienced [31]. As seen in the technology acceptance model, some causal relations is defined in that one's attitude toward using a technology leads to behavioral intention to use and behavioral intention to use leads to actual system use. In addition, the dimension of attitude

towards using are explained with three variables - relative advantage, perceived usefulness and perceived ease of use.

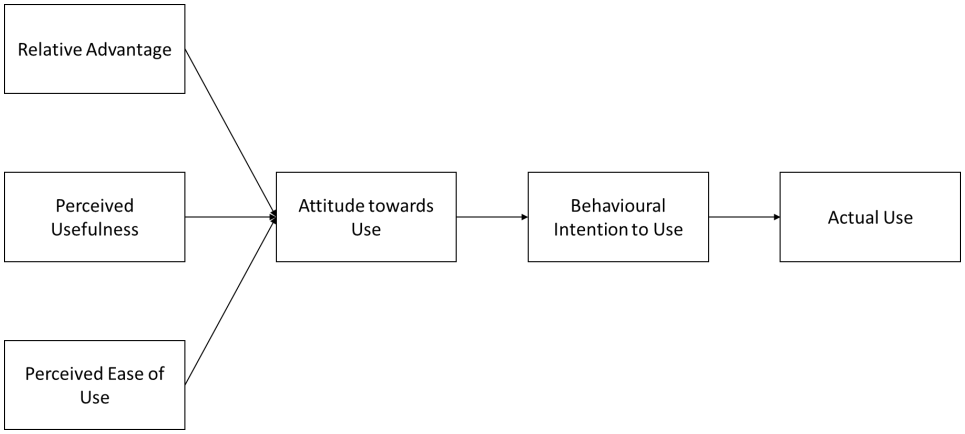


Figure 2: The extended TAM

This study proposes a revision to TAM in that the attitude towards use be replaced with AI anxiety. The relations remain in that AI anxiety is inversely related to behavioral intention. This means the lower the anxiety experienced by the user, the higher the likelihood of the users to intentionally use the technology. Likewise, AI anxiety forms an inverse relationship with the three variables - relative advantage, perceived usefulness and perceived ease of use. Figure 3 shows the revised TAM.

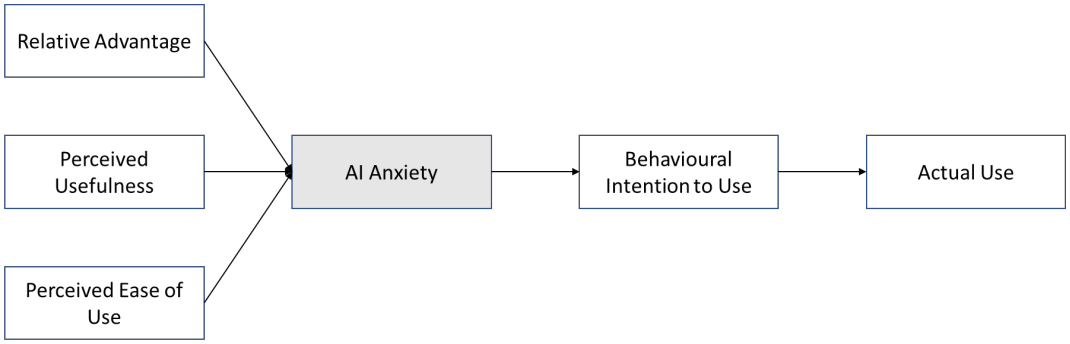


Figure 3: The revised TAM

Automated marking versus human marking of formative assessments

The most highly valued activity for a lecturer is to teach; but in reality, a lecturer's role is not only to teach but includes a repertoire of non-teaching duties like preparation of lessons, marking/correcting students' work, administrative duties, student counselling, involvement in extracurricular activities, communications with parents/guardians, and other tasks. The distribution of class time during an average lesson includes actual teaching and learning, administrative tasks and keeping order in the classroom [14]. In Singapore, school teachers spend only 74% of classroom time on actual teaching, which is lower than the OECD average of 78% [1]. Lesson planning and marking/correcting students' work form the next big chunk of lecturers' time. There is no doubt that teaching and learning should comprise as big a portion as possible of lecturers' time each day, which means that lecturers' would welcome any reliable and efficient way to reduce their load for other non-instructional tasks like marking assessment papers, so that they could devote more time to actual teaching.

Since lecturers spend the next large amount of time on marking/correcting students' work, we must find an efficient technology that lecturers will trust, to mark short answer/essay type responses. For this to happen, the steps required by the technology to mark and derive a score for a formative paper must be sufficiently close to one that is marked and scored by a human marker. Artificial Intelligence (AI) technologies are now a common feature in institutional assessments and instructional tools [4]. The advancement of natural-language processing (NLP) has enabled researchers to look for new approaches to extract measures that directly correlate to writing quality [15]. There has been multiple attempts over the past years to automate marking of short answer/essay type of questions, to achieve a level close enough to that of a human marker [15 - 25]. Grading natural language responses to short answer/essay questions can be considered much more difficult, as an understanding of the natural language is required. NLP techniques are required to analyse the language in student answers and

perform textual manipulation, or statistical techniques based on the features extracted from them [21]. On submission, the student's response is compared against recommended answers to derive a number of parameters which reflect knowledge and understanding as exhibited by the student [13]. This mimics the way a human marker would mark students' responses. Most lecturers would be confident in marking student responses for a content they did not themselves know, provided they read the appropriate recommended answers first, and were given a marking scheme. The human marker then looks for the coverage and evidence of knowledge and understanding of the student against the recommended answer. In that perspective, training a machine would somewhat be similar.

Personalised learning

Learning is a natural human activity that is shaped by personal experiences, cognitive awareness, personal bias, opinions, cultural background, and environment [26]. Hence, the notion that every student learns differently is an accurate statement. Personalised learning in the context of education, is easier to achieve when the number of students is small like in the case of apprenticeship, mentoring, and one-on-one tutoring. It becomes an onerous task in large student-to-lecturer ratio. On the other hand, lecturers know the advantages of and the need for personalised learning. Customising instructions for students can effectively help students to meet their needs and goals. Accepting mass education and acknowledging the advantages of personalised learning, the only way to balance the two is to leverage other models of learning and technologies to help make personalised learning possible in mass education. Bloom [2] observed that students who received one-on-one tutoring performed two standard deviations better than peers who learnt in conventional setting. In response to this reality, a digital solution that is as effective as human one-on-one tutoring could be the solution.

Personalised learning system

A personalized learning system can adapt itself when providing learning support to different learners to defeat the weakness of one-size-fits-all approaches in technology-enabled learning systems. The goal is to have a learning system that can dynamically adapt itself based on a learner's characteristics and needs. Human one-on-one tutors can do this and now it is possible for digital systems to do so as well [26]. Personalised learning systems allow students to advance at different paces. If the pace of learning is personalized, students with different abilities are allowed different amounts of time to learn the same material [27]. Which is what human lecturers try to do albeit laboriously in large classes. Close to 70% of lecturers in Singapore let students practice similar tasks until every student has understood the subject matter, in an attempt to achieve clarity of instructions [1].

Methodology

This study is based on a quantitative approach and a set of questionnaires is utilised for the purpose of meeting the objectives of the study. To operationalize the proposed revised TAM used, the questionnaire design was adopted based on the literature reviewed (refer to Annex A and Annex B). Students were administered with 20 questions, while lecturers were administered with an extra 3 sentiment-type questions, totaling 23 questions for lecturers. A Likert Scale is applied for each set of items in the questionnaire. The Likert scale is designed to examine how strongly respondents agree or disagree with statements on a seven-point scale with the following anchors: (1) Strongly disagree, (2) Moderately disagree, (3) Slightly disagree, (4) Neither disagree nor agree, (5) Slightly agree (6) Moderately agree, (7) Strongly agree [29] (refer to Annex C). The questionnaire was designed using Microsoft Forms and was administered to lecturers and students of the School of Business & Services, College Central, ITE. The pilot study was conducted over the course of 9 months from July 2021 to March 2022, covering 26 classes and 4 modules.

Research hypotheses

As stated in the objectives and consistent with related literature, this study tested the following hypotheses:

- H1: Lecturers' and students' AI anxiety is negatively affected by the relative advantage they see in their use of AI.
- H2: Lecturers' and students' AI anxiety is negatively affected by the perceived usefulness of AI.
- H3: Lecturers' and students' AI anxiety is negatively affected by the perceived ease of use of AI.
- H4: Lecturers' and students' behavioural intention to use AI is negatively affected by their AI anxiety.
- H5: Lecturers' and students' actual use of AI is affected by their behavioural intention to use AI.

Findings

The descriptive statistical analyses applied to responses from students, in Table 1, show that the mean scores of all items ranged from ranged from 3.68 to 5.75. The standard deviations ranged from 1.272 to 1.954 and the skew and kurtosis indices from -1.327 to .128 and -1.087 to 1.818 respectively. The data in Table 1 were considered to be univariate normal [32].

Item		Mean	SD	Skewness	Kurtosis
1	AI Chatbot would enable me to accomplish my tasks (revision) more quickly.	5.59	1.377	-1.233	1.691
2	AI Chatbot would improve the quality of my work (revision).	5.50	1.365	-1.080	1.308
3	AI Chatbot would enhance my effectiveness on the job (revision).	5.55	1.344	-1.168	1.599
4	AI Chatbot would make my job (revision) easier.	5.63	1.408	-1.327	1.818
5	AI Chatbot gives me greater control over my work (revision).	5.50	1.346	-0.933	0.800
6	In general, I doubt about using AI Chatbot for fear of committing errors that I cannot correct.	4.73	1.566	-0.464	-0.130
7	In general, the use of AI Chatbot causes me stress.	3.79	1.909	0.058	-1.023
8	In general, I feel apprehensive (worried) when using AI Chatbot.	3.68	1.954	0.128	-1.087
9	Using AI Chatbot for my assessment allows me to learn effectively.	5.44	1.313	-0.912	1.028
10	Using AI Chatbot for my assessment improves my academic productivity.	5.30	1.351	-0.809	0.816
11	Using AI Chatbot increases my assessment opportunities.	5.46	1.272	-0.878	1.151
12	Incorporating AI Chatbot for my assessment is useful.	5.51	1.273	-0.976	1.291
13	My interaction with AI Chatbot is clear and understandable.	5.53	1.294	-0.779	0.504
14	I find it easy to get AI Chatbot to do what I want them to do	5.48	1.355	-0.790	0.346
15	I find AI Chatbot easy to use.	5.75	1.294	-1.164	1.440
16	I would participate in AI-based assessment processes.	5.26	1.398	-0.717	0.460
17	I plan to participate in AI-based assessment processes often.	5.03	1.455	-0.573	0.204
18	I hope to participate in AI-based assessment processes in the future.	5.08	1.430	-0.600	0.304
19	Using AI Chatbot to support my learning is a good idea.	5.39	1.370	-0.858	0.798
20	I have positive feelings about AI Chatbot for my learning.	5.50	1.372	-0.967	1.014

Table 1: Descriptive statistics of the 20 items used in Student survey

The same descriptive statistical analyses were applied to lecturer responses, as seen in Table 2. The mean scores of all items ranged from ranged from 2.17 to 6.50. The standard deviations ranged from .786 to 1.653 and the skew and kurtosis indices from -1.679 to 1.667 and -1.040 to 2.480 respectively. The data in Table 2 were considered to be univariate normal [32].

Item		Mean	SD	Skewness	Kurtosis
1	AI Chatbot would enable me to accomplish my tasks (marking) more quickly.	6.39	0.916	-1.437	1.335
2	AI Chatbot would improve the quality of my work (marking).	6.06	0.938	-0.602	-0.533
3	AI Chatbot would enhance my effectiveness on the job (marking).	6.06	0.873	-0.713	0.197
4	AI Chatbot would make my job (marking) easier.	6.50	0.786	-1.227	-0.069
5	AI Chatbot gives me greater control over my work (marking).	5.83	1.150	-0.942	0.636
6	In general I doubt about using AI Chatbot for fear of committing errors that I cannot correct.	3.44	1.653	0.505	-0.188
7	In general, the use of AI Chatbot causes me stress.	2.17	1.618	1.667	1.996
8	In general, I feel apprehensive (worried) when using AI Chatbot.	2.33	1.572	1.522	1.725
9	AI Chatbot will enable me to get the information of the students (their scores) quickly.	6.00	1.237	-1.679	2.480

10	AI Chatbot is useful in the rapid retrieval of information from the students. (eg. number of times students attempt the paper)	5.94	1.259	-1.476	1.747
11	AI Chatbot will save the time of lecturers.	6.22	0.943	-1.443	1.874
12	Using AI Chatbot would improve my tracking of students' performance.	5.94	0.998	-0.677	-0.370
13	My interaction with AI Chatbot is clear and understandable.	5.61	1.037	-0.155	-1.007
14	I find it easy to get AI Chatbot to do what I want them to do.	5.56	1.097	-0.761	0.335
15	I find AI Chatbot easy to use.	5.89	0.832	-0.465	0.112
16	I would participate in AI-based assessment processes.	5.83	0.924	-0.140	-0.910
17	I plan to participate in AI-based assessment processes often.	5.72	1.074	-0.335	-1.040
18	I hope to participate in AI-based assessment processes in the future.	5.72	1.074	-0.335	-1.040
19	Using AIs to support the assessment of the students is a good idea.	6.22	0.943	-0.969	0.036
20	I have positive feelings about implementing AI-based tools for the assessment of the students.	6.17	0.924	-0.868	0.012

Table 2: Descriptive statistics of the 20 items used in Lecturer survey

In this study, Cronbach's alpha internal consistency coefficient was calculated in order to determine the reliability of the scale. Table 3 and Table 4 shows the reliability coefficients of the items presented to students and lecturers respectively. Cronbach's alpha reliability coefficient normally ranges between 0 and 1. The closer Cronbach's alpha coefficient is to 1.0 the greater the internal consistency of the items in the scale. Cronbach's alpha of > .7 is considered 'Acceptable' while > .8 is considered 'Good', > .9 is 'Excellent' [33].

Factor	Number of Items	Cronbach alpha
Relative Advantage (RA)	5	0.94570738
AI Anxiety (AN)	3	0.86818892
Perceived Usefulness (PU)	4	0.95280773
Perceived Ease of Use (PEU)	3	0.912644118
Behavioural intention (BI)	3	0.953065383
Actual use (AU)	2	0.916345007

Table 3: Reliability coefficients of items presented to students

Factor	Number of Items	Cronbach alpha
Relative Advantage (RA)	5	0.856044724
AI Anxiety (AN)	3	0.703051571
Perceived Usefulness (PU)	4	0.893678161
Perceived Ease of Use (PEU)	3	0.893730547
Behavioural intention (BI)	3	0.975406872
Actual use (AU)	2	0.947680158

Table 4: Reliability coefficients of items presented to lecturers

In this study, a total of 24 lecturers across 26 classes, with about 40 students per class, and across 4 modules used the AI Chatbot for a total of 9 months. The lecturers and students were from a combination of Nitec and Higher Nitec classes. The 4 modules picked for this pilot study were modules with accountable theory tests. Only modules with theory tests were picked, as the AI Chatbot's strength lies in preparing students for theory tests through multiple practice of mock test papers, which is automatically marked by the AI's NLP, giving students immediate feedback on their given answers to both multiple choice and short answer questions. The AI's NLP attempts to mark the SAQ questions as close to how a human marker would mark. The response rate to the survey questionnaire was about 75% from lecturer respondents, and 48% from student respondents. The respondents' replies were collated and charted according to the 6 constructs and 5 relational hypotheses.

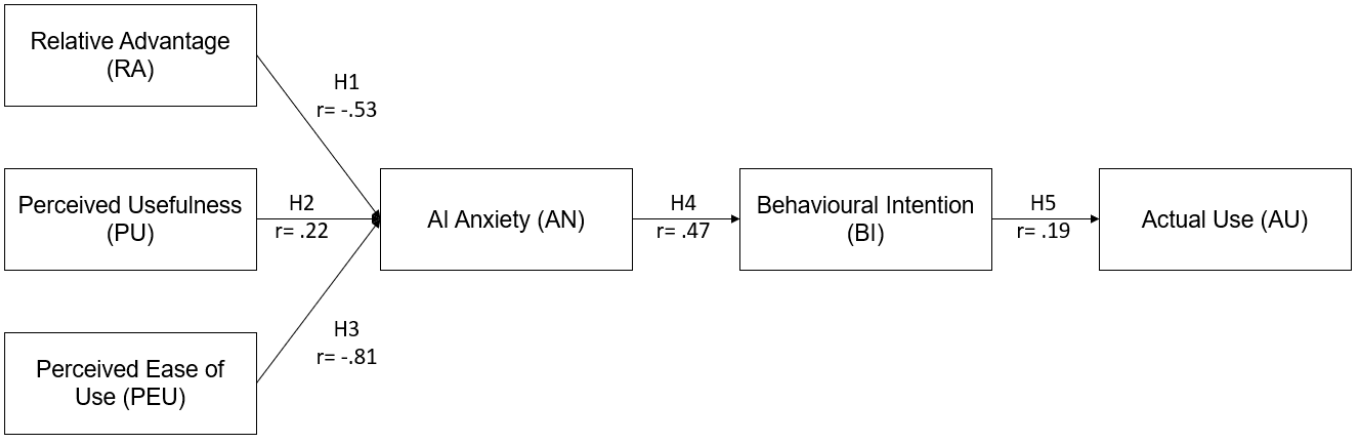


Figure 4: Correlation coefficient of the measured constructs

Analysis

H1: The constructs of RA and AN were found to be negatively correlated, r (496) = -.53, p < .02.

H2: The constructs of PU and AN were found to be a weak positive correlation, r (496) = .22, p < .01.

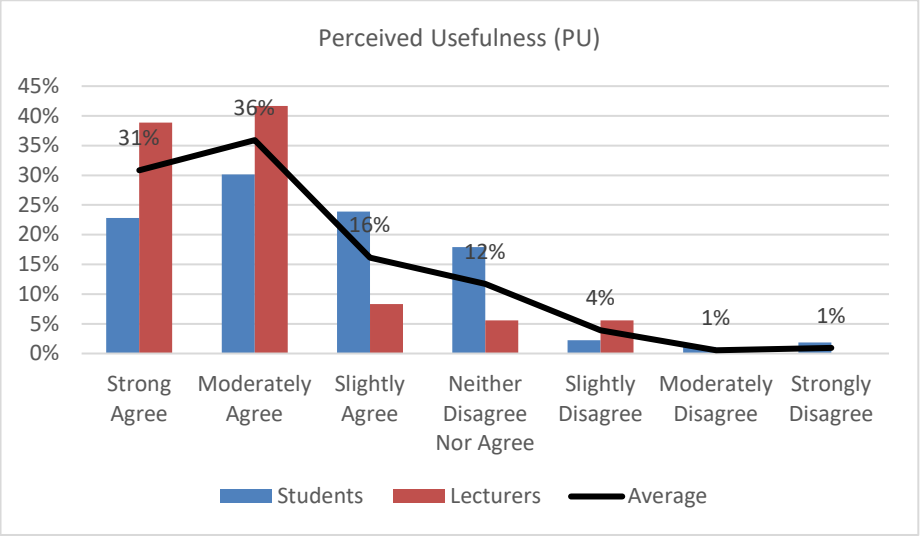
H3: The constructs of PEU and AN were found to be negatively correlated, r (496) = -.81, p < .03.

H4: The constructs of AN and BI were found to be moderately correlated, r (496) = .47, p < .01.

H5: The constructs of BI and AU were found to be a weak positive correlation, r (496) = .19, p < .05.

H1 to H3 were found to be consistent with the hypothesis which stated that the respondents' anxiety to use the AI Chatbot is negatively correlated to its relative advantage, perceived usefulness and perceived ease of use, although we see a very weak correlation for H2. It goes to show that if the respondents see its advantage and its usefulness and it is easy to use, they would generally have less anxiety in using the AI Chatbot. H4 and H5 correlation is found to be somewhat beyond the expectation when it is found that anxiety to use the AI Chatbot is positively correlated to the respondents' behavioural intention to use the AI Chatbot, albeit a moderate correlation. This may have to do with the increasing ubiquitous nature of AI technologies. While on one hand, there may still be resistance to use the technology, but the intention to use the technology could be driven by the presence of this technology everywhere. The study found no significant impact of behaviour intention on actual use with weak correlations between them. The measurement of 2 items for actual use could explain this inconsistency. Further to the statistical analysis, the findings were charted and described as follows:

Perceived Usefulness (PU)



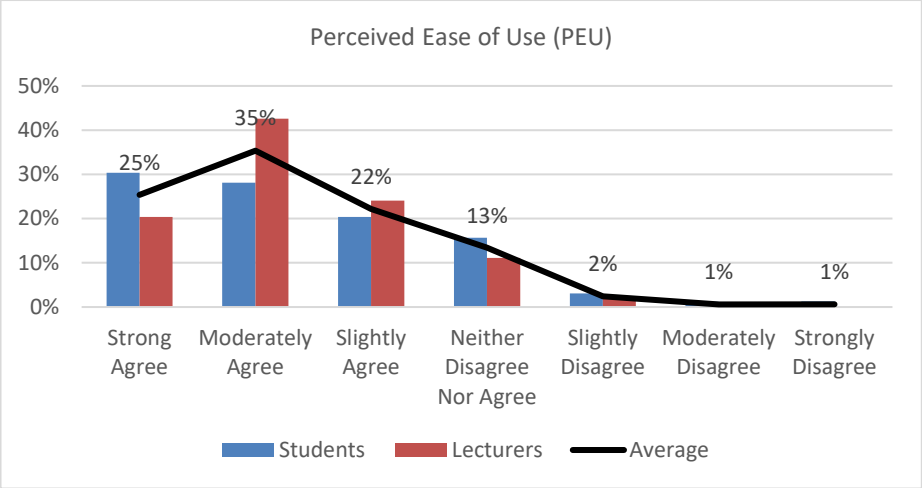
Average values derived from questions pertaining to Perceived Usefulness (PU) from Lecturers and Students

At the start of this pilot project, lecturers and students were introduced to the various features available on AI Chatbot to complement their learning in the 4 selected modules. Knowledge bases and topical tests were pre-loaded to AI Chatbot to supplement students learning at their own pace. The students and lecturers were on the hybrid approach, alternating between home-based learning and campus-based learning. AI Chatbot platform made it easier for students keep pace with their learning by conducting their own revision after the physical lessons or during the home-based learning weeks. The lecturers were able to retrieve records of the students' attempts and performance, and this helped to close the gaps of learning.

In the chart above, the positive sentiments of Slightly Agree, Moderately Agree and Strongly Agree were summed up as agreeable. Close to **90%** of the lecturers felt that the AI Chatbot have helped in efficient retrieval of the information to understand the students' learning progress and their attempts of the topical quizzes/ tests. By analysing the students' attempts and results, lecturers can map out the students' performances and evaluate the topics in which students were faring poorly. Artificial Intelligence (AI) and Machine Learning (ML) were also used in the automated marking of the topical quizzes/ tests, thus reducing the time taken for the lecturers to grade the assessments. With the time saved, the lecturers could improve their productivity in designing lessons, scale tutoring and better guide students in their learning.

About **77%** of the student respondents felt that the AI Chatbot was useful as it was available 24/7 to answer their questions instantly and consistently. This also helped to build up their learning pace and gain full control to complete the assessments to enhance the whole learning experience and improve learning outcomes.

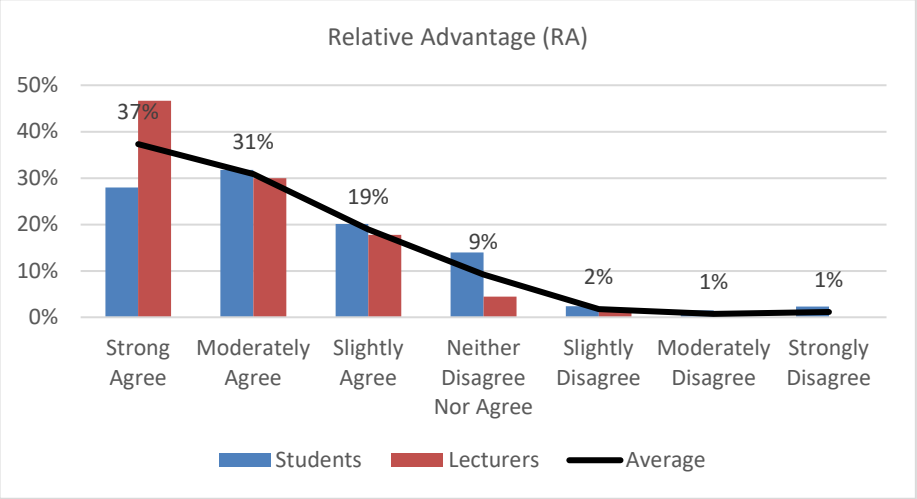
Perceived Ease of Use (PEU)



Average values derived from questions pertaining to Perceived Ease of Use (PEU) from Lecturers and Students

AI Chatbot accounts were created to facilitate the students and lecturers to access the platform. The conversational design allows the AI Chatbots to communicate with students on all the 4 selected modules (pre-loaded knowledge bases). Both the lecturers and students agreed that the AI Chatbot was easy to use, able to provide instant feedback, thus easing the whole teaching process and taking some workload off from the lecturers, resulting in good user experience in both groups of users. About **87%** of the lecturers and **78%** of the students concur with this, based on the survey results.

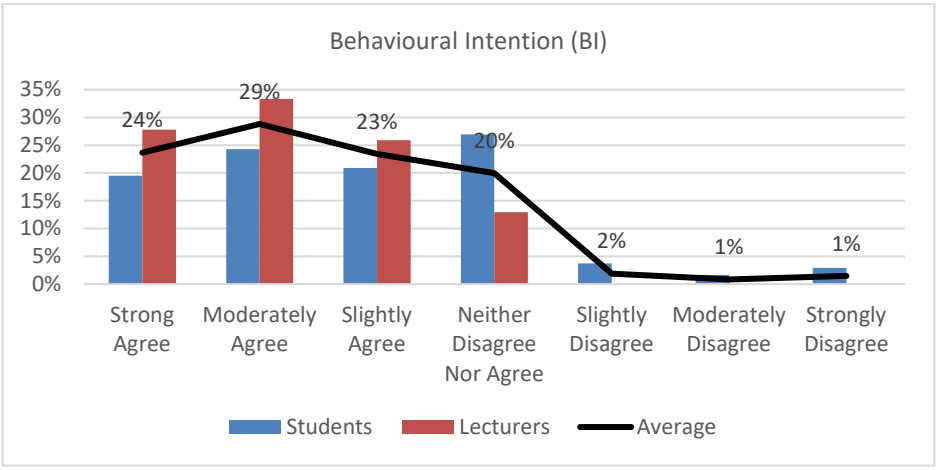
Relative Advantage (RA)



Average values derived from questions pertaining to Relative Advantage (RA) from Lecturers and Students

As the students and lecturers familiarised themselves with the knowledge-based AI Chatbot platform, the user experience was far-reaching as both groups of users were able to make use of their time more effectively. Lecturers found that they were able to conduct the marking process easily and looked through the answers to analyse and evaluate how they could reinforce the learning process. Students on the other hand felt that the platform provided excellent learning support as they could get instant feedback. Students could also alert the lecturer for help to check on their responses. About **95%** of the lecturers and **80%** of the students concurred.

Behavioural Intention (BI)

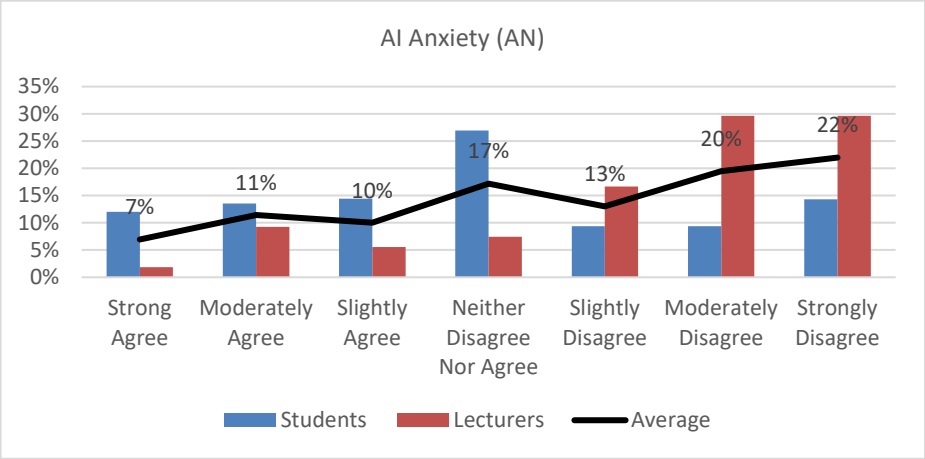


Average values derived from questions pertaining to Behavioural Intention (BI) from Lecturers and Students

The leap of digitalisation has made it easier for the students and lecturers to embrace current trends and technology. The tech-savvy students are quick to accept and learn new platforms as a medium to support their learning. Covid-19 has also accelerated the popularity of utilising Artificial Intelligence to supplement learning and assessment. With the convenience to access the virtual platform with only a mobile device and internet connectivity, education institutions are keen to continue offering online platforms as a promising reinforcement to facilitate self-directed learning, and deliver personalised learning experience.

Both groups of users had positive experiences using the AI Chatbot and this greatly reduced the resistance of using AI-based assessment. About **87%** of the lecturers and **65%** of the students would love to have such learning experience again. The advancement of technology can benefit teaching and learning. Deploying AI in education can support student learning and build in stealth assessment to help students learn better and faster. AI can transform learning to optimise the learning outcomes and deliver a superior and proactive user experience.

AI Anxiety (AN)



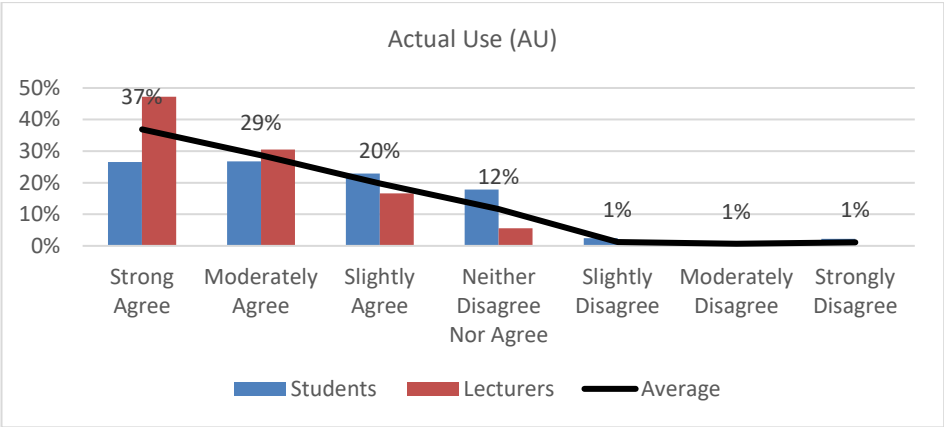
Average values derived from questions pertaining to AI Anxiety (AN) from Lecturers and Students

When it comes to technology, the efficacy of AI Chatbots is promising. However, there are potential challenges to the acceptance and use of chatbots in education by both educators and students. Lecturers lacking in digital competencies may be apprehensive to adopt and utilise the chatbot technology due to complex backend processes required to create knowledge bases and accurate content integration. Lecturers also fear that the AI Chatbot might replace their jobs (although AI is meant to relieve the lecturers' work in the classroom and not to replace the lecturers).

On the other hand, students may also worry about getting incorrect advice from the chatbot. The AI anxiety arises as the chatbot technology is still in the early stages of development and the chatbots are often confused by typos or unable to understand certain nuances of diction used by the students.

The AI feature of AI Chatbot was easy to acclimate for most of the lecturers. The students were also kept motivated and engaged with the easy usability of the learning platform. Congruent with the easy access, good learning and user experience from using AI Chatbot, the anxiety of using digital intervention had also greatly decreased during this pilot study. About **84%** of the lecturers and **69%** of the students did not feel stress or worried about using the AI chatbot.

Actual Use (AU)



Average values derived from questions pertaining to Actual Use (AU) from Lecturers and Students

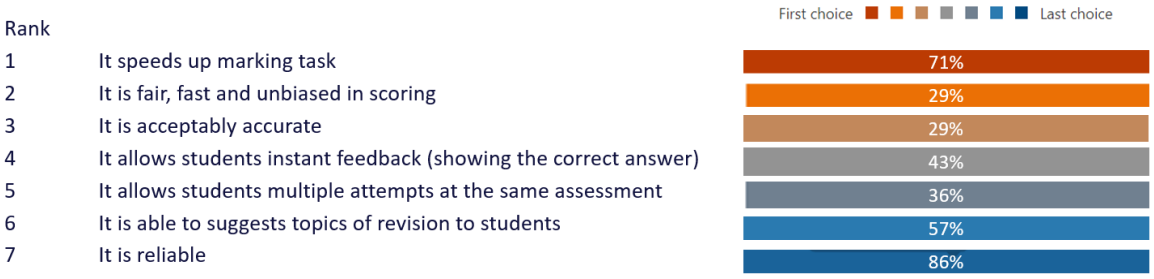
AI in education has brought about several beneficial changes, including the power to optimise and automate both learning and teaching tasks for improved learning outcomes for students and educators. The 24/7 access made quality learning accessible to all students, anytime and anywhere. The chatbot also made asynchronous communication and interaction convenient and comfortable for both students and lecturers. With the use of AI, students had personalised and accessible learning experiences and were able to enhance their skills and knowledge outside of classroom hours. The conversational intelligence of the AI Chatbot also helped students to find immediate answers to frequently asked questions. Similarly, AI Chatbot enabled lecturers to monitor and analyse the students' learning progress and address common gaps in learners' knowledge, freeing up more time to focus on lesson planning, lesson engagement and teaching core competencies.

AI Chatbot complemented their teaching / learning experiences. Lecturers were able to upload quizzes or assignments to supplement the lesson delivery and students were able to log in anytime, anywhere to access the learning portal and receive instant grading upon completion of the assessment. About **95%** of the lecturers and **77%** of the students agreed that the AI Chatbot had a positive impact on their teaching and learning experience, and would actually use them.

AI Chatbot features most preferred by lecturers

Apart from 20 Likert-scale questions, lecturers were also asked some opinion-based questions. A preferential question of the features most favoured by lecturers were asked. A list of 7 most common features of AI-driven marking were presented to the lecturers, of which they were to rank them in terms of their preferred choices of feature starting with 1 to 7.

About **71%** of lecturers ranked speeding up marking task as the most liked number 1 choice of the features of AI Chatbot. This demonstrated that the real value of automation, in speeding up manual task like marking was indeed a feature that would drive adoption of AI Chatbot up. On the other extreme, about 87% of lecturers ranked reliability as the last feature they considered when they use the AI Chatbot. That was probably due to some level of anxiety that still existed in the minds of the lecturers, on the accuracy of the marking. It could also be perhaps that reliability as a feature, should already have been addressed in any solution that offers itself as marking solution. The reliability feature can only move up the ranks if machine learning over time improves and accurately marks and truly presents itself as a solution that really mimics a human marker.

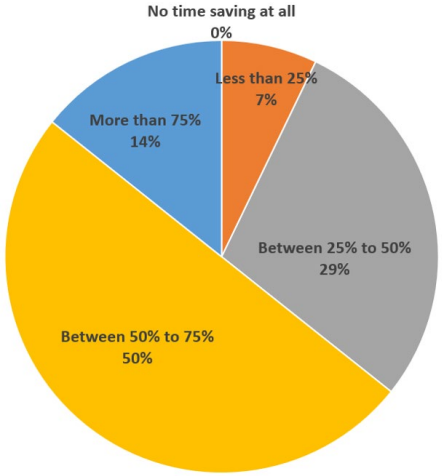


Average values derived from question pertaining Features most preferred of the AI Chatbot

Amount of marking time saved by AI Chatbot

This pilot study saw 24 lecturers use the AI – driven marking tool to mark mock test papers for their class. An average class would comprise about 40 students. In a typical mock paper, there would be 30 multiple choice and 5 short answer questions with parts. On an electronic platform, it would typically take an average of 3 hours to manually mark a softcopy mock paper in full for a class. The respondents were asked to compare that with the task of eyeballing a marked mock test paper by the Chatbot, and derived a time-saved in terms of percentage.

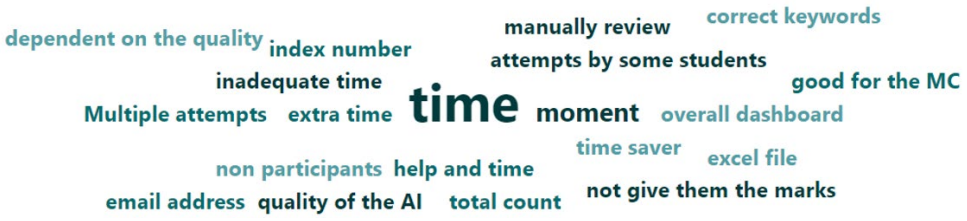
About **64%** of lecturers said they experienced lower marking fatigue with about **50% to 75%** of perceived time savings per marking process.



Average values derived from question pertaining How much marking time AI Chatbot saved

Open comments about AI - driven marking tool

Lastly, lecturers were asked to leave a voluntary open comment about the AI-driven marking tool. About **36%** of the lecturers' comments involved the word **time**. It was the most frequently occurring word submitted by the lecturers. The word time could mean time saved, or that time was needed for the machine to learn the way human marked and award marks. Whatever it is, what's certain is that time is a factor that an AI-driven marking tool could address and help lecturers address.



Open comments about AI Chatbot from Lecturers

Conclusion and Discussion

With the traction gained from the increasing availability of artificial intelligence, there is a growing need to introduce an educational technology platform to amalgamate the ease of self-directed learning and the speed of personalized feedback in the form of formative assessment. This pilot study is the very first-time lecturers and students in the School of Business & Services used the AI-driven tutoring and AI-driven marking tool, collectively known as the AI Chatbot. Hence, a set of survey questions were set based on the Technology Acceptance Model, to get a sense of the lecturers and students' intention to actually use this novel solution. Our findings revealed that in general, lecturers and students were open to adopting AI for teaching and learning purposes, in particular, for the purpose of formative learning, in personalized tutoring and immediate marking for short answer questions.

Future studies could look at designing measurements for AI Chatbots that measure the causality and not just the correlations, between time spent on the platform, frequency of practice and students' module performance. It will also be helpful if future studies could suggest if AI for education technologies can develop self-directed learning, and which level of self-directed learning can AI for education technologies help to develop, since self-directed learning is not an "all or nothing" situation, Grow (1991).

Annex A – List of survey items administered to students

List of items by construct to students.

Construct	Measurement Item	Reference
Relative Advantage (RA)	AI Chatbot would enable me to accomplish my tasks (revision) more quickly.	Moore & Benbasat, (1991); Karahanna et al. (1999)
	AI Chatbot would improve the quality of my work (revision).	
	AI Chatbot would enhance my effectiveness on the job (revision).	
	AI Chatbot would make my job (revision) easier.	
	AI Chatbot gives me greater control over my work (revision).	
AI Anxiety (AN)	In general I doubt about using AI Chatbot for fear of committing errors that I cannot correct.	Á. Hernández García. 2008.
	In general, the use of AI Chatbot causes me stress.	
	In general I feel apprehensive (worried) when using AI Chatbot.	
Perceived Usefulness (PU)	Using AI Chatbot for my assessment allows me to learn effectively.	Sánchez-Prieto, José & Cruz-Benito, Juan & Therón, Roberto & García-Peñalvo, Francisco. (2020).
	Using AI Chatbot for my assessment improves my academic productivity.	
	Using AI Chatbot increases my assessment opportunities.	
	Incorporating AI Chatbot for my assessment is useful.	
Perceived Ease of Use (PEU)	My interaction with AI Chatbot is clear and understandable.	
	I find it easy to get AI Chatbot to do what I want them to do.	
	I find AI Chatbot easy to use.	
Behavioral intention (BI)	I would participate in AI-based assessment processes.	
	I plan to participate in AI-based assessment processes often.	
	I hope to participate in AI-based assessment processes in the future.	
Actual use (AU)	Using AI Chatbot to support my learning is a good idea.	
	I have positive feelings about AI Chatbot for my learning.	

Annex B – List of survey items administered to lecturers

List of items by construct to lecturers.

Construct	Measurement Item	Reference
Relative Advantage (RA)	AI Chatbot would enable me to accomplish my tasks (marking) more quickly.	Moore & Benbasat, (1991); Karahanna et al. (1999).
	AI Chatbot would improve the quality of my work (marking).	
	AI Chatbot would enhance my effectiveness on the job (marking).	
	AI Chatbot would make my job (marking) easier.	
	AI Chatbot gives me greater control over my work (marking).	
AI Anxiety (AN)	In general I doubt about using AI Chatbot for fear of committing errors that I cannot correct.	Á. Hernández García. (2008).
	In general, the use of AI Chatbot causes me stress.	
	In general I feel apprehensive (worried) when using AI Chatbot.	
Perceived Usefulness (PU)	AI Chatbot will enable the lecturers to get the information of the students quickly.	Abu-Dalbouh, Hussain. (2013).
	AI Chatbot is useful in the rapid retrieval of information from the students.	
	AI Chatbot will save the time of lecturers.	
	Using AI Chatbot would improve my tracking students' performance.	
Perceived Ease of Use (PEU)	My interaction with AI Chatbot is clear and understandable.	Sánchez-Prieto, José & Cruz-Benito, Juan & Therón, Roberto & García-Peñalvo, Francisco. (2020).
	I find it easy to get AI Chatbot to do what I want them to do.	
	I find AI Chatbot easy to use.	
Behavioral intention (BI)	I would participate in AI-based assessment processes.	
	I plan to participate in AI-based assessment processes often.	
	I hope to participate in AI-based assessment processes in the future.	
Actual use (AU)	Using AIs to support the assessment of the students is a good idea.	
	I have positive feelings about implementing AI-based tools for the assessment of the students.	
Features of AI Chatbot	What are some of the AI-driven marking features you like best?	Sentiment analysis of users' opinion about the use of AI Chatbot
Perceived Time Saving	How much time do you think is saved for you, using this AI - driven marking tool versus manually marking each paper per class?	
Open Comments	Any other comments about the AI-driven marking tool?	

Annex C – The two sections of questionnaire administered to lecturers

Items	Continuum of agreement							Statistical Evaluation
	Strongly Disagree	Moderately Disagree	Slightly Disagree	Neither Disagree Nor Agree	Slightly Agree	Moderately Agree	Strongly Agree	
AI Chatbot would enable me to accomplish my tasks (marking) more quickly.	Relative Advantage.(RA)							Mean, Standard Deviation, Skewness, Kurtosis
AI Chatbot would improve the quality of my work (marking).								
AI Chatbot would enhance my effectiveness on the job (marking).								
AI Chatbot would make my job (marking) easier.								
AI Chatbot gives me greater control over my work (marking).								
In general I doubt about using AI Chatbot for fear of committing errors that I cannot correct.	AI Anxiety (AN)							
In general, the use of AI Chatbot causes me stress.								
In general I feel apprehensive (worried) when using AI Chatbot.								
AI Chatbot will enable the lecturers to get the information of the students quickly.	Perceived Usefulness (PU)							
AI Chatbot is useful in the rapid retrieval of information from the students.								
AI Chatbot will save the time of lecturers.								
Using AI Chatbot would improve my tracking students' performance.								
My interaction with AI Chatbot is clear and understandable.	Perceived Ease of Use (PEU)							
I find it easy to get AI Chatbot to do what I want them to do.								
I find AI Chatbot easy to use.								
I would participate in AI-based assessment processes.	Behavioral intention (BI)							
I plan to participate in AI-based assessment processes often.								
I hope to participate in AI-based assessment processes in the future.								
Using AIs to support the assessment of the students is a good idea.	Actual use (AU)							
I have positive feelings about implementing AI-based tools for the assessment of the students.								
What are some of the AI-driven marking features you like best?	Features of AI Chatbot							Mode
How much time do you think is saved for you, using this AI - driven marking tool versus manually marking each paper per class?	Perceived Time Saving							Average
Any other comments about the AI-driven marking tool?	Open Comments							Sentiment Analysis

References

[1] Singapore - OECD. (n.d.). Retrieved June 15, 2022, from https://www.oecd.org/education/talis/TALIS2018_CN_SGP.pdf

[2] BLOOM, B. E. N. J. A. M. I. N. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13(6), 4–16. https://doi.org/10.3102/0013189x013006004

[3] Bryant, J., Heitz, C., Sanghvi, S., & Wagle, D. (2021, November 11). *How artificial intelligence will impact K-12 teachers*. McKinsey & Company. Retrieved June 14, 2022, from https://www.mckinsey.com/industries/education/our-insights/how-artificial-intelligence-will-impact-k-12-teachers

[4] 2022 Educause Horizon Report: Teaching and learning edition. 2022 EDUCAUSE Horizon Report | Teaching and Learning Edition. (2022, April 18). Retrieved June 16, 2022, from https://library.educause.edu/resources/2022/4/2022-educause-horizon-report-teaching-and-learning-edition

[5] When will computer hardware match the human brain? When will computer hardware match the human brain? by Hans Moravec. (n.d.). Retrieved June 16, 2022, from https://www.jetpress.org/volume1/moravec.htm

[6] Artificial Intelligence in education: A Review. IEEE Xplore. (n.d.). Retrieved June 16, 2022, from https://ieeexplore.ieee.org/abstract/document/9069875

[7] Rotenberg, V. S. (2013). Moravec's paradox: Consideration in the context of two brain hemisphere functions. Activitas Nervosa Superior, 55(3), 108–111. https://doi.org/10.1007/bf03379600

[8] Sánchez-Prieto, J. C., Cruz-Benito, J., Therón, R., & García-Peñalvo, F. J. (2019). How to measure teachers' acceptance of ai-driven assessment in Elearning. Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality. https://doi.org/10.1145/3362789.3362918

[9] Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. Management Science, 35(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

[10] (PDF) teachers' Perspectives on Artificial Intelligence. (n.d.). Retrieved June 19, 2022, from https://www.researchgate.net/publication/337716601_Teachers'_Perspectives_on_Artificial_Intelligence

[11] Examining teachers' decisions to adopt new technology. (n.d.). Retrieved June 19, 2022, from https://www.researchgate.net/publication/220374696_Examining_Teachers'_Decisions_To_Adopt_New_Technology

[12] Silva, P. (n.d.). Davis' Technology Acceptance Model (TAM) (1989). Information Seeking Behavior and Technology Adoption, 205–219. https://doi.org/10.4018/978-1-4666-8156-9.ch013

[13] Mason, O., Grove-Stephensen, I. (2019, August 11). Automated Free Text Marking with paperless school. figshare. Retrieved June 19, 2022, from https://repository.lboro.ac.uk/articles/conference_contribution/Automated_free_text_marking_with_Paperless_School/9490169

[14] Education at a glance 2014: Indicators by Chapter. OECD. (n.d.). Retrieved June 20, 2022, from https://www.oecd.org/education/education-at-a-glance-2014-indicators-by-chapter.htm

[15] Automated Essay Scoring Versus Human Scoring: A Comparative Study. (n.d.). Retrieved June 20, 2022, from https://files.eric.ed.gov/fulltext/EJ838612.pdf

[16] Mohler, M., Bunescu, R. and Mihalcea, R. (2011) “Learning to grade short answer questions using semantic similarity measures and dependency graph alignments,” in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies 1: 752–762.

[17] Leacock, C. and Chodorow, M. (2003) “C-rater: Automated scoring of short-answer questions.” Computers and the Humanities 37 (4): 389–405.

[18] Mitchell, T., Russel, T., Broomhead, P., Aldridge N. (2002) “Towards robust computerized marking of free-text responses,” in M. Danson (Ed.), Proceedings of the Sixth International Computer Assisted Assessment Conference, Loughboroug University, Loughborouh, UK

[19] Landauer, T.K.(2003) “Automatic essay assessment.” Assessment in education: Principles, policy and practice 10 (3): 295–308.

[20] Shermis, M.D., Burstein, J., Higgins, D. and Zechner, K. (2010) “Automated essay scoring: Writing assessment and instruction.” International encyclopedia of education 4 (1): 20–26

[21] Burrows, S., Gurevych, I. and Stein, B. (2015) “The eras and trends of automatic short answer grading.” International Journal of Artificial Intelligence in Education 25 (1): 60–117.

[22] Jordan, S. and Mitchell, T. (2009) “e-Assessment for learning? The potential of short-answer free-text questions with tailored feedback.” British Journal of Educational Technology 40 (2): 371–385.

[23] Bachman, L.F., Carr, N., Kamei, G., Kim, M., Pan, M.J., Salvador, C. and Sawaki, Y., (2002) “A reliable approach to automatic assessment of short answer free responses,” in Proceedings of the 19th international conference on Computational linguistics 2: 1–4. Association for Computational Linguistics.

[24] Pulman, S.G. and Sukkarieh, J.Z. (2005)“Automatic short answer marking,” in Proceedings of the second workshop on Building Educational Applications Using NLP (pp. 9–16). Association for Computational Linguistics.

[25] Cutrone, L. and Chang, M. (2011)“Auto-Assessor: Computerized assessment system for marking student’s short-answers automatically,” in Technology for Education (T4E), 2011 IEEE International Conference on (pp. 81–88). IEEE.

[26] Shemshack, A., Spector, J. M. (2020). A systematic literature review of personalized learning terms. Smart Learning Environments, 7(1). https://doi.org/10.1186/s40561-020-00140-9

[27] Hao, K. (2022, January 10). China has started a grand experiment in AI education. it could reshape how the world learns. MIT Technology Review. Retrieved June 21, 2022, from https://www.technologyreview.com/2019/08/02/131198/china-squirrel-has-started-a-grand-experiment-in-ai-education-it-could-reshape-how-the/

[28] Saleh Zolait, A. H., & Sulaiman, A. (2008). Incorporating the innovation attributes introduced by rogers’theory into theory of reasoned action: An examination of internet banking adoption in Yemen. Computer and Information Science, 1(1). https://doi.org/10.5539/cis.v1n1p36

[29] (PDF) what is the best response scale for survey and questionnaire ... (n.d.). Retrieved June 26, 2022, from https://www.researchgate.net/publication/343994538_What_Is_the_Best_Response_Scale_for_Survey_and_Questionnaire_Design_Review_of_Different_Lengths_of_Rating_Scale_Attitude_Scale_Likert_Scale

[30] Dönmez-Turan, A., & Kır, M. (2019). User anxiety as an external variable of technology acceptance model: A meta-analytic study. Procedia Computer Science, 158, 715–724. https://doi.org/10.1016/j.procs.2019.09.107

[31] Saadé, R. G., & Kira, D. (2007). Mediating the impact of technology usage on perceived ease of use by anxiety. Computers & Education, 49(4), 1189–1204. https://doi.org/10.1016/j.compedu.2006.01.009

[32] Kline, T. (2012). Psychological testing: A practical approach to design and Evaluation. SAGE Publications, Inc.

[33] Calculating, interpreting, and reporting cronbach’s alpha reliability ... (n.d.). Retrieved August 23, 2022, from https://www.researchgate.net/publication/31591315_Calculating_Interpreting_And_Reporting_Cronbach's_Alpha_Reliability_Coefficient_For_Likert-Type_Scales