

2021

Developing an Open-Book Online Exam for Final Year Students

Keith Quille

Technological University Dublin, keith.quille@tudublin.ie

Keith Nolan

Technological University Dublin, keith.nolan@tudublin.ie

Brett Becker

University College Dublin

See next page for additional authors

Follow this and additional works at: <https://arrow.tudublin.ie/ittscicon>



Part of the [Computer Engineering Commons](#)

Recommended Citation

Quille, K., Nolan, K. & Becker, B.A. (2021). Developing an Open-Book Online Exam for Final Year Students. *26th ACM Conference on Innovation and Technology in Computer Science Education*, Virtual Event, Germany, June 26 - July 1. doi:10.1145/3430665.3456373

This Conference Paper is brought to you for free and open access by the School of Science and Computing (Former ITT) at ARROW@TU Dublin. It has been accepted for inclusion in Conference Papers by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, vera.kilshaw@tudublin.ie.

Authors

Keith Quille, Keith Nolan, Brett Becker, and Sean McHugh

Developing an Open-Book Online Exam for Final Year Students

Keith Quille

Dept. of Computing, TU Dublin
Tallaght, Dublin, Ireland
Keith.Quille@TUDublin.ie

Brett A. Becker

School of Computer Science, University College Dublin
Dublin, Ireland
Brett.Becker@ucd.ie

Keith Nolan

Dept. of Computing, TU Dublin
Tallaght, Dublin, Ireland
Keith.Nolan@TUDublin.ie

Seán McHugh

Dept. of Computing, TU Dublin
Tallaght, Dublin, Ireland
Sean.McHugh@TUDublin.ie

ABSTRACT

Like many others, our institution had to adapt our traditional proctored, written examinations to open-book online variants due to the COVID-19 pandemic. This paper describes the process applied to develop open-book online exams for final year (undergraduate) students studying Applied Machine Learning and Applied Artificial Intelligence and Deep Learning courses as part of a four-year BSc in Computer Science. We also present processes used to validate the examinations as well as plagiarism detection methods implemented. Findings from this study highlight positive effects of using open-book online exams, with ~85% of students reporting that they either prefer online open-book examinations or have no preference between traditional and open-book exams. There were no statistically significant differences reported comparing the exam results of student cohorts who took the open-book online examination, compared to previous cohorts who sat traditional exams. These results are of value to the CSEd community for three reasons. First, it outlines a methodology for developing online open-book exams (including publishing the open-book online exam papers as samples). Second, it provides approaches for deterring plagiarism and implementing plagiarism detection for open-book exams. Finally, we present feedback from students which may be used to guide future online open-book exam development.

CCS CONCEPTS

• **Social and professional topics** → **Student assessment.**

KEYWORDS

Online exam; Open-book; Artificial Intelligence; Machine Learning

ACM Reference Format:

Keith Quille, Keith Nolan, Brett A. Becker, and Seán McHugh. 2021. Developing an Open-Book Online Exam for Final Year Students. In *26th ACM Conference on Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2021), June 26–July 1, 2021, Virtual Event, Germany*. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3430665.3456373>



This work is licensed under a Creative Commons Attribution International 4.0 License.

ITiCSE 2021, June 26–July 1, 2021, Virtual Event, Germany.

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8214-4/21/06.

<https://doi.org/10.1145/3430665.3456373>

1 INTRODUCTION

March 2020 saw the beginning of the COVID-19 pandemic, leading to a host of changes in the delivery and assessment of academic programmes. Where possible all work and educational programmes moved to online formats using tools such as Microsoft Teams, Zoom, WebEx, Slack and a myriad of others. All sectors experienced significant change, and the education sector was not alone in facing unique, unprecedented challenges. Lecturers were forced to deliver content and assessments online.

Within computing disciplines, the initial move to online delivery was perhaps easier than it was for many other disciplines. Lecturers had the option to pre-record lectures, provide supplementary material and host live streams of lectures where the lecturer could fairly easily employ techniques such as live coding, arguably providing students with lectures that were somewhat comparable to traditional experiences. However, recent evidence showed that the shift to online learning led to changes in student sense of belonging for computing students [18]. It is also likely that assessment was an aspect of online delivery that computing disciplines faced challenges in.

In the assessment of course work, especially for high-stakes final year undergraduate courses such as, the courses presented in this paper, the generally accepted method to summative end-of-semester assessment is closed-book, proctored exams. Such closed-book assessments encourage students to “predict” what questions might appear on the exam, which ultimately encourages rote, surface learning (lower-order questioning). This system is still in place perhaps due to the slow pace of change in large institutions. Nonetheless, the pandemic quickly forced assessment to take a dramatically different form. However, the need to provide assessment instruments that fairly and validly assesses student knowledge (employing high standards of academic integrity) in an open-book format was still a desired requirement.

2 RELATED WORK

2.1 Closed- vs Open-Book Assessment

Closed-book proctored assessment is a widely accepted approach for assessing how a student can demonstrate gained knowledge, with no additional material [31, 32]. Closed-book assessments are often invigilated to ensure there is no extra material being utilised by students, maintaining academic integrity and fairness. Traditional exams in computing typically assess multiple concepts in one item and have high interdependence between them, which can

be difficult to assess and likely work against approaches such as mastery learning [14].

On the other hand, open-book assessments allow students to consult course material during the assessment and can be proctored or not [34]. This is less conventional, partly due to issues preventing widespread adaption such as students spending too much time looking up texts and formulating or honing their answers. There is also a general consensus that students will spend less time studying prior to the assessment and so a reduction in preparedness for the assessment occurs [4, 30, 32]. This could possibly damage long-term retention of concepts.

2.2 Student Experiences of open-book Assessment

While the literature appears to be mixed in terms of the benefits of open-book assessment over closed-book assessment, for students, open-book assessments are most often seen as a benefit.

It is known that computing students, like others, are susceptible to mental health issues during their course, with anxiety (including test anxiety) and stress becoming a wider concern of late [11, 19–23, 26–28]. Deloatch et al. investigated how exam modality relates to students' perceptions of test anxiety and performance during programming exams [9] with a survey administered to measure student perception of test anxiety experienced with paper-based exams and online exams. Of 391 students participating, 22% ($n=61$, $\bar{x}=4.26$, $SD=1.51$) perceived high test anxiety for paper-based exams while 23% ($n=64$, $\bar{x}=4.15$, $SD=1.67$) experienced high test anxiety for online exams.

De Raadt proposed a method of allowing students to create 'cheat sheets' for exams. With 89 students taking part, exam marks did improve marginally when allowed a cheat sheet, and students that utilised a cheat sheet reported reduced levels of test anxiety before and during the exam [8]. It is also possible that creating cheat sheets reinforces learned material and aids retention.

Through research investigating enhancing student performance through the use of open-book assessment, Green et al. found that open-book assessment can be instrumental in strengthening understanding in both cooperative learning and more traditional classrooms [12]. Following the assessment, students reported that being able to reference material during the assessment reduced feelings of stress and anxiety. However, they did report that timing was an issue.

2.3 Development of open-book Assessments

It should be possible to design open-book assessments with higher-order questions to test critical thinking and higher-order skills [2]. While traditional assessments are designed around student recall, open-book assessments encourage educators to ask questions that go beyond the general cognitive levels of recall. While there is no definitive guide to creating open-book assessments in computing, the literature is consistent on the benefits of using Bloom's Taxonomy to develop higher-order questions [5, 12, 13].

The original Bloom's Taxonomy has six classes (acknowledging multiple variants of the taxonomy) which are widely recognised and have been the stimulus for changes within the education sector including assessment. The progression within the taxonomy goes

from knowledge-based (lower-order questioning typical of closed book assessments) to evaluation-based behaviours (higher-order questioning more typical of open book assessments). Other research has used Socratic questioning with Bloom's Taxonomy to develop questions [25]. Through incorporating the Socratic Method and Bloom's Taxonomy the aim is to enhance online discussion, critical thinking, higher-order questioning and student learning [33].

3 COURSES OVERVIEW

Our work focussed on two courses (Applied Machine Learning and Applied Artificial Intelligence & Deep Learning) with three student cohorts. The rationale for the selection of these two courses were that they included final year high-stakes examinations. Other courses were considered (such as CS1, Advanced Routing and Switching and Enterprise Database Technologies) however these typically consisted of large components of continuous assessment reducing the value and the stakes of the examination component, if any existed. The students sitting open-book exams within these two courses consisted of both part-time (PT) and full-time (FT) students. Full-time students in this study are students who were majoring in Computing with Software Development. Details on these courses and the student cohorts are discussed in the following subsections. All student cohorts were in their final year of study, at Hons Bachelor Degree level in spring of the 2019-20 academic year (just as COVID-19 forced a national lockdown).

3.1 Applied Machine Learning

The Applied Machine Learning (AML) course descriptor [29] is based on the following learning outcomes:

- (1) Apply data pre-processing and data exploration techniques in the context of the machine learning process.
- (2) Demonstrate knowledge of machine learning techniques, their methods and application.
- (3) Determine the machine learning techniques and methods for particular scenarios.
- (4) Evaluate the models produced, using relevant performance metrics.

The programming language used is Python, with Azure Notebooks as the IDE, which is a cloud based Jupyter notebook [17]. Assessment is broken down into two in-class assignments, which require students to explore a dataset and carry out appropriate data pre-processing techniques for the machine learning process. In addition, students are given a pen and paper assessment where they are provided a particular dataset or problem, with model outputs and performance measures (such as a confusion matrix), from which they must interpret results and produce findings, including statistical testing. The two assessments have a combined weighting of 50% with the end of term exam worth the remaining 50%. This course was delivered in the fall and spring semesters, where only PT students who sat the exam in the spring semester (when COVID-19 resulted in open-book assessment) are included in this study for AML. The PT students took both the AML and the Applied AI & Deep Learning course in spring in block form.

3.2 Applied AI and Deep Learning

The Applied AI and Deep Learning (AAIDL) course descriptor [29] is based on the following learning outcomes:

- (1) Discuss and explain the general principles of artificial neural networks and deep learning networks.
- (2) Distinguish between and select appropriate hyper-parameters for training artificial neural networks.
- (3) Demonstrate an understanding in determining performance of artificial neural networks.
- (4) Apply artificial neural networks and deep learning techniques to several contexts.

This course is delivered only in the spring semester. This course is also taught with Python, using Azure Notebooks. The assessment for the module is broken down into two parts: (1) one in class pen and paper assignment involving a forward pass and one back-propagation of an artificial neural network and (2) one elapsed (take home assessment completed over a time period) assessment where students are given a large dataset and must develop an appropriate deep artificial neural network. They must interpret the results and produce findings. This included hyper-parameter tuning and a grid search. The two assessments have a weighting of 25% each, with an end of term exam, with a weighting of 50%. This course was delivered in the spring to both FT and PT students who sat the online open book exam.

4 EXAM DEVELOPMENT METHODOLOGY

In our institution examination papers are submitted many months prior to the examinations. In this case, before COVID-19 forced the University to move to online delivery, the examination papers for AML and AAIDL were already created for a traditional exam setting. The exam had a time limit of two hours. This time limit is a University-wide standard. There are exceptions available for SEN (Special Educational Needs) students where an additional time of 20 minutes is applied with additional supports if required.

The institutions management team conducted a review of useful resources for developing open-book online examinations (in response to the COVID situation that led to online assessments) and recommended the work of the Centre of Teaching and Learning at the University of Newcastle, Australia [13] to all academic staff. This work presented two resources for developing open-book questions: One based on Bloom’s Taxonomy [1], and the other based on Socratic Questioning [25]. Bloom’s taxonomy typically has several levels of objectives that increase in complexity or specificity – these levels become more higher-order questioning as they progress. Examples of these levels (from lower order to higher order questioning, as included in the Newcastle guide) are: Remember, Understand, Apply, Analyse, Evaluate and Create [1]. Socratic questioning targets higher-order thinking and complex concepts, and is comparable to some of the higher-order questions from Bloom’s Taxonomy. Based on the work of Paul [25], there are six question categories: questions for clarification, questions that probe assumptions, questions that probe reasons and evidence, questions about origins or sources, questions that probe implications and consequences and finally viewpoint questions. All categories were referenced in the Newcastle guide [13].

One of the main goals for creating our first set of open-book online exams was transparency of the process and methodology. Therefore we used the pre-written traditional exams as a starting point in terms of topics (not specific questions), to provide evidence of the mapping from traditional closed-book examinations to an open-book online examination using the Newcastle guide as the mapping tool. This mapping was presented to students for exam preparation and to external examiners who validate our assessments. The traditional AML and AAIDL exam papers are published online [29]. The majority of the traditional exam questions were appropriate for adaption to open-book online learning. An example of a closed book question deemed not appropriate for adaption, was Q1 from the the AAIDL examination paper. This question involved an artificial neural network, with weights presented (Figure 1) and two input values (In_1 and In_2). Students have to conduct a forward pass (where the activation functions A_1, A_2, A_3 and \hat{y} are provided) and then update the weights using one back-propagation pass.

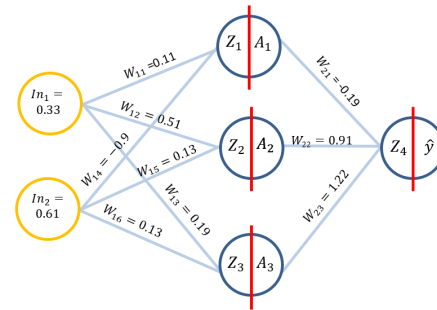


Figure 1: AAIDL Traditional Exam Paper Q1

As all students (if correct) would arrive at the same answer, there was a possibility of undetected plagiarism in an online format. Unlike in a proctored exam setting, students would have access to a Python script (which was part of a prior lab assignment) that could generate the solution swiftly, even if each student was given unique weight values. While AAIDL Q1 was not suitable for the above reasons, the majority of the traditional questions were suitable for adaption/mapping into open-book online format. When mapping the examination questions to open-book, special efforts were made to adapt questions to higher-order questions using Bloom’s taxonomy and suitable mapping questions from the sample Socratic questions, both from the Newcastle guide. The final open-book online exam papers are published online [29] with some sample mappings provided next:

Example 1 (AAIDL Q2b): Explain Learning Rate Decay, and why it would be used over a normal learning rate. **mapped to:** “Learning Rate Decay performs better than using a set learning rate value” Is this statement true? By what reasoning did you come to that conclusion?

Example 2 (AAIDL Q3b): Explain how a convolutional layer works, in a Convolutional Neural Network (CNN). Aid your answer with an example/diagrams, using a 5x5 pixel image as the input. **mapped to:** “For image classification of faulty products on a production line, a Convolutional Neural Network (CNN) always has significantly stronger

performance than using a dense Artificial Neural Network (ANN) Is this a true statement (explain your reasoning)? Explain what additional information you might need about the problem situation, to agree or disagree with this statement.

Example 3 (AML Q2b): *Explain in detail with the aid of a diagram, the 10 fold Cross Validation technique. mapped to: Ten-fold Cross Validation is often referred to as the “Gold Standard” of Machine Learning model validation techniques (the best technique to use). Explain what is the most important feature of this technique, that makes it the “Gold Standard”? Explain an alternative to Ten-fold Cross Validation? Compare and contrast the two techniques (10-fold Cross Validation and the alternative technique), giving examples of problem situations where each technique may be more suitable.*

Prior to the open book examination, all exam papers were inspected by an external examiner to validate if the exam papers used met the learning outcomes of the two modules given their adaptation. The feedback was very positive and no issues were identified regarding open book assessment or that any learning outcomes from the modules were not met.

5 EXAM DELIVERY AND PLAGIARISM

5.1 Exam Delivery Methods

A priority when delivering the open-book online exam was the format, and how it would be delivered. We took a student centred approach, with an open discussion with the students prior to submitting our format to the University for approval. As this was a first for our institution, varying approaches were implemented by different lecturers, ranging from making the exam an elapsed piece of work, to using a quiz (the VLE used in our institution is Moodle). All of the approaches implemented had to be submitted and approved by the Head of Department and/or the Academic Council.

The method for delivery that we selected for the AML and AAIDL courses was an open-book online exam. Based on student concerns voiced during the open discussion, the duration of the exam was extended by 30 minutes to allow for uploading and navigating the exam itself. Additionally, we reduced each question from a typical four parts per question to three (to further alleviate time concerns). This was in addition to the regular format of choose any three of four questions from the traditional exam. A sample paper was also created using an open-book online exam format for students to use as revision (this was adapted from a previous sample closed book paper, where both papers were presented to the students). The Newcastle guide was also provided to students and the mapping explained, so students' could map historical (traditional) exam paper questions for additional revision examples. This too was welcomed by students. Finally, several study/revision sessions were provided to demonstrate the mapping process using the sample paper and to allow students practice the open book questions.

Before and during the exam, the lecturer was available using MS Teams if any questions arose. Additionally, students were asked to keep their audio on as, if any questions arose, the response could be announced to the class if required. The examination was provided in PDF format, which could be downloaded in case of any internet or technology issues. In some cases students printed it. In addition, the exam paper was opened for viewing 10 minutes prior to the exam. During this ten minutes, each question from the exam paper

was read out to the students via MS Teams. While this was primarily to aid SEN students (including the printing of the paper), the overall feedback was very positive on this, from all students.

SEN students had an additional 20 minutes for the examination. In addition, an optional session was provided prior to the exams on using Office 365 accessibility tools such as screen readers and dictation to further prepare students. The majority of students (SEN and Non-SEN) availed of this.

5.2 Plagiarism

One of the department's primary concerns for academic integrity was plagiarism during open-book online exams. One of the initial approaches developed was centred around academic honesty principles. This consisted of students signing (virtually) the University wide plagiarism policy, coupled with a session during each revision class about the importance of understanding the policy and what it means (both for consequences and ethically). The use of highlighting policy was in some cases found to reduce the amount for plagiarism in computer science courses [15, 16]. Another proactive (and perhaps also reactive) approach was the inclusion of an exam viva after the examination. This consisted of a ten minute viva style session with 20% of each exam cohort. The students were randomly pre-selected prior to the exam (but were not told until after the exam to avoid additional stress or anxiety). The pre-selected students took part in the viva straight after the exam. The students were not assessed on the correctness of their answers but were asked Socratic questions such as, "Where did you get that idea?" or "By what reasoning did you come to that conclusion?". Students were provided details about the process prior to the exam. With GDPR compliance, students selected for the post exam viva were asked for their consent for recording of the session. 100% of the students opted for this. This was a positive response, and additional future work could unpack the underlying reasons for opting in for this. No students were identified as plagiarism concerns based on the viva responses. Finally Urkund (<https://www.orkund.com/>) plagiarism detection software was used for the final student uploads which were in word or PDF format. This tool reports the percentage of plagiarism and for each instance, the source of the plagiarism. In some cases the tool reported a relatively high amount of plagiarism, but on further investigation, several students copied the exam questions (perhaps for a placeholder or guide) and this resulted in a higher plagiarism score. In Urkund you can manually remove these cases from the scores. This may be a useful insight for future development of online open-book exams using plagiarism software. No students had a concerning plagiarism score (greater than 25%) after manual processing. This, coupled with the viva also finding no plagiarism concerns, was a positive outcome.

6 STUDENT RESULTS

The results of the previous cohorts (2018-19 academic year) who sat the traditional closed-book written examinations were compared to the cohort of students who sat the open-book online exam format (spring of the 2019-20 academic year). This section presents an investigation to determine if any significant differences exist. The overall average result, the standard deviation and number of students in each course are presented in Table 1. Next a Welch's *t*-test [10] was used to investigate if any differences between the

traditional and open-book exam results were statistically significant. A Welch's t -test is used to test the hypothesis that two populations with unequal variance and/or sample sizes have equal means, and for this reason was selected over a Student's t -test. For each result presented in Table 1, a Welch's t -test was conducted using a confidence interval of 95%. While the average results are lower for each cohort completing the online exam, none of the differences were statistically significant (where each p -value was > 0.05).

Year	Mode	Course	Avg %	Std %	N
2019	PT	AML Traditional	60.54	13.07	14
2020	PT	AML Online	54.29	12.85	17
2019	PT	AAIDL Traditional	68.70	9.51	7
2020	PT	AAIDL Online	60.20	11.65	15
2019	FT	AAIDL Traditional	60.21	11.63	27
2020	FT	AAIDL Online	56.00	15.98	15

Table 1: Comparison of traditional and online exam results. No significant differences were found between courses.

7 STUDENT EXPERIENCES

After each exam students completed an online survey that consisted of 9 questions (with Q4 having 9 Likert sub-questions) for a total of 17 data points where all questions are presented in the following sections. The survey took an average of 7.5 minutes to complete. The primary aim of this survey was to capture experiences (positive and negative) for future development of open book exams. 27 of 47 students responded (~60% response rate). The survey contained both quantitative (Q1-4) and qualitative (open ended) questions (Q5-9). Next, we present a summary of these responses. For the qualitative survey questions, the authors present responses that may be of value to other CS educators when developing open-book examinations (both positive and negative experiences). The responses are presented on a per survey question basis.

7.1 Quantitative

7.1.1 Q1: What exam was this survey for. This question was to gauge what students responded to the survey for which exam. Applied AI and Deep learning had 20 responses and Applied Machine Learning had 7 responses.

7.1.2 Q2: Was this your first ever open-book exam. The authors hypothesised that the vast majority of students may not have sat an open-book online exam previously. However, survey responses reported otherwise, with 9 students (33%) reporting that they had sat an open-book exam previously. 18 students reported it was their first experience.

7.1.3 Q3: Would you prefer (if given the choice, under normal circumstances). this question asked students if they would prefer; closed-book exams (traditional), open-book online exams (like this exam), either (no preference), or other. Positively, ~85% of students reported that they would prefer an online open-book exam or have no preference either way. The one student who selected other said: "Prefer to see my grade before deciding".

7.1.4 Q4: How would you rate the following for the open-book exam experience. Question four had nine Likert questions to gauge student experiences. These questions and their responses are presented

in Table 2 where the results are largely positive. For exam questions, one student reported fair, with four students reporting good and 22 students reporting excellent. The exam time limit was a concern for students, even with the additional 30 minutes and reduction in question sub-parts. Almost 25% of students reported average, fair or poor for this. Also, one student reported fair for the use of Urkund. Cronbach's alpha (α) is a measure of internal consistency [3, 6, 7] and was investigated to determine the reliability, or internal consistency of the students responses to the Likert scale. α ranges between 0 and 1, where an α value of 0.7 or above is appropriate for most early-stage research studies [24]. α for the Q4 Likert questions was 0.73, which represents 'good' internal consistency or reliability of the student responses.

Survey Question	0	1	2	3	4
Clarity of instruction for the entire process	0	0	0	3	24
Online process (teams and moodle for instructions and upload)	0	0	0	1	26
Exam Questions (structure, wording and clarity, please ignore difficulty)	0	1	0	4	22
Exam time limit, including additional time for uploading	2	2	2	4	17
Upload Procedure	0	0	1	4	22
The use of Urkund Plagiarism	0	1	6	7	13
The use of a VIVA for plagiarism purposes	0	0	5	14	8
Lecturer availability prior to exams for questions	0	0	0	1	26
Lecturer availability during the exam for questions, and in case issues arise	0	0	0	1	26

Table 2: Student experiences for open-book online exams and corresponding number of students selecting each likert choice. 0: Poor, 1: Fair, 2: Average, 3: Good, 4: Excellent

7.2 Qualitative

7.2.1 Q5: Did you experience any difficulties during the online exam (technical, etc)? Twenty students reported no difficulties at all. The difficulties that were reported varied, however a theme of timing issues arose, as indicated by the following examples:

- "Yes, issues locally with laptop. Had to switch laptops and email study docs. Took some time."
- "When reviewing the answers I type I realised that some sections of one or two answers were missing. I am not sure if this was my typing skills not keeping up with my thought process or if I had accidentally deleted / overwritten some parts"
- "I struggled with timing and having to type out the questions. in written exams I usually get to attempt all questions but with this one, I struggled to get the required questions done on time."

7.2.2 Q6: Any positive feedback re online exams? This was to examine any positive feedback. An interesting pattern emerged where students cited less stress/anxiety. A second pattern indicated a fairness element, where students preferred not having to memorise as part of studying. Some responses are presented here:

- "The open-book format for the online exams worked well and reduced the overall stress of the exam"
- "More relaxed doing at home. Less stress overall. open-book exams still require you to understand the topic but does not rely on memorising large chunks of information."

- *"I think the open-book-style questions put much less focus on memorisation, which is good"*
- *"The biggest positive I found was it removed stress because you can work in a comfortable environment and the fact I didn't have to take as much time off work or have to commute to the college for exams also meant I didn't have to worry about transport, traffic conditions, finding my seat etc."*
- *"It does change the concept of the exams, from a closed-book is usually memorise the information. In this open-books the most important is understand what are you talking. I think that its better in this type of module."*

7.2.3 Q7: *Any negative feedback re online exams.* Twenty students reported no negative experiences. For those who did respond, two themes emerged, time and the environment at home:

- *"I was racing for time at the end, I am never stuck for time in an exam and I always attempt all the questions, but I only had time for three"*
- *"Should be allowed to write the answers out and scan it for the upload. Unfair on people who type a lot slower than others as its hard to give the same quality of answers then."*
- *"I'm faster at writing than typing so this was a challenge to get all I wanted typed in the time frame provided."*
- *"Hard to concentrate with partner at work and kids at home"*
- *"You have less control over your surroundings re outside noise, I had kids playing in neighbours garden and someone cutting grass during the exam."*

7.2.4 Q8: *Any comments on the Lecturer, positives, or to improve upon with respect to online exams.* There were no negative responses to this question, and while many responses were positive about lecturer effort, one response was notable as singling out the reading of the exam questions prior to the exam as a benefit:

- *"Reading out the questions at the start helps a lot. being there for questions take away some of the unease about what if there are any problems"*

7.2.5 Q9: *Any final comments you would like to add.* Two students commented on the perceived difficulty of the examinations due to the open-book nature.

- *"The exams feel way harder when they are open-book and with the format having to be changed (no calculations for example)."*
- *"I've generally found open-book exams to be more difficult because of this because a lot of time is wasted checking stuff you have written or think you know."*

Other responses were positive but varied.

8 THREATS TO VALIDITY

This study presented potential insights for developing future open-book examinations in computing courses. However several threats to validity should be noted: First, the sample size was relatively small. Second, the authors acknowledge the extensive academic debate over the applicability of Bloom's Taxonomy and the multiple versions that exist, where we were bound by the University recommendation of the Newcastle guide. Thus future work could investigate additional techniques for mapping closed book assessments to open book assessments. Finally, only highlights of the qualitative

data was presented. Efforts were made to present balanced qualitative responses that might provide insights to educators developing open-book assessment (both positive and negative student experiences). Future work could unpack these further with more detailed analysis such as encoding the responses and the identification of key themes, and the validity of the questions themselves.

9 CONCLUSIONS AND LESSONS LEARNED

This paper presents an approach for adapting traditional in-class examinations into online open-book examinations. When developing the open-book exam, transparency of the mapping (based on the Newcastle guide [13]) was achieved by publishing both the traditional exam paper as well as the online open-book exam paper. Both the mapping and methodology may be of value to educators developing similar open-book online examinations. The plagiarism approaches such as the post examination viva seemed to have been successful with no cases of plagiarism found. The students' experiences on a whole were very positive, where almost ~85% of students reporting that they would prefer an online open-book examination or have no preference between an open-book exam and traditional exam. Several students also reported less stress and pressure when sitting the online open-book exam. In addition students acknowledged that the focus was more on understanding rather than memorisation, which they noted as positive. The most common concern noted by students was the time allowed for the exam, which agrees with the literature [12], even with an additional 30 minutes and a reduction in the number of sub-questions applied. Some students also reported that they felt the exam was more difficult, however the difference in exam results were not statistically significant when compared with the previous student cohort who sat a traditional closed-book examination. It is worth noting, while a higher demand was placed on students, a higher demand is also placed on educators when teaching open book material and marking of exams. In addition, upon external review (as per institution processes) the external examiner felt that the open-book online examinations met the learning outcomes of the courses. Overall this paper may be of value to the CSEd community, as the process, methodologies and lessons learned could provide a foundation for the development of future open-book examinations.

For the authors who are facing another two semesters of online open book assessment, some valuable lessons learned have been identified which will be implemented this semester (future work will be conducted to investigate if they are of value to the students). First, discussions have begun to increase further the allotted time for assessment to reduce student anxiety. Second, the pre-exam preparation classes discussing the open book assessment format and the mapping used will be repeated as anecdotally students reported this was a key element in preparation for the assessment. Third, students will be informed that the previous cohort of students taking the open book assessment did not perform worse than their predecessors taking the traditional closed book examination. This is in an effort to reduce the perception that open book assessments are more difficult than closed book assessments. Finally the efforts made to preserve academic integrity have now be adapted across our school, based on the work in this study, especially the post exam viva, and this will again be deployed for future open book assessments, and not just examinations.

REFERENCES

- [1] Lorin W Anderson, Benjamin Samuel Bloom, et al. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- [2] Caroline Baillie and Susan Toohey. 1997. The 'Power Test': its impact on student learning in a materials science course for engineering students. *Assessment & Evaluation in Higher Education* 22, 1 (1997), 33–48. <https://doi.org/10.1080/0260293970220103> arXiv:<https://doi.org/10.1080/0260293970220103>
- [3] B Becker and Orla Butler. 2011. Learner-to-Educator Feedback–Acquiescence Bias, Reliability and Learner Opinion. In *International Conference on Engaging Pedagogy*.
- [4] David Boniface. 1985. Candidates' use of notes and textbooks during an open-book examination. *Educational Research* 27, 3 (1985), 201–209.
- [5] Dr Richard Brightwell, Ms Janine-Helen Daniel, and Dr Angus Stewart. 2004. Evaluation: is an open book examination easier? *Bioscience Education* 3, 1 (2004), 1–10. <https://doi.org/10.3108/beej.2004.03000004> arXiv:<https://doi.org/10.3108/beej.2004.03000004>
- [6] Lee J Cronbach. 1951. Coefficient alpha and the internal structure of tests. *psychometrika* 16, 3 (1951), 297–334.
- [7] Lee J Cronbach and Paul E Meehl. 1955. Construct validity in psychological tests. *Psychological bulletin* 52, 4 (1955), 281.
- [8] Michael de Raadt. 2012. Student Created Cheat-sheets in Examinations: Impact on Student Outcomes. In *Proceedings of the Fourteenth Australasian Computing Education Conference (ACE '12)*. Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 71–76. <http://dl.acm.org/citation.cfm?id=2483716.2483725>
- [9] Robert Deloatch, Brian P Bailey, and Alex Kirlik. 2016. Measuring Effects of Modality on Perceived Test Anxiety for Computer Programming Exams. *SIGCSE '16 Proceedings of the 47th ACM Technical Symposium on Computing Science Education* (2016), 291–296. <https://doi.org/10.1145/2839509.2844604>
- [10] Ben Derrick, Deirdre Toher, and Paul White. 2016. Why Welch's test is Type I error robust. *The Quantitative Methods in Psychology* 12, 1 (2016).
- [11] Barbara Dooley, clíodhna o'connor, Amanda Fitzgerald, and Aileen O'Reilly. 2019. *My World Survey 2: The National Study of Youth Mental Health in Ireland*.
- [12] Tor Vidar Eilertsen and Odd Valdermo. 2000. Open-book assessment: A contribution to improved learning? *Studies in Educational Evaluation* 26, 2 (2000), 91–103. [https://doi.org/10.1016/S0191-491X\(00\)00010-9](https://doi.org/10.1016/S0191-491X(00)00010-9)
- [13] Centre for Teaching and Learning. [n.d.]. *A guide for academics - A guide for open book exams, available at https://www.newcastle.edu.au/_data/assets/pdf_file/0006/268980/Open-Book-Exams.pdf*.
- [14] Andrew Luxton-Reilly, Brett A. Becker, Yingjun Cao, Roger McDermott, Claudio Mirolo, Andreas Mühlhling, Andrew Petersen, Kate Sanders, Simon, and Jacqueline Whalley. 2018. Developing Assessments to Determine Mastery of Programming Fundamentals. In *Proceedings of the 2017 ITiCSE Conference on Working Group Reports (ITiCSE-WGR '17)*. Association for Computing Machinery, New York, NY, USA, 47–69. <https://doi.org/10.1145/3174781.3174784>
- [15] David J. Malan, Brian Yu, and Doug Lloyd. 2020. Teaching Academic Honesty in CS50. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20)*. Association for Computing Machinery, New York, NY, USA, 282–288. <https://doi.org/10.1145/3328778.3366940>
- [16] Tony Mason, Ada Gavrilovska, and David A. Joyner. 2019. Collaboration Versus Cheating: Reducing Code Plagiarism in an Online MS Computer Science Program. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19)*. Association for Computing Machinery, New York, NY, USA, 1004–1010. <https://doi.org/10.1145/3287324.3287443>
- [17] Microsoft. [n.d.]. *Azure Notebooks*. <https://notebooks.azure.com/>.
- [18] Catherine Mooney and Brett A. Becker. 2021. Investigating the Impact of the COVID-19 Pandemic on Computing Students' Sense of Belonging. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE '21)*. Association for Computing Machinery, New York, NY, USA, 612–618. <https://doi.org/10.1145/3408877.3432407>
- [19] Keith Nolan and Susan Bergin. 2016. The Role of Anxiety when Learning to Program: A Systematic Review of the Literature. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research (Koli Calling '16)*. ACM, New York, NY, USA, 61–70. <https://doi.org/10.1145/2999541.2999557>
- [20] Keith Nolan, Susan Bergin, and Aidan Mooney. 2019. An Insight Into the Relationship Between Confidence, Self-efficacy, Anxiety and Physiological Responses in a CS1 Exam-like Scenario. In *Proceedings of the 1st UK & Ireland Computing Education Research Conference (UKICER)*. ACM, New York, NY, USA, Article 8, 7 pages. <https://doi.org/10.1145/3351287.3351296>
- [21] Keith Nolan, Aidan Mooney, and Susan Bergin. 2015. Facilitating student learning in Computer Science: large class sizes and interventions. *International Conference on Engaging Pedagogy* (2015).
- [22] Keith Nolan, Aidan Mooney, and Susan Bergin. 2019. An Investigation of Gender Differences in Computer Science Using Physiological, Psychological and Behavioural Metrics. In *Proceedings of the Twenty-First Australasian Computing Education Conference (ACE '19)*. ACM, New York, NY, USA, 47–55. <https://doi.org/10.1145/3286960.3286966>
- [23] Keith Nolan, Aidan Mooney, and Susan Bergin. 2019. A Picture of Mental Health in First Year Computer Science. In *Proceedings of the 10th International Conference on Computer Science Education: Innovation and Technology (CSEIT'19)*. Global Science and Technology Fourm.
- [24] Jum C Nunnally. 1994. *Psychometric theory 3E*. Tata McGraw-hill education.
- [25] Richard Paul and Linda Elder. 2007. Critical thinking: The art of Socratic questioning. *Journal of developmental education* 31, 1 (2007), 36.
- [26] Keith Quille. 2019. *Predicting and Improving Performance on Introductory Programming Courses (CS1)*. Ph.D. Dissertation. National University of Ireland Maynooth.
- [27] K Quille and S Bergin. 2015. Programming: Factors that Influence Success Revisited and Expanded. In *International Conference on Enguaging Pedagogy (ICEP), 3rd and 4th December, College of Computing Technology, Dublin, Ireland*.
- [28] Keith Quille, Susan Bergin, and Keith Quille. 2019. CS1: how will they do? How can we help? A decade of research and practice research and practice. *Computer Science Education* 29, 2-3 (2019), 254–282. <https://doi.org/10.1080/08993408.2019.1612679>
- [29] Keith Quille, Keith Nolan, Sean McHugh, and Brett A. Becker. 2020. Associated Exam Papers and Module Descriptors. <http://tiny.cc/ITiCSE21OpenBook>.
- [30] Glenda C Rakes. 2008. Open book testing in online learning environments. *Journal of Interactive Online Learning* 7, 1 (2008), 1–9.
- [31] Beth Simon, Mike Clancy, Robert McCartney, Briana Morrison, Brad Richards, and Kate Sanders. 2010. Making Sense of Data Structures Exams. In *Proceedings of the Sixth International Workshop on Computing Education Research (ICER '10)*. Association for Computing Machinery, New York, NY, USA, 97–106. <https://doi.org/10.1145/1839594.1839612>
- [32] Christos Theophilides and Mary Koutselini. 2000. Study Behavior in the Closed-Book and the Open-Book Examination: A Comparative Analysis. *Educational Research and Evaluation* 6, 4 (2000), 379–393. <https://doi.org/10.1076/edre.6.4.379.6932> arXiv:<https://doi.org/10.1076/edre.6.4.379.6932>
- [33] T Rick Whiteley. 2014. Using The Socratic Method and Bloom's Taxonomy of the Cognitive Domain to Enhance Online Discussion, Critical Thinking, and Student Learning. In *Developments in Business Simulation and Experiential Learning: Proceedings of the Annual ABSEL conference*, Vol. 33.
- [34] Larry Yueli Zhang, Andrew K. Petersen, Michael Liut, Bogdan Simion, and Furkan Alaca. 2021. A Multi-Course Report on the Experience of Unplanned Online Exams. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE '21)*. Association for Computing Machinery, New York, NY, USA, 17–23. <https://doi.org/10.1145/3408877.3432515>