



---

# Evaluating the efficacy of different types of in-class exams

Wenya Cheng<sup>1</sup> and Geethanjali Selvaretnam<sup>1</sup>

## Abstract

In this research project, students were asked to compare their experiences of two types of exams in an undergraduate economics course with closed-book exams. The first type of exam had a timed group discussion session followed by individual work while the second type had a conventional open-book format. We find that a vast majority of students prefer our new assessment methods, but the group discussion session received mixed reviews. Post-exam feedback on exam preparation methods vindicates our hypothesis that closed-book exams may not encourage deeper learning and are not always an effective means of assessing students' knowledge and skills. Group discussions are valued by students to brainstorm ideas, clarify questions, and formulate arguments. However, allowing discussion just before a written exam is disruptive to students who want a serene atmosphere to gather and organise their thoughts.

## Keywords

open-book exam, closed-book exam, assessments, group discussion, higher education

---

<sup>1</sup> University of Glasgow, Glasgow, UK

### Corresponding Author:

Geethanjali Selvaretnam, Adam Smith Business School, University of Glasgow  
University Avenue, Glasgow, G12 8QQ, UK  
Email: [Geethanjali.Selvaretnam@glasgow.ac.uk](mailto:Geethanjali.Selvaretnam@glasgow.ac.uk)

## Introduction

In years past, it used to be common sight to see examination halls enveloped in pin-drop silence with invigilators ensuring the integrity of the assessment is not jeopardised by cheat notes, mobile phones, copying, or discussions. The question arises whether such an anxiety-stricken environment the most effective in assessing students' knowledge and skills. There is a rich literature in the past decade promoting the idea of assessment *for* learning and analysing the effects of such assessments (Boud, 2000, 2007, 2014; Brown, 2015, 2019; Brown et al., 2005; Brown & Race, 2012; Carless, 2015; Sambell et al., 2013; Wiliam, 2011; Wiliam & Thompson, 2008; Winstone & Boud, 2020).

The evidence comparing closed-book and open-book exams is mixed (Agarwal et al., 2008). Anaya et al. (2010) ran an experiment with engineering and business students and found that the former group learnt better with closed-book exams which required knowledge of specific quantitative methods while the latter had better learning with open-book exams which simulated the real world. Durning et al. (2016) conducted a systematic review of studies comparing the two types of exams. They found no significant evidence to suggest one is favoured over the other and there is room for both types to be used according to the learning outcomes of the course, considering the benefits that each type of assessment brings.

An argument in favour of closed-book exam is that it is more aligned with learning objectives which require quick retrieval of correct answers to act effectively in future career such as medicine, and science, technology, engineering and mathematics (STEM). Several studies have found that more preparation, deeper learning, quicker responses, and better performance can be expected in closed-book rather than open-book exams (Block, 2012; Durning et al., 2016). Heijne et al. (2008) carried out a study on medical students' learning approaches and found that closed-book exams resulted in more deep information processing than open-book exams. The exams involved information to be retained, processed, and appropriately applied quickly, which was an essential learning outcome being developed and assessed. Moore and Jensen (2007) also found that students in a biology course performed worse in a closed-book final exam when their in-course exam was open-book compared with students whose in-course exam were closed-book. Rummer et al. (2019) conducted a field experiment in two parallel introductory psychology courses where students were given preparatory exams for a small percentage of the total grade. These exams were followed by a surprise closed-book exam in week eight and another exam after a further six weeks. Both exams mostly required retrieval of information, and students who had closed-book exams as the preparation exams performed better.

However, existing literature also suggest some advantages in open-book exam over closed-book exam. Open-book exams may enable deeper level learning with less commitment to short-term memory, reduce negative effects of anxiety, enable wider level of assessment and can be more representative of certain professional settings (Anaya et.al., 2010; Green et al., 2016). Theophilides and Koutselini (2000) compared the study behaviour of students of education majors before and during exams. Their results suggested that closed-book exams resulted in students postponing preparation until closer to the exam, studying selected sections, and memorisation without necessarily understanding, while open-book exams resulted in students reading more sources to

expand their knowledge, being creative, delving deeper when answering questions and simulating the real world.

Previous studies have reported the benefits of open-book exams in STEM subjects. It has been found that short open-book exams throughout an introductory biology course cultivated the habit of reading more widely and engaging in deeper learning, evidenced by significantly improved performance, especially students who scored low at the beginning of the course (Phillips, 2006). Alumni of a postgraduate animal health course, who were mainly deep learners, regarded open-book exams to be less stressful and a more realistic method to examine their ability to apply their knowledge to solve problems (Dale et al., 2009). Ramamurthy et al. (2016) explored the learning approaches of students in a pharmacy course and found that open-book exams promote less memorisation, less anxiety, and more problem solving. A recent review of literature about the advantages and disadvantages of open-book exams in medical studies concluded that the benefits outweigh the downsides, and it is time to use open-book exams more so that learners do not spend too much time on memorising facts (Teodorszuk et al., 2018).

In some cases, a variety of assessments may be preferred. Based on a systematic literature review, Johanns et al. (2017) concluded that nursing students would benefit from a mix of open-book and closed-book exams during the programme to develop different sets of essential skills, although there was higher student satisfaction and less anxiety in open-book exams. Gharib and Phillips (2013) found that students of an introductory psychology course performed slightly better in open-book exams but were better prepared and less anxious in exams with crib sheets.

Open-book exams can come in different formats, such as an in-person exam with access to resources, take-home exam, and online exam. Compared to an in-person exam, take-home and online exams may be more susceptible to cheating as one cannot ensure the answers submitted by students are entirely their own work (D'Souza & Siegfeldt, 2017; King et al., 2009). In fact, the transition to online exams have also raised questions about the effectiveness of open-book exams (Clark, et al., 2020; Gamage et al., 2020). BBC (2018) reported that one in seven students had paid for essay mill services. Now steps are being taken to ban essay mills (Department for Education, 2021).

Research on exams which allow students to discuss questions with their peers before answering individually is rather limited, as this method is not used much in summative assessments. However, there are plenty of studies finding the positive learning benefits of group discussions, such as in two-stage exams and team-based learning (Bruno et al. 2017; Fatmi et al., 2013; Jang et al., 2017; Koles et al., 2010; Michaelsen et al., 2004). These studies suggest that group discussions encourage critical thinking and enhance understanding. van Blankenstein et al. (2011) found that engaging in discussion is more likely to result in long-term benefit than short-term benefit. Knierim et al. (2015) studied the learning outcomes of two-stage exams in an introductory geology course. Exam scores were higher when a closed-book exam was followed by a collaborative group submission (typical two stage) instead of an open-book exam, suggesting that peer collaboration and discussion had more learning benefits. Nicol and Selvaretnam (2021) analysed the internal feedback being self-generated during group discussion and showed a high level of self-regulatory feedback being generated. They described not only the exam design but also the suitability of the venue to conduct such exams, where students

are able to discuss without disturbing the others too much and comfortably complete the individual exams.

In this study, we focus on two types of in-person open-book exams to encourage deeper learning and reduce stress during preparation and assessment. The first exam allowed students to discuss with their peers in groups before writing their own answers while the second exam only allowed students to refer to any reading materials. The exams were conducted in an undergraduate economics course in 2019. Both exams took place at a set time in a set place, and therefore ruled out any possibility of plagiarism and ghost-writing. This study adds to the literature which explores different types of exam styles to enhance student learning.

## Research questions

Our research objective is to evaluate the efficacy of two types of in-class exams. In particular:

1. How do these types of exams affect academic performance?
2. What are the preferences of students regarding different types of exams?
3. What are the advantages and disadvantages of these two exams from students' perspectives?

## Methodology

### Description of exams

Students in an undergraduate honours optional course (FHEQ level 6) in the Business School, *Economics of Poverty*, were required to attempt two unseen in-class exams as summative assessments with structured questions in the middle and at the end of a 10-week teaching semester. Each in-class exam, which were the focus of our study, lasted for 90 minutes and accounted for the same weight. These exams were designed so that examiners were primarily facilitators of the exams – handing out papers, maintaining a comfortable atmosphere, etc.

Exam 1 in week 5 covered material from the first half of the course. The exam allowed students to discuss the exam questions in groups of three to four with any reading materials before writing their own answers. The maximum time for group discussion was 30 minutes. Students could decide how to utilise this discussion time. If students preferred to work individually throughout the exam, they would have 90 minutes to write their answers. If they spent 30 minutes on group discussion, they would have 60 minutes to complete the exam. While students had access to any reading material they brought during the group discussion, they could consult only one A4 size crib sheet when completing the individual work. The rationale for this design is to introduce group discussions which could promote different perspectives and critical thinking but not widely used in summative assessments (e.g., Fatmi et al., 2013; Koles et al., 2010).

Exam 2 in week 10 was a typical open-book exam which allowed students to take any reading materials into the exam hall and work individually for 90 minutes. The exam focused on material from the second half of the course.

### Data collection

To answer the first research question, exam grades are a natural choice for comparison. For the remaining two questions, the methodology hinges on eliciting information from

students about how a particular type of exam compares with other types of exams they have experienced, in terms of preparation and effective performance (Cohen et al., 2018; De Vaus et al., 2008). Learner experience data were collected through anonymous survey forms which were completed by students immediately after each exam. The survey questions are provided in Table 1. Exam 1 and Exam 2 were attended by 35 and 36 students respectively, who in turn returned 32 and 33 forms and gave informed consent to participate in our research. The completed forms and answer scripts were collected separately to ensure anonymity. This research has been approved by the College of Social Science ethics committee, University of Glasgow.

**Table 1:** Survey questions

<p><b>Questionnaire for Exam 1</b></p> <p>Rank your preference (1 to 4, 1 being the highest and 4 the lowest) on the following assessment formats:</p> <p>___ Closed-book individual exam (usual exam style)</p> <p>___ Open-book group discussion followed by individual work (Exam 1)</p> <p>___ Open-book individual work - one A4 size crib sheet allowed</p> <p>___ Open-book individual work - all paper material allowed (Exam 2)</p> <p>Comment on this assessment format (i.e., group discussion followed by individual work) – what are the advantages and disadvantages, the suitability of the assessment etc.</p> <p>How did you prepare for this assessment? Would you prepare differently if it's a closed-book exam?</p> <p>How did you make use of the group discussion time at the beginning of this assessment?</p>
<p><b>Questionnaire for Exam 2</b></p> <p>Re-rank the following assessment formats based on your preference (1 being the highest and 4 the lowest):</p> <p>___ Closed-book individual exam (usual exam style)</p> <p>___ Open-book group discussion followed by individual work (Exam 1))</p> <p>___ Open-book individual work - one A4 size crib sheet allowed</p> <p>___ Open-book individual work - all paper material allowed (Exam 2)</p> <p>Comment on this assessment format (open-book individual work with all paper material allowed) – what are the advantages and disadvantages, the suitability of the assessment etc.</p> <p>How did you prepare for this assessment? Would you prepare differently if it's a closed-book exam?</p> <p>Would you prepare differently if it were an open-book group discussion followed by individual work (like the previous in-course exam)?</p>

Question 1 of both surveys asks students to rank four exam styles: closed-book individual exam; open-book group discussion followed by individual work (Exam 1); open-book individual work with all paper materials allowed (Exam 2); open-book individual work (only one A4 size crib sheet allowed). This provides us with some quantitative data about their preferences. As students have experienced closed-book exams in other courses, this exam style was included in the surveys for comparison. The third exam type listed in the questionnaire combines the two in-class exams – it resembles Exam 1 in the sense that students have to work individually but can bring only one A4 crib sheet, and at the same time removes the group discussion component like Exam 2.

After ranking their preferences, students were asked to briefly explain the advantages and disadvantages of the exam format they had just experienced and their preparation method. We performed a thematic analysis where the qualitative data from the surveys were coded to find out the number of comments under various categories (Cohen et al., 2018; Ryan & Bernard, 2003). For example, when answering Question 2, a student might write ‘it is less stressful and relies on all knowledge instead of purely on memory’. This could fall into two categories – ‘less stressful’ and ‘test overall knowledge’. The number of comments under each category and the quotations throw some useful insights.

Observations of examiners are used to complement our analysis. Two examiners, who are also the authors of this paper, were present at the exam hall. They observed the behaviour of the students: how they formed groups, engaged in discussions and when they finished discussion in Exam 1; what type of material they brought into the exam hall; the suitability of the exam halls, etc. Notes were taken during their observation.

The analysis and findings on the students’ exam performance, preference of exam format, exam preparation, advantages and disadvantages of the exams, and examiners’ observations are presented in the following sections.

## Results

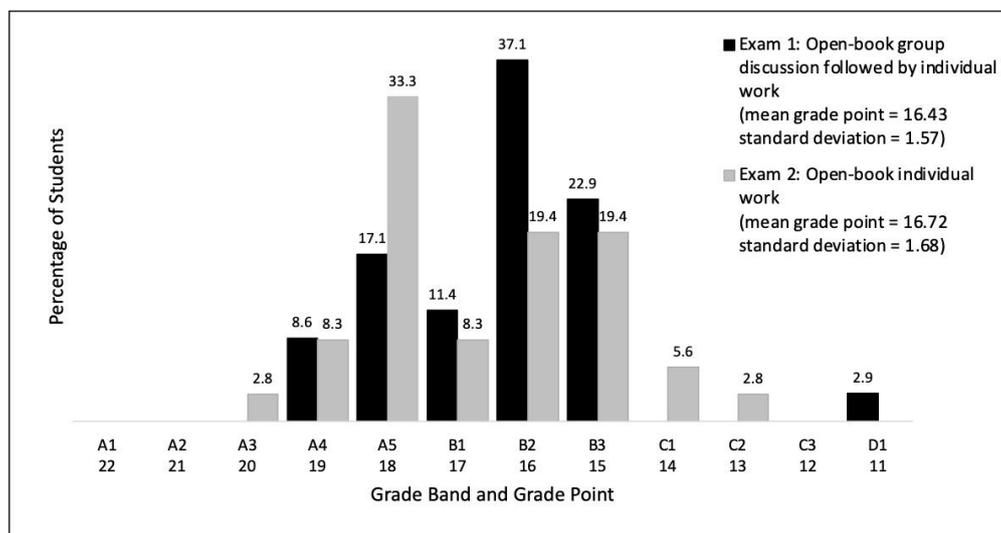
### Exam performance

The first research question asks how the two types of exams affect student performance. In order to minimise the potential bias caused by differences in exam syllabus, questions were set to achieve the same set of intended learning outcomes (see Table A1 in the appendix).

The grade distributions of the two exams are shown in Figure 1. Students were awarded grade points on a 22-point scale, where A1 is the highest at 22, and H is the lowest with 0. The grade descriptors are provided in Table A2 in the appendix. Students performed well in the two exams with no one receiving less than D1 (grade point 11).

The mean (standard deviation) of Exam 1 and Exam 2 are B2 at 16.43 (s.d. 1.57) and B1 at 16.72 (s.d. 1.68) respectively. The mode also increased from B2 at 16 to A5 at 18 between exams. Given the small class size, it is reasonable to find that the grades are not normally distributed. The grade distributions of both exams seem bimodal with distinct peaks at A5 and B2, yet more students received grade A in Exam 2 than in Exam 1. Bimodality indicates that the class had two distinct groups of students, a high-performing group, and a low-performing group. One may notice an outlier in Exam 1 (D1) which could have a disproportionate effect on statistical results. Excluding this outlier increases the mean of Exam 1 to 16.5 and reduces the standard deviation to 1.29.

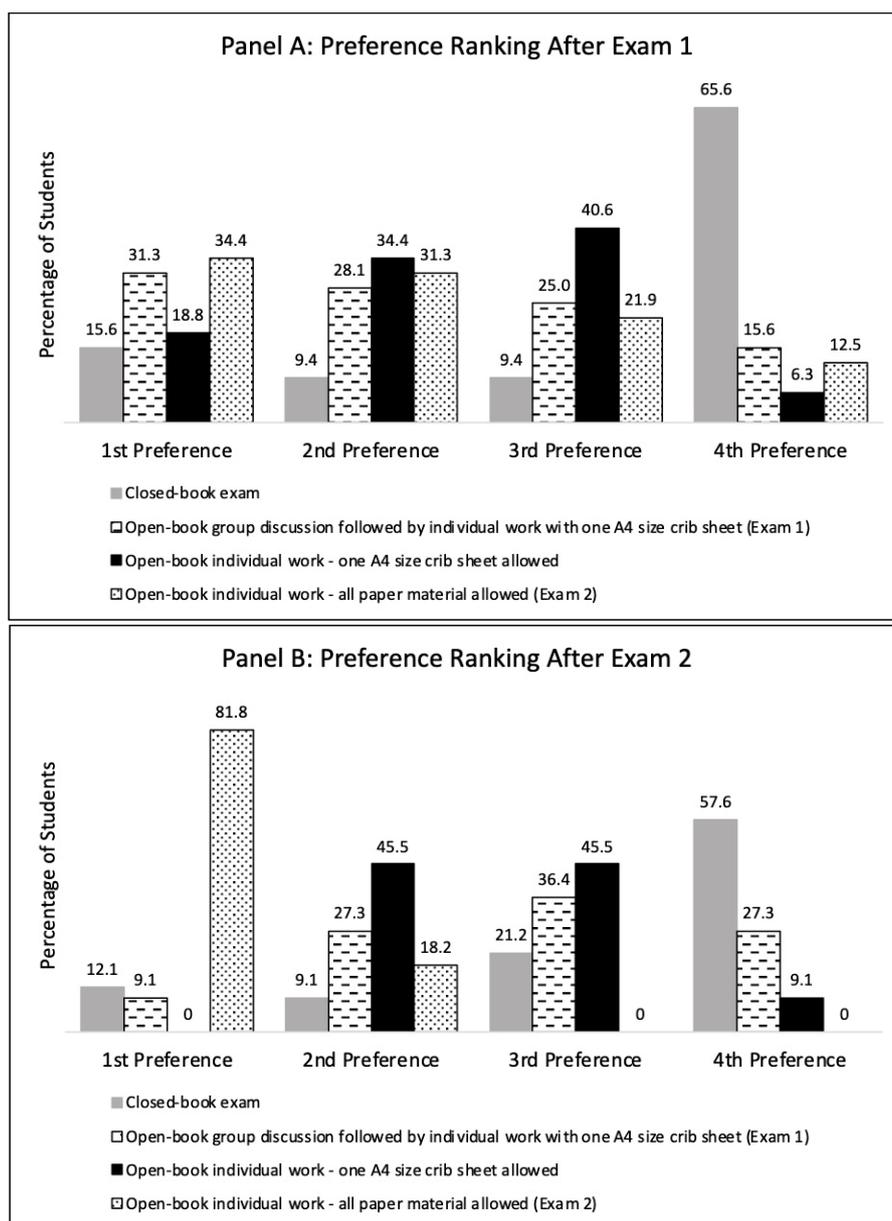
While the mean and mode of grades improved in Exam 2, we would like to know if there was also an improvement in the overall grade distribution. Considering the grades are not normally distributed, we conducted the nonparametric Mann-Whitney U test to test for equivalence of the two exam results. The p-value is 0.396 which suggests that the difference in exam performance is not statistically significant. This conclusion does not change after excluding the outlier in Exam 1. A potential explanation for this result is the grades of the low-performing students in Exam 1 became more dispersed in Exam 2: some achieved better grades while some scored worse than before.



**Figure 1:** Grade distributions of Exam 1 and 2

### Ranking of preference

The second research question is about students' preference of exam formats. Understanding how their preferences change with exam experience is also important. Results of survey question 1 are presented in Figure 2. The bars representing the four exam formats are grouped by students' preference ranking. Panel A shows students' preference after Exam 1 which had a group discussion component while Panel B shows the rankings after Exam 2 which was the usual open-book exam with individual work and all paper material allowed. In both surveys, Exam 2 was the most preferred and closed-book exam was the least favoured. Some students still preferred closed-book exam, perhaps because they have a comparative advantage in this exam format.



**Figure 2:** Preference ranking of exam formats

Table 2 presents the median and inter-quantile range of the preference rankings for the four exam formats. A lower median indicates stronger preference. The results reassure our previous finding that open-book exam was more preferred to closed-book exam, and Exam 2 was the most popular format. An interesting finding is after experiencing Exam 2, all students chose this exam format as one of their top two preferences. The median preference of Exam 2 changed from 2 to 1, indicating an increase in preference.

To investigate if there is a significant change in student preferences, we computed the p-value of the Mann-Whitney U test for each exam format. The last column of Table 2 shows that the shift in preference towards Exam 2 is statistically significant ( $p=0.0001$ ) and is mainly at the expense of Exam 1. The median preference of Exam 1 changed from 2 to 3, indicating a decrease in preference and is statistically significant ( $p=0.0307$ ). Students' preference for closed-book exam was consistent across surveys while the

change in preference for open-book individual exam with one A4 crib sheet allowed is not statistically significant ( $p=0.2016$ ).

**Table 2:** Preference strength of different exam formats

	Survey 1 median (inter-quantile range)	Survey 2 median (inter-quantile range)	p-value of Mann-Whitney U test
Closed-book exam	4 (2.5-4)	4 (3-4)	0.7236
Open-book group discussion followed by individual work with one A4 size crib sheet (Exam 1)	2 (1-3)	3 (2-4)	0.0307
Open-book individual work - one A4 size crib sheet allowed	2 (2-3)	3 (2-3)	0.2016
Open-book individual work - all paper material allowed (Exam 2)	2 (1-3)	1 (1-1)	0.0001

### Advantages and disadvantages of the two exam formats

The final research question is about the advantages and disadvantages of the format of Exam 1 and Exam 2. To analyse this, we use the responses to the open-ended questions in both surveys as well as examiners' observations.

#### Advantages and disadvantages of Exam 1

We coded the responses to question 2 in Survey 1 by dividing students' comments to short phrases. The number of comments per category is presented in Table 3.

**Table 3.** Advantages and Disadvantages of Exam 1

Exam 1: Open-book group discussion followed by individual work with one A4 size crib sheet	No. of comments
<b>Advantages</b>	
Sharing ideas/additional points/clarify/being reassured	17
Teamwork	3
Enjoyable exam	2
Reduce exam pressure	2
Allows notes/no need to memorise	2
Be better prepared to discuss	2
Realistic	2
<b>Disadvantages</b>	
Group discussion was chaotic	16

Discussion time would be better spent for individual preparation	7
Time for discussion was too short	4
Free-rider problem	4
May not know others with whom they could discuss	2
Layout of the exam hall	1

Regarding Exam 1, the most popular positive comment (from 17 students) was related to the benefits of group discussion such as being reassured and getting some additional ideas. Six students among these explicitly mention that allowing discussion among the peers before answering the exam was a good format. Another two students even said that they enjoyed the exam. Two students mentioned that this exam format was more akin to what they would have to face in reality and develops teamwork skills.

Some quotes which capture these sentiments are given below:

The group discussion helped me remember what the representation of the symbols were, and also allowed me a chance to clarify my answers and questions of a classmate.

It is good to be able to discuss how each of you interpret the questions, as well as adding more potential arguments to your answers if needed.

I enjoyed this style of exam as I feel it is more realistic to real life. It is unlikely in real life that we would have to answer a question/perform a task without discussing it with people or use notes.

Exams were considered less stressful and required less memorisation because students were allowed to take an A4 crib sheet to the exam hall:

Good for in-course assessment, takes off some of the extra exam pressure.

Contrary to a concern that students prepare less for open-book exams, two students commented that in Exam 1, they had to be well prepared to explain their thoughts to peers, as commented by a student:

It also adds more pressure to be more prepared (a good thing) as you don't want to disappoint your group.

On the downside, the most negative comments (from 16 students) were that the group discussion caused stress, chaos, disruption, and confusion before answering the exam. Another seven students thought the group discussion time could have been spent better working on their own. The following quotations from students capture these sentiments.

very stressful.

Very chaotic, time consuming.

I felt like I got confused at times by what other people were saying.

However, one must be mindful of the 17 positive comments about the benefits of group discussion. It is important to understand the causes of negative atmosphere.

Four comments were related to free-rider problem. Related to this issue, four students reflected on how they could have performed better by being more prepared and thus the time given for discussion was too short.

free-rider problem where some individuals copy points made by others but don't suggest own points.

I didn't learn the material as thoroughly as I would've for a regular exam as I was relying on the first 30 minutes of discussion.

Not enough time, discourages revision of material for some.

Two students raised the question about those who did not feel comfortable being part of any of the groups.

The disadvantage is if you know no one in your class to discuss with or have some form of social anxiety. And also if your group is as clueless as you.

We recognise the importance of providing information about the assessment format at the beginning of the course so that the students are aware of the use of group discussions. The course was also conducted with many opportunities during the classes so that students get used to discussing with peers.

One student raised an issue about the exam hall, which was an important administrative issue and should be considered by the examiners:

I didn't enjoy working from the chairs with the table arms (unsure what to call them), as they are all right-handed and a little unstable.

### Advantages and disadvantages of Exam 2

The coded responses to question 2 in Survey 2 are presented in Table 4. Exam 2 received more positive feedback from the students, which is consistent to their preference ranking. Moreover, the content of positive comments was more diverse with explanations. One comment summed up the general opinion:

love it. I have no comment on the exam!!

**Table 4:** Advantages and Disadvantages of Exam 2

<b>Exam 2: Open-book individual work - all paper material allowed</b>	<b>No. of comments</b>
<b>Advantages</b>	
Tests overall knowledge and understanding/applications	10
Better preparation without just memorising	9
Less stressful than closed-book exam and Exam 1	8
Able to write a good answer, better organised and detailed	6
Better than Exam 1	5
Good to have notes	4
Read more widely	4

<b>Disadvantages</b>	
Exam could be too challenging and stricter marking!	4
Students could be too reliant on notes	3
Cannot discuss ideas with peers	1
Paper wastage	1

Several students remarked the ability of Exam 2 to test overall knowledge and application so that students could invest in deeper understanding without rote memorisation. There were six comments related to better answer quality in terms of structure and details. A further four comments argued that such an exam encouraged students to engage in wider reading. The following quotes are examples of such views:

I like this style of exam as rather than testing what I can remember, I feel I am being tested on my understanding.

Allows for more applicable use of information. It is rare in any other situation other than an exam, that you have to write a lot of information that you have revised.

This is a nice style of assignment since you should learn the big picture well before the exam but you can still check the details during the exam to help you elaborate more and build a better answer. It is well suitable for this course since you can discuss in your answers and there's no clear-cut right answer. It's also good for practicing clear and concise answers were you to get your thoughts well collected.

I believed I learned more. I was able to read more widely, as I took notes on the things I would have had to memorise otherwise and could therefore learn more in depth on different topics.

Eight comments specifically mentioned this exam was less stressful, which could be linked to the four comments on benefits to make notes.

Having access to individual notes is certainly useful and helps to pinpoint useful details. In practice, it does not change things substantially if one has done the work in advance.

This was less stressful than last time - I have time to think and prepare careful outlines with the help of the note I had prepared.

It is noteworthy that five students were compelled to explicitly stated that Exam 2 was better than Exam 1 with comments such as:

Much better than the previous! ; that was awful - I would go straight to the seats in the middle or wouldn't talk to anyone.

On the downside, four students questioned whether the exam itself would be more challenging and the marking stricter:

Harsher marked? - Harder to get good grade!

There were three comments about students becoming too reliant on notes without imprinting important information in their minds:

it may encourage people to rely heavily on notes rather than learn the material.

One student mentioned the reverse sentiment that group discussion was beneficial:

Cannot bounce ideas off others; no reaffirmation that I was going down the right path with what I was thinking or writing.

One student was concerned that this exam promoted wasting paper. This issue should be considered by the examiners if we are to promote environmentally friendly practices.

### Allowing group discussion

The final question in Survey 1 was about how students made use of the group discussion at the beginning of Exam 1. Unsurprisingly, almost all students mentioned aspects such as clarified questions, improved own note with examples and details, some examples of which are:

We discussed how to interpret questions, then we discussed a variety of different arguments for the answers, allowing us to have better rounded responses to the questions and deeper analyse our own arguments.

Talked through the questions and brainstormed ideas.

As an English as a second language student, discussion help me to avoid the possible error of misunderstanding the question's purpose.

### Exam preparation

After each exam, the students were asked how they prepared for that exam and whether they would have prepared any differently if it were a closed-book exam. The responses in Surveys 1 and 2 are summarised in Tables 5.

**Table 5:** Exam preparation compared to closed-book exams

	No. of comments	
	Exam 1	Exam 2
Same, but would have spent more time if it was closed-book exam.	14	14
Took notes/post-it and organised them to take to the exam	14	16
Spent less time on just memorising for this exam	12	13
Read more widely	4	7
Study and prepare in groups	1	1
Less stressed before exam	0	2

Compared to closed-book exam, preparations for Exam 1 and Exam 2 were mostly the same, other than more organised notes and wider reading for the latter. One student pointed out the incentive problem in Exam 2:

I felt I had less need to prepare although I knew I should (hence the incentive problem).

Survey 2 also asked whether the students had prepared differently to Exam 1. It is noteworthy that 22 out of 36 students said they prepared the same way, and 10 students did not say anything. Two students mention more focused notetaking while another said:

I would have to prepare individual sheets with notes for every type of question.

Another student summed up the anxiety of closed-book exams by saying, *pray for myself!* There were a couple of comments about how they might have approached Exam 1 differently, which was not the question being asked.

### Examiners' observations

We observed and analysed students' behaviour during the two exams, such as how they utilised their group discussion time and what kind of reading materials they brought to the exams. During the exams we acted as facilitators rather than invigilators preventing malpractice which created a better assessment atmosphere.

The hall for Exam 1 was spacious and had separate areas for group discussions and individual work. The group discussion areas had several large tables with chairs round it, while the individual work area had a standard exam room set-up. Students who preferred to work alone could go straight to the individual work area and attempt the exam without too much disruption. However, most students did not seem to realise they were allowed to move to the individual work area before the maximum group discussion i.e., initial 30 minutes of the exam. A couple of students explained that they would have felt uneasy to do so because others might think they were being selfish. As most students rushed to the individual work area right after 30 minutes, the exam hall was rather chaotic. Although the exam format and exam hall were designed well, the exam process could be improved. If there were clearer instructions about what the students were expected or allowed to do during the exam, it might have reduced their perception of 'chaos'.

There was a concern about Exam 2 that students may not prepare sufficiently with a false sense of security because they could take any reading material into the exam hall. It was encouraging to see that was not the case. Many students prepared either handwritten or typed notes and had sticky notes indicating the different sections in their books or lecture notes. Except for a few who were wading through the reading materials, most students were comfortable to start tackling the exam questions. Several students had books which were not the recommended textbook, indicating wider reading and references.

### Discussion and conclusion

This study investigates students' experiences in terms of preparing and performing in different exam formats. Two types of exams were introduced in the course. Exam 1 had group discussion followed by individual work with an A4 crib sheet while Exam 2 was a typical open-book exam that allowed all paper material. Since both exams were open-book, questions were designed with emphasis on information synthesis and problem-solving rather than memorisation, which were more appropriate to assess student achievement of the intended learning outcomes. Both exams were in-person instead of take-home or online to avoid inappropriate collaboration and contract cheating which increased the credibility of exam results. It is evident that students preferred the exam formats in this course to closed-book exams and Exam 2 was overwhelmingly preferred

over Exam 1, especially after experiencing both exam types. The analysis of the grades, preference ranking and written comments in the post-exam surveys point to some useful findings. We reflect on the messages that can impact academic practice.

The first research question is about student performance in the two exams. To the best of our knowledge, comparison of such exam types is absent in the literature. In our study, we find that the mean and mode of grades were higher in Exam 2, but the grade distribution was more dispersed. The improvement of average grades can be explained by two reasons. First, students could consult more paper material when writing answers in Exam 2 than Exam 1. Second, students were likely to benefit from the experience and feedback in Exam 1 regarding exam preparation and answering. Indeed, the examiners' observation and survey feedback confirmed that many students prepared properly for Exam 2 such as creating self-made notes, highlighting, and annotating. The larger grade dispersion of Exam 2 could be because students could not share ideas and therefore had more diversity in answer quality. This result is consistent with prior studies in two-stage exams which show that group stage helps to reduce performance gap between high-performing and low-performing students (Bruno et al., 2007; Knierim et al., 2015; Koles et al., 2010; Phillips, 2006). Using the Mann-Whitney U test, we did not find a statistically significant difference between the grades of the two exams. Therefore, it is hard to conclude which exam type is more effective in enhancing overall class performance. Another interesting aspect was the grade distribution in both exams were bimodal, indicating that changing the exam format and having more exam experience are insufficient to close the performance gap between the distinct groups of high- and low-performing students.

The second research question is related to student preference of different assessment types. In addition to the two exams in this course, we asked students to consider closed-book exams which they were familiar with and an open-book exam with only an A4 crib sheet. In line with the findings Gharib and Phillips (2013) and Johanns et al. (2017), open-book exam was more preferred to closed-book exam in both surveys, and Exam 2 was the most popular format. An interesting finding is that after experiencing Exam 2, all students chose this exam format as one of their top two preferences and the popularity of Exam 1 significantly dropped. A possible explanation for the statistically significant changes in preferences is that students were unfamiliar with these two exam formats, and they could make a more informed comparison after experiencing both exams.

The third research question examines the advantages and disadvantages of these exams. Student responses to the survey questionnaire in our study, especially for Exam 2, confirmed the findings from previous studies that open-book exams reduce anxiety, enable assessment of relevant knowledge and its application rather than committing information to short-term memory (e.g., Gharib & Phillips, 2013; Green et al., 2016; Johanns et al., 2017; Moore & Jensen, 2007; Teodorczuk et al., 2018). Existing studies show evidence of deep learning when students prepare for open-book exams by reading additional resources, critically evaluating these sources of information, and being creative, while closed-book exams can result in students postponing studies, memorising without proper understanding, and sticking to assigned readings only (Dale et al., 2009; Green et al., 2016; Theophilides & Koutselini, 2000). In fact, several students in our study reported reading beyond the assigned readings to develop their thinking and ideas. Therefore, preparation for open-book exams seems more suitable for deeper learning than what a closed-book exam would achieve in this course.

Free riding was a potential problem in Exam 1 as it did not reward the group discussion, so students might feel that others could take advantage of their ideas when they write their individual answers without sharing their own ideas. Although only four out of the 36 students raised this as an issue, it is worth bearing in mind as a potential challenge in this exam design.

A few students had raised the issue whether both Exams 1 and 2 would lead to over-reliance on notes and impede learning of material. It is important that these types of exams are chosen only if they are suitable to achieve and assess the intended learning outcomes of the course. Closed-book exams should be used in certain subjects which require accurate application of technical knowledge in a time sensitive manner (Block, 2012; Heijne, 2008; Rummer et al., 2019). Exams which allow reading material and/or group discussions would be appropriate when the intended learning outcomes emphasise critical thinking and wider reading, without the negative effects of anxiety (Green et al., 2016; Koles et al, 2010; Theophilides & Koutselini, 2000). As this course focused on critical analysis, evaluation, and synthesise of knowledge rather than recall of information, our exam formats were appropriate to the skills examined.

We expected disruptions in Exam 2 with students wading through too much reading material, not having enough room on the desk, dropping things etc. (Block, 2012; Theophilides & Koutselini, 2000). However, these issues did not arise because students had prepared beforehand, and had planned and organised the material which they brought into the exam hall. This adds to Carrier's (2003) finding that students who prepare appropriately for open-book exams, knowing where to find the necessary information, will be able to perform well.

Despite the lower preference, many students appreciated the benefits of Exam 1, where group discussions before individual work allowed them to brainstorm and be inspired. However, rather than reducing anxiety, a number of students felt more anxious and unsettled because of the disruption to their train of thought when they moved from group discussion to individual work. It is vital to maintain a serene atmosphere in an exam hall if we want to assess the capabilities and understanding of the students comprehensively. This form of assessment and issue are not prevalent in academic discussions, but it is clear from studies about two-stage exams that a relaxing environment for discussions without disturbance inspire immediate feedback and learning (Jang et al., 2017; Nicol & Selvaretnam, 2021).

This study threw some light into inclusive assessments with diverse students. An advantage of group discussion is that students could clarify what exactly is expected from the questions, which would especially benefit non-native students. If Exam 1's format is adopted, clearer instructions should be provided in terms of expectations and logistics. It is also important to have an appropriate venue for the exam, with enough space for designated areas for discussions and exam writing. Some students raised the issue that chairs fixed with writing board did not provide enough space. Not only should we ensure reasonable writing space, but also bear in mind that left-handed students are not disadvantaged. Another issue is about students who may not fit in easily into a group for the discussion. We had conducted the whole course with students discussing in small groups during lectures and the students being alerted to this form of exam. Surprisingly, some students had not made any pre-arrangements and just sat in a random group for

discussion. When designing such assessments, teachers should be alert to students who might be left out of group discussions.

This study contributes to the existing literature on assessment in several ways. First, we introduced a new assessment format which uses discussion as a learning tool to promote better understanding (two-stage exams have a different design). Exam 1 was designed considering the pedagogical benefits of discussion and open-book exams. Second, we analysed learners' responses regarding their preference of the different types of exams and reasons related to preparation, learning and exam experience. Third, we discussed the advantages and disadvantages of two types of exams from learners' perspectives.

We acknowledge there are some limitations in this study. An ideal experiment would consist of students attempting the same exam questions in different assessment formats. However, this was not possible for our study given the time constraint and class size. We also realise that the findings of this study may not be generalised with one inquiry comprising of a class of less than forty students in a particular economics course. Also, using exam grades to evaluate student learning outcomes may have its limitations. Further research in different disciplines, class size and types of courses would enable wider suggestions for academic practice. A drawback of this analysis is that we were unable to compare the performance of these students in a closed-book exam in this course. Such an analysis will be a useful addition to this strand of literature.

## References

- Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger, H. L. III, & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied of Cognitive Psychology, 22*, 861–876. <https://doi.org/10.1002/acp.1391>
- Anaya, L., Evangelopoulos, N., & Lawani, U. (2010). Open-book vs. closed-book testing: an Experimental Comparison. *American Society for Engineering Education*, 2010. <https://strategy.asee.org/open-book-vs-closed-book-testing-an-experimental-comparison.pdf>
- BBC News. (2018, August 31). *Essay mills: 'One in seven' paying for university essays*. <https://www.bbc.co.uk/news/uk-wales-45358185>
- Block, R.M. (2012). A discussion of the effect of open-book and closed-book exams on student achievement in an introductory statistics course. *PRIMUS, 22*(3), 228 – 238. <https://doi.org/10.1080/10511970.2011.565402>
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151-167. <https://doi.org/10.1080/713695728>
- Boud, D. (2007). Reframing assessment as if learning were important. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education* (pp.181-197). Routledge. <https://doi.org/10.4324/9780203964309>
- Boud, D. (2014). Shifting views of assessment: From secret teachers' business to sustaining learning. In C. Kreber, C. Anderson, N. Entwistle & J. McArthur (Eds.) *Advances and innovations in university assessment and feedback* (pp. 13-31). Edinburgh University Press. <https://doi.org/10.3366/edinburgh/9780748694549.003.0002>
- Brown, S. (2015). *Learning, Teaching and Assessment: Global perspectives*. Palgrave.
- Brown, S. (2019). Using assessment and feedback to empower students and enhance their learning. in *Innovative assessment in higher education: A handbook for academic practitioners*. Routledge. <https://doi.org/10.4324/9780429506857-5>

- Brown, S., & Race, P. (2012). Using effective assessment to promote learning. In D. Chalmers & L. Hunt (Eds.), *University teaching in focus: A learning-centred approach*. Australian Council for Educational Research. <https://doi.org/10.4324/9780203079690-5>
- Brown, S., Race, P., & Smith, B. (2005). *500 tips on assessment* (2nd edition). Routledge. <https://doi.org/10.4324/9780203307359>
- Bruno, B.C., Engels, J., Ito, G., Gillis-Davis, J., Dulai, H., Carter, G., Fletcher, C., & Böttjer-Wilson, D. (2017). Two-stage exams: A powerful tool for reducing the achievement gap in undergraduate oceanology and geology classes. *Oceanography*, 30(2), 198 – 208. <https://doi.org/10.5670/oceanog.2017.241>
- Carless, D. (2015). *Excellence in university assessment: learning from award-winning practice*. Routledge. <https://doi.org/10.4324/9781315740621>
- Carrier, L. M. (2003). College students' choices of study strategies. *Perceptual and Motor Skills*, 96, 54–56. <https://doi.org/10.2466/pms.2003.96.1.54>
- Clark, T., Callam, C., Paul, N., Stoltzfus, M., & Turner, D. (2020). Testing in the time of COVID-19: A sudden transition to unproctored online exams. *Journal of Chemical Education*, 97(9), 3413–3417. <https://doi.org/10.1021/acs.jchemed.0c00546>
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8<sup>th</sup> edition). London: Routledge. <https://doi.org/10.4324/9781315456539>
- Dale, V. H. M., Wieland, B., Pirkelbauer, B., & Nevel, A. (2009). Value and benefits of open-book examinations as assessment for deep learning in a post-graduate animal health course. *Journal of Veterinary Medical Education*, 36, 403–410. <https://doi.org/10.3138/jvme.36.4.403>
- De Vaus, D., Petri, P. A., Kidd, S., & Shaw, D. (2008). *Constructing questionnaires. Surveys in Social Research* (pp. 93 – 120). Taylor & Francis. <https://doi.org/10.4324/9780203519196>
- Department for Education. (2021, October 5). *Essay mills to be banned under plans to reform post-16 education* [Press release]. <https://www.gov.uk/government/news/essay-mills-to-be-banned-under-plans-to-reform-post-16-education>
- D'Souza, K.A., & Sigfeldt, D.V. (2017). A conceptual framework for detecting cheating in online and take-home exams. *Decision Sciences Journal of Innovative Education*, 15(2), 370–391. <https://doi.org/10.1111/dsji.12140>
- Durning, S., Dong, T., Ratcliffe, T., Schuwirth, L., Artino, A., Boulet, J., & Eva, K. (2016). Comparing open-book and closed-book examinations: A systematic review. *Academic medicine. Journal of the Association of American Medical Colleges*, 91(4), 583–99. <https://doi.org/10.5281/zenodo.4256825>
- Fatmi, M., Hartling, L., Hillier, T., Campbell, S., & Oswald, A. E. (2013). The effectiveness of team-based learning on learning outcomes in health professions education: BEME guide no. 30. *Medical Teacher*, 35(12), e1608–e1624. <https://doi.org/10.3109/0142159X.2013.849802>
- Gamage, K.A.A., Silva, E.K.d., & Gunawardhana, N. (2020). Online delivery and assessment during COVID-19: Safeguarding academic integrity. *Education Science*, 10(11), 301. <https://doi.org/10.3390/educsci10110301>
- Gharib, A., & Philips, W. (2013). Test anxiety and performance on open-book and cheat sheet exams in introductory psychology. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 9(1). <http://www.ipedr.com/vol53/001-BCPS2012-C00006.pdf>
- Green, S. G., Ferrante, C. J., & Heppard, K. A. (2016). Using open-book exams to enhance student learning, performance, and motivation. *The Journal of Effective Teaching*, 16(1), 19–35. <https://files.eric.ed.gov/fulltext/EJ1092705.pdf>

Heijne, M., Kuks, J., Hofman, W., & Cohen-Schotanus, J. (2008). Influence of open- and closed-book tests on medical students' learning approaches. *Medical Education*, 4, 967-74.

<https://doi.org/10.1111/j.1365-2923.2008.03125.x>

Jang, H., Paulson, J. A., Lasri, N., Miller, K., Mazur, E., & Paulson, J. A. (2017). Collaborative exams: Cheating? Or learning? *American Journal of Physics*, 85, 223-227. <https://doi.org/10.1119/1.4974744>

Johanns, B., Dinkens, A., & Moore, J. (2017). A systematic review comparing open-book and closed-book examinations: Evaluating effects on development of critical thinking skills. *Nurse Education in Practice*, 27(2017), 89-94. <https://doi:10.1016/j.nepr.2017.08.018>

King, C. G., Guyette, R. W. Jr., & Piotrowski, C. (2009) Online exams and cheating: An empirical analysis of business students' views. *The Journal of Educators Online*, 6(1), 1-11.

<https://doi.org/10.9743/IEO.2009.1.5>

Knierim, K., Turner, H., & Davis, R. K. (2015). Two-stage exams improve student learning in an introductory geology course: logistics, attendance, and grades. *Journal of Geoscience Education*, 63(2), 157-164. <https://doi.org/10.5408/14-051.1>

Koles, P. G., Stolfi, A., Borges, N. J., Nelson, S., & Parmelee, D. X. (2010). The impact of team-based learning on medical students' academic performance. *Academic Medicine*, 85(11), 1739-1745.

<https://doi.org/10.1097/ACM.0b013e3181f52bed>

Michaelsen, L. K., Knight, A. B., & Fink, L. D. (2004). Team-based learning: A transformative use of small groups in college teaching. *Centers for Teaching and Technology - Book Library*, 199.

<https://digitalcommons.georgiasouthern.edu/ct2-library/199>

Moore, R., & Jensen, P. A. (2007). Do open-book exams impede long-term learning in introductory biology courses? *Journal of College Science Teaching*, 36(7), 46-49.

<https://www.questia.com/library/journal/1G1-169164816/do-open-book-exams-impede-long-term-learning-in-introductory>

Nicol, D., & Selvaretnam, G. (2021). Making internal feedback explicit: harnessing the comparisons students make during two-stage exams. *Assessment & Evaluation in Higher Education*, 47, 1-16.

<https://doi.org/10.1080/02602938.2021.1934653>

Phillips, G. (2006). Using open-book tests to strengthen the study skills of community-college biology students. *Journal of Adolescent & Adult Literacy*, 49(7), 574-582.

<https://doi.org/10.1598/JAAL.49.7.3>

Ramamurthy, S., Er, H., Nadarajah, V., & Pook, P. (2016). Study on the impact of open and closed book formative examinations on pharmacy students' performance, perception, and learning approach. *Currents in Pharmacy Teaching and Learning*, 8, 364-374.

<https://doi.org/10.1016/j.cptl.2016.02.017>

Rummer, R., Schweppe, J., & Schwede, A. (2019) Open-book versus closed-book tests in university classes: a field experiment. *Frontiers in Psychology*, 10, 463.

<https://doi.org/10.3389/fpsyg.2019.00463>

Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field Methods*, 15(1), 85-109.

<https://doi.org/10.1177/1525822X02239569>

Sambell, K., McDowell, L., & C. Montgomery. (2013). *Assessment for learning in higher education*.

Routledge. <https://doi.org/10.4324/9780203818268>

Teodorczuk, A., Fraser, J., & Rogers, G. D. (2018). Open-book exams: A potential solution to the "full curriculum"? *Medical Teacher*, 40(5), 529-530. <https://doi.org/10.1080/0142159X.2017.1412412>

Theophilides, C., & Koutselini, M. (2000). Study behavior in the closed-book and the open-book examination: a comparative analysis. *Educational Research Evaluation*, 8(4), 379–393.

<https://doi.org/10.1076/edre.6.4.379.6932>

van Blankenstein, F.M., Dolmans, D.H.J.M., & van der Vleuten, C.P.M. (2011). Which cognitive processes support learning during small-group discussion? The role of providing explanations and listening to others. *Instructional Science*, 39, 189–204. <https://doi.org/10.1007/s11251-009-9124-7>

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14.

<https://doi.org/10.1016/j.stueduc.2011.03.001>

Wiliam, D., & M. Thompson, (2008). Integrating assessment with learning: What will it take to make it work? In C. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53-84).

Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781315086545-3>

Winstone, N. E., & Boud, D. (2020): The need to disentangle assessment and feedback in higher education. *Studies in Higher Education*, 47(3), 656–667.

<https://doi.org/10.1080/03075079.2020.1779687>

## Appendix

**Table A1:** Course Intended Learning Outcomes

<ul style="list-style-type: none"> <li>Clearly explain and critically analyse concepts used in the economics of poverty, discrimination and development</li> <li>Critically examine the context in which the poor undertake decisions and analyse how economic development is affected by human behaviour</li> <li>Critically evaluate policies which alleviate poverty and discrimination as well as economic development from economic and social perspectives</li> </ul>
---

**Table A2:** Grade descriptors at the University of Glasgow

Grade band	Grade point	Grade Descriptor	Grade band	Grade point	Grade Descriptor
A1	22	Excellent	D1	11	Satisfactory
A2	21		D2	10	
A3	20		D3	9	
A4	19		E1	8	Weak
A5	18	E2	7		
B1	17	Very Good	E3	6	Poor
B2	16		F1	5	
B3	15		F2	4	
C1	14	Good	F3	3	Very Poor
C2	13		G1	2	
C3	12		G2	1	
			H	0	