

ISSN: 1755-6031

Journal homepage: <https://www.cambridgeassessment.org.uk/our-research/all-published-resources/research-matters/>

Comparing open-book and closed-book exams in international literature in English assessments

Emma Walland

To cite this article: Walland, E. (2025). Comparing open-book and closed-book exams in international literature in English assessments. *Research Matters: A Cambridge University Press & Assessment publication*, 40, 28–49. <https://doi.org/10.17863/CAM.121935>

To link this article: <https://www.cambridgeassessment.org.uk/Images/research-matters-40-comparing-open-book-and-closed-book-exams-in-international-literature-in-english-assessments.pdf>

Abstract:

When designing assessments such as examinations, it is important to consider which materials students should be permitted to access during the assessment, for example, whether to allow students access to a relevant book, such as a literary text. Debates among teachers and assessment professionals highlight differing views about the advantages and disadvantages of open-book and closed-book exams, and how each approach could influence the constructs being assessed, student performance, exam preparation and students' experiences. However, the effects of open- and closed-book exams vary based on how they are implemented and enacted. This small-scale study explored differences between open-book and closed-book exams in international secondary school literature in English qualifications. Variations in the nature, extent, and quality of textual references in samples of essays written under different conditions were explored. Additionally, the research explored a small group of assessors' perspectives on these differences and analysed the differences in student marks in some regions based on examination format. This study provides valuable insights into the ongoing debate about open- and closed-book exams within this context, informing syllabus development and future research. The findings highlight the complexities and interactions among various assessment principles.

Cambridge University Press & Assessment is committed to making its documents accessible in accordance with the WCAG 2.1 Standard. We're always looking to improve the accessibility of our documents. If you find any problems or you think we're not meeting accessibility requirements, contact our team:

Research Division, ResearchDivision@cambridge.org

If you need this document in a different format contact us, telling us your name, email address and requirements and we will respond within 15 working days.

© Cambridge University Press & Assessment 2025

Full Terms & Conditions of access and use can be found at

[T&C: Terms and Conditions | Cambridge University Press & Assessment](#)

Comparing open-book and closed-book exams in international literature in English assessments

Emma Walland (Research Division)

Introduction and background

High-stakes exams typically operate with restrictions on the materials students may access. A debate in literature in English¹ relates to whether students should be allowed to take their set texts (free from added notes and annotations) into their exams, and the benefits and drawbacks of this. Questions surround the possible impact of an open-book policy on the constructs being assessed, as well as the potential influence on teaching, learning, exam preparation and students' experiences.

This research focused on an international literature in English syllabus taken by students around the world aged approximately 16 years – Cambridge IGCSE Literature in English (Cambridge International Education, 2024). The research aimed to explore any differences in open-book and closed-book exams in this context, to provide insights that can inform syllabus development and future research. The research focused on differences in student outcomes, essay features, and assessor perspectives, and facilitated some exploration of the constructs being assessed.

In particular, this research explored essay questions from a drama component requiring students to compose essays about dramatic texts in time-constrained conditions as part of an exam. The essay aims to assess a combination of knowledge and skills: knowledge in the form of detailed knowledge of the texts, supported by references; and skills in the form of understanding the meanings of the texts, exploring beyond surface meaning, appreciating how language, structure and form are used to create and shape meanings and effects, and communicating a sensitive and informed personal response (Cambridge International Education, 2024). In Cambridge IGCSE Literature in English, there are open-book and closed-book versions of the relevant exam, where open-book refers to a copy of a clean (unmarked and unannotated) text. The essay questions, time constraints, other assessment conditions and assessment objectives remain the same between versions. In the relevant exam, students have some choice over which questions they answer.

¹ Note the term *literature in English* in this article is used to refer to the study of written works in the English language.

While existing research has identified the differences between open-book and closed-book exams across various dimensions in differing contexts, few studies have explored the specific example of literature in English. Table I shows a summary of the themes explored in the research literature in the UK and abroad and discusses how they may apply to the international literature in English context.

Table I: Themes found in existing research on open-book and closed-book exams, as applied to the context of literature in English

1	Materials permitted	Different contexts may allow different forms of open-book exams, for example, ranging from students being allowed their textbooks, notes, or free use of the Internet. In the context of Cambridge IGCSE Literature in English, open-book exams permit students access to a clean copy of a literary text without any added notes or annotations. Research from the UK context highlights some challenges with this approach, including the cost and administrative implications for schools to purchase and provide the clean texts for the exams. There could also be variations among schools as to which version of the text the students have access to, and some have publisher-made notes or summaries, which students could use during their exams (Al Malki et al., 2023; Didau, 2017; UK Parliament Debate, 2017).
2	Constructs assessed	Proponents of open-book exams argue that they facilitate the assessment of higher-order skills rather than memorisation (Abu-Snoubar et al., 2022; Al Malki et al., 2023; Ben-Chaim & Zoller, 1997; UK Parliament Debate, 2017). However, this will depend on how open-book and closed-book exams are implemented – including the assessment questions, material allowed, and time constraints. In settings where providing access to a text leads to exam questions being designed to be more applied, for example, then major differences in the constructs assessed in open-book versus closed-book exams may be observed. In contrast, in other cases, adopting an open-book policy may not automatically alter the constructs being assessed in the exam (Francis, 1982). In the case of international Literature in English (Cambridge IGCSE), the constructs assessed are intended to be the same in both open-book and closed-book versions of the assessment (Cambridge International Education, 2024). Both versions assess the same combination of knowledge and skills (also noted by Stock, 2017 in the UK context), and the exam questions and marking criteria remain the same in both conditions. However, one potential nuance in the assessment constructs could arise if long word-for-word quotations are expected and valued in students’ essays. This might create a small difference in assessment constructs between the two versions. For example, the closed-book exam might promote more memorisation of long verbatim quotations.
3	Time constraints	Students are commonly perceived by teachers to “waste time” in open-book exams searching for information (Francis, 1982; Marsh, 2017; Pena, 2023; Quentin, 2022; UK Parliament Debate, 2017). This could occur if students do not know their texts thoroughly and rely on exam time to read through the text and search for references or content. Teachers are concerned about how this might influence their students’ performance.

4	Exam preparation and learning	Research has shown that students may spend less time preparing for open-book exams compared with closed-book exams in some contexts (Abu-Snoubar et al., 2022; Al Malki et al., 2023; Eilertsen & Valdermo, 2000; Francis, 1982; Pena, 2023; UK Parliament Debate, 2017), and this could also apply to literature in English. Students may perceive open-book exams as easier and requiring less preparation, which may not be the case, and this could impact negatively on their performance. This, however, can arguably be mitigated by teaching strategies and exam practice in open-book conditions (Feller, 1994; Marsh, 2017).
5	Anxiety	Research in other settings has shown that students perceive open-book exams as less anxiety-provoking (Abu-Snoubar et al., 2022; Ben-Chaim & Zoller, 1997; Eilertsen & Valdermo, 2000; Feller, 1994; Francis, 1982), and this could also apply to the case of literature in English, although there is no direct evidence in this setting.
6	Student and teacher perceptions	Stakeholders have differing views about open- and closed-book exams, but there is some relevant evidence suggesting that students view open-book exams more positively (Abu-Snoubar et al., 2022; Ben-Chaim & Zoller, 1997; Permzadian & Cho, 2023; Pravini et al., 2019), which could also apply to international literature in English. Some stakeholders equate closed-book exams with memorisation, and open-book exams with higher-order skills (although, as noted, this depends on implementation). Some perceive that open-book exams are more progressive and authentic (Abu-Snoubar et al., 2022; Al Malki et al., 2023; Feller, 1994; Francis, 1982; UK Parliament Debate, 2017), while others might argue that closed-book exams are more rigorous (UK Parliament Debate, 2017). These various perceptions are also likely to exist in the Cambridge IGCSE Literature in English context.
7	Question design and marking	If open-book exams are designed specifically to take advantage of the book, the exam design and target constructs could be different. This could have implications for assessment design, validity and marking (Al Malki et al., 2023; Eilertsen & Valdermo, 2000; Feller, 1994). This does not appear to be the case in Cambridge IGCSE Literature in English, where the exam design and constructs are intended to be the same in both formats. However, an area to explore further is marking, and whether it differs for open-book and closed-book responses.
8	Student outcomes	There is a lack of recent and relevant evidence about the impact of different types of open-book and closed-book exams on student outcomes. Some research has found better outcomes in closed-book exams (e.g., Rummer et al., 2019), others found no difference (as noted by Abu-Snoubar et al., 2022; Eilertsen & Valdermo, 2000), and some found open-book exams resulted in better outcomes (as noted by Al Malki et al., 2023; Ben-Chaim & Zoller, 1997). The mix in findings likely reflects the fact that outcomes may differ depending on many factors such as the constructs being assessed in each exam, the design of the exams, student experience with each format, motivation levels and so on. Thus, studies particular to literature in English and referring only to access to a literary text are needed.

Thus, as Table I shows, the literature review reveals the complex and multifaceted nature of this topic. Differences between open-book and closed-book exams will depend heavily on methods of implementation and enactment, within an overarching assessment design and context. The research literature evidences a wide range of meanings of open-book and closed-book exams, with variations as to what materials are permitted and how. There is a lack of evidence about how open-book or closed-book exams may influence the characteristics and quality of

student essays, marking, teaching and learning, or student results in literature in English. These areas were explored further in this research.

Research aim

This research was an initial attempt at exploring differences between open-book and closed-book exams in the context of international literature in English. The following three research questions were of interest:

1. What might the differences in essays written in open-book and closed-book conditions be with regard to the nature, length and quality of references to the text?
2. How might expert assessors in international literature in English perceive the impact of open-book or closed-book examinations on marking, teaching and learning?
3. What were any differences in student outcomes internationally?²

Methods

As already described, this research was conducted in the context of Cambridge IGCSE Literature in English using essay questions about dramatic texts. There are open-book and closed-book versions of the relevant exam, where open-book refers to a copy of a clean (unmarked and unannotated) text. Time constraints and assessment objectives remain the same between versions, however, the essay responses from students in open-book versus closed-book conditions were written as part of different exam papers. In the open-book exam version, students wrote one essay which constituted the entire paper, and in the closed-book exam version, students wrote the essay as well as one other extended response exam question worth the same number of marks. However, any effects of this would likely be very minor, as the number of marks and the amount of time available per question were the same. In the exams, students have a choice among different question options.

A multimethod approach was taken, using different methods to explore different aspects of the research aim (Hunter & Brewer, 2016). The following paragraphs describe the methods for research questions 1 and 2, followed by research question 3.

For research questions 1 and 2, four expert assessors were recruited to evaluate samples of essays taken from open-book and closed-book conditions in Cambridge IGCSE Literature in English. In early 2025, they worked remotely to identify textual references made by students in the essays, to evaluate them according to the quality of selection and application, and to record the type of textual reference. Quality criteria were provided by the researcher for the assessors to use. These were guided by the assessment objectives and adapted for this research in consultation with an expert. The quality of selection and use of references were evaluated on a scale ranging from Excellent to Poor, with brief

² The research questions were separated into different studies, but are reported together here for ease of reading.

descriptors provided for each level³. References were also categorised into type, for example, direct reference or paraphrase⁴. Before assessors began identifying and evaluating textual references, training was conducted, definitions agreed, and practice examples completed together to ensure the criteria were applied similarly by the different assessors. Each assessor evaluated a different set of essays, consisting of similar proportions of open-book and closed-book essays. They were not told which essays were from open-book or closed-book conditions. Ethical research principles according to the British Educational Research Association (BERA) were adhered to throughout the study⁵.

Two samples were drawn from each exam condition, each consisting of 17 essay responses, taken from the Cambridge International exam series in June 2023. The essay question was about Shakespeare's *Othello*⁶. Each sample was selected using random stratified sampling from the total pool of essays in each condition. This helped ensure that they were broadly representative of different student ability levels and represented different regions and schools internationally⁷. Each of the two samples had a similar mean to the overall population of students who took the relevant exams, for the essay mark as well as prior attainment. Results on a similar assessment on poetry and prose, taken in the same exam session as part of the same qualification, were used as a proxy for prior attainment. The poetry and prose exam has the same assessment objectives and weightings, and also requires extended writing responses from students. For the drama exams that were the focus of this research, the open-book sample scored approximately one mark higher for their essays (out of a total of 25 marks) on average compared with the closed-book sample. The data for research question 1 was analysed predominantly with descriptive statistics.

To provide insights for research question 2 (assessor perspectives), after completing their evaluations, the assessors took part in individual 30-minute semi-structured interviews held via video conferencing. They were asked to share their views, experiences and perspectives of the impact of format on essay writing, teaching and learning, as well as marking. The interviews were transcribed and analysed thematically according to the research question.

For research question 3 (differences in outcomes), results data was obtained from our internal databases for essays written in the June 2023 and 2024 exam

3 For example, excellent selection of reference was "The reference is selected skilfully with flair, and is highly relevant" and excellent use of reference was "The discussion of the references is sensitive and with considerable detail to the way the writer uses language to achieve effects". The scale was created based on the relevant assessment objectives for the essay, in consultation with an expert assessor.

4 *Direct references* were when students quoted from the text, whereas *paraphrase* referred to students describing parts of the text in their own words.

5 The research did not require a formal ethical review by the Cambridge University Press & Assessment Research Ethics Committee.

6 The question was, "What do you think makes Shakespeare's portrayal of Iago so fascinating?"

7 Regions covered were Denmark, Uganda, South Africa, Vietnam, Netherlands, France, UK, India, Czechia, Tanzania, Indonesia, Romania, Thailand, Ghana, Malawi and Kenya. Initially there were 20 essays in each condition, however, the samples were restricted to ensure a better match of prior literature in English attainment.

series⁸. The data was cleaned, prepared and analysed using SAS Enterprise Guide. Descriptive statistics were computed by year, region and question for the essay options with relatively larger numbers of students. The descriptive statistics explored the mean marks for different essay questions, as well as the age, gender and prior attainment of the students. The marks on the similar poetry and prose assessment (explained previously) were again used as a proxy for prior attainment.

Descriptive statistics are limited because comparing mean marks does not account for the fact that other variables might influence student marks, such as school or prior attainment. Therefore, regression analyses were carried out to control for some (though not all) potentially influential variables. The regression analyses aimed to explore the effect of open-book and closed-book exams on essay performance, after controlling for prior literature in English attainment age, gender and school. The control variables were selected based on relevance and availability. The significance level was set at 0.05. The regions covered included the UK, India, Kenya, France, Malaysia, Canada, Spain and Thailand, which provides a reasonably good spread of data across different geographical areas. The data was assessed to ensure it was appropriate for regression analysis before proceeding. Hierarchical regression (or multilevel modelling) was chosen because the data comes from students in different schools (i.e., has a hierarchical or clustered structure⁹) (University of Bristol Centre for Multilevel Modelling, 2024).

Separate models were fitted for each region, as different questions were the most popular in different regions. It was also possible that there might have been regional differences in the findings. The questions were modelled separately, as they were different questions about different drama texts. The models allowed for fixed and random effects. School was included as a random effect, as there were many schools per dataset and the intention was to generalise beyond these specific schools. The essay questions covered Shakespeare and non-Shakespeare texts. Some of the options included in the analysis had relatively smaller sample sizes and thus lower statistical power, but were included in the analysis and reporting to provide some additional insights over a wider range of questions. However, these results should be interpreted with caution and seen as exploratory.

Findings

The main findings for each research question are presented, followed by a discussion. Values were rounded where relevant to aid interpretation.

Research question 1: Differences in essays

Analysis of the assessors' evaluations of the characteristics of the 34 essays showed that the essays in the open-book sample performed more favourably

⁸ This data was collected as part of the usual marking and processing of students' examination scripts and has been stored and used in line with Cambridge University Press & Assessment's Data Privacy notice (<https://www.cambridge.org/legal/candidate-privacy-notice>).

⁹ Students from a school would tend to be more similar in terms of their outcomes than students from different schools, due to a variety of school characteristics (such as teaching approach or geographical location).

according to the descriptive statistics that were calculated. Overall, the essays written in open-book conditions in this sample had a slightly higher number of references to the text than those written in closed-book conditions (as shown in Table 2), although this trend did not reach statistical significance¹⁰. On average, each open-book essay had about one more reference (based on the median) and 1.9 more references (based on the mean) than the closed-book essays. There was also more variability (indicated by the wider range and higher standard deviation) for the essays in the open-book sample compared with the closed-book sample. Although the samples were selected using stratified random sampling to have similar prior attainment and attainment, and they covered a range of schools and regions, there could be sample bias affecting the results.

Table 2: Number of references to the text in each condition.

Sample	N	Total number of references to the text	Mean	SD	Median	Range
Open	17	95	5.59	3.66	5	1 to 13
Closed	17	63	3.71	2.20	4	0 to 8

The findings from the descriptive statistics further showed that essays in the open-book condition had longer direct references to the text, and less paraphrasing, as illustrated in Figure 1.

¹⁰ A Wilcoxon rank-sum test was conducted, and the difference was not statistically significant ($z = -1.3058$, $p = 0.0958$). The analysis was, however, limited by small sample sizes.

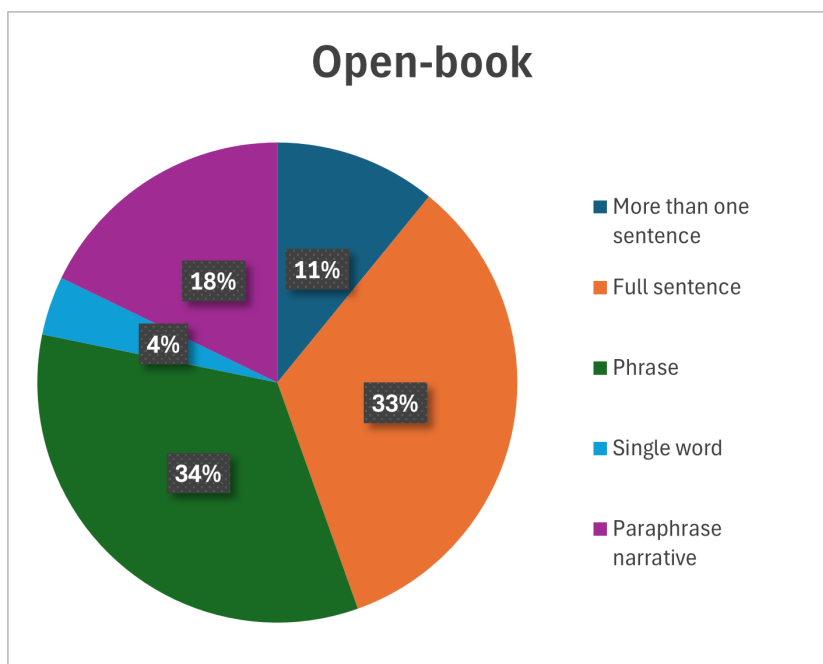
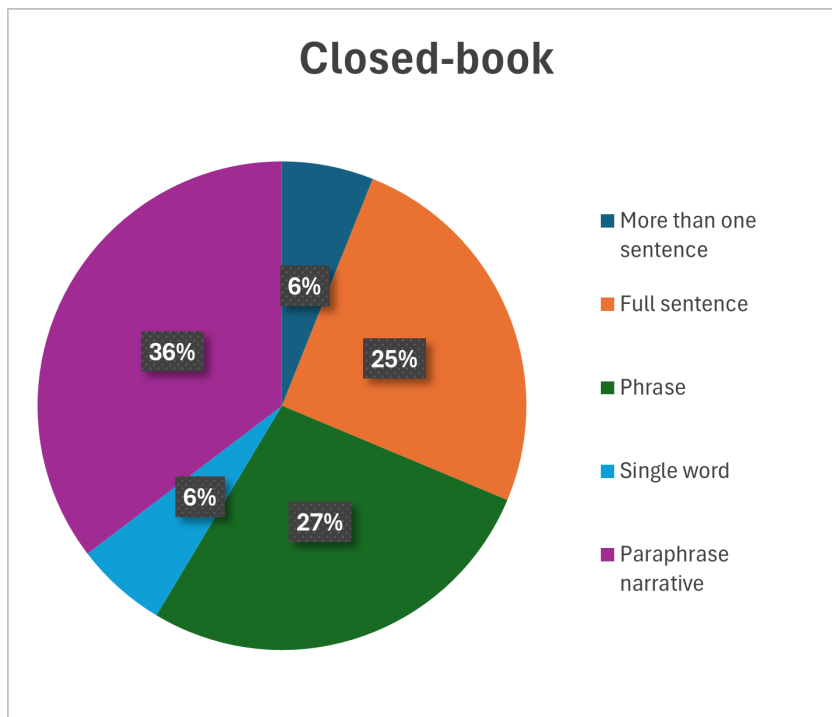


Figure 1: Pie charts showing the percentages of different types of references in open-book and closed-book conditions. (All categories except “paraphrase narrative” refer to direct quotations.)

Table 3 shows that the quality of selection of references was generally slightly better in the open-book sample of essays.

Table 3: Quality of selection of the references as percentages (N in brackets)

Sample	Excellent	Good	Average	Poor	Missing	N total
Open	33% (31)	43% (40)	21% (20)	3% (3)	(1)	95
Closed	29% (18)	27% (17)	31% (19)	13% (8)	(1)	63

Table 4 shows that the quality of use of the references was also generally better in the sampled essays written in open-book conditions, although a similar percentage scored “Excellent”.

Table 4: Quality of use of the references as percentages (N in brackets)

Sample	Excellent	Good	Average	Poor	Missing	N total
Open	14% (13)	33% (31)	26% (24)	28% (26)	(1)	95
Closed	13% (8)	11% (7)	34% (21)	42% (26)	(1)	63

Thus, in summary, essays written in the open-book conditions in this sample had slightly more and longer direct references to the text, and they also appeared to be better quality in terms of how well they were selected and how well they were applied in general. Although this is based on relatively small samples, these exploratory descriptive insights show that students in open-book conditions were able to select and apply a number of references well, with access to the book.

Research question 2: Assessor perspectives

Assessor perspectives of the differences between open-book and closed-book exams, and their potential impact on teaching, learning and marking, were explored in research question 2. In addition to their assessment roles, the assessors had teaching experience, and they drew upon this experience when answering the interview questions.

In relation to the differences between essays, the assessors noted that they could not necessarily differentiate between essays written in each condition, and they perceived that differences in the quality of the essays depended far more on teaching and the students themselves rather than the format.

The perceived advantages and disadvantages they shared reflected common themes found in the literature review, for example, the idea that students might experience reduced stress or anxiety in open-book conditions. For example, one assessor said:

“... thinking about the poor students and the horrible pressure of doing exams. Yeah, I get why they would want the book.”

They also spoke about the possible impact of open-book exams on how students prepare for and write their exams. The concern about wasting time arose, as noted by one assessor:

“... there are a number of pitfalls that students can fall into if they’re not careful, one of which is not knowing the text well enough to be able to pinpoint the particular quotation they’re looking for in the exam. And it can waste time, especially if they’ve got a very long text. And they’re not too sure

where the incident is that they want to refer to. So it doesn't necessarily make it easier for the candidate ... If they are copying chunks from the text with no comment, sometimes not particularly relevant, so it's time wasted and nothing gained."

The assessors also noted the financial impact of schools supplying clean texts, and the potential impact on students if they are not able to annotate their texts to keep them clean for the exams (if schools cannot afford two copies). Assessors stated the following, for example:

"It's a financial burden".

"I think that there's disadvantages for the centres as well because the text that's taken into the exam can't be marked, can't be annotated by the student, they can't be highlighted. So the students can't mark the text as they're learning."

A concern about fairness arose. Some assessors felt that, in cases where students have access to versions of the text containing publisher notes, annotations and glossaries, this could put them at an advantage over students without that. This was noted as a particular concern for Shakespeare texts. One assessor stated:

"It's impossible to buy a Shakespeare text that doesn't have some notes ... That leaves me feeling very uncomfortable."

However, another assessor questioned how much help these notes would actually be for achieving the higher marks.

When reflecting on the impact on teaching and learning, some assessors felt that preparation for both modes would be very similar, and good knowledge of the text is needed for both, thus supporting the idea that the same constructs are assessed in both. For example, one assessor stated:

"Yeah, I'm not so sure it would affect the teaching because when I've been preparing students for open and closed books exactly the same way and said you need to learn these, you know these, you know, sit and learn them like you were learning lines for a play."

There were differing views regarding the importance of learning quotations by heart, and one assessor raised an interesting concept of different teaching methods for quotations that are not rote learning approaches. For example, hot seating, where students assume the role of different characters and interact with each other using words from the text.

In terms of marking, the assessors reported that marking exams is generally the same regardless of whether the essay is from an open-book or closed-book exam, due to the assessor training processes and the mark scheme. They noted that they are trained to focus on the mark scheme and the assessment objectives when marking.

“We use the same mark scheme. And we don’t make any concessions to whether it’s open or closed books, we mark on the strength of their response. Well, you know how far they are hitting the assessment objectives. So it is not something that we consider. We don’t consider that when we’re awarding a mark.”

“We mark what’s in front of us, regardless of where it’s come from.”

During live exam marking, assessors are aware which essays are from open- and closed-book conditions, but they felt this did not influence their marking. Some of them did not rule out the possibility of bias entirely but perceived that any chance or impact of bias in this regard would be negligible or very small.

Research question 3: Differences in outcomes

Research question 3 explored, through statistical analyses, differences in the outcomes for students taking different options. The rows in Table 5 “Which mean was higher?” and “By how much?” show the differences in means for open-book versus closed-book exams in each region and option (i.e., exam question) included in the analysis. These descriptive statistics show that there were generally small differences in student outcomes depending on whether exams were open-book or closed-book. The differences in means sometimes favoured open-book exams, and other times favoured closed-book exams. The differences ranged from 0.23 marks for Kenya in 2024 (option 2b) to 6.05 marks in Canada in 2023 (option 5b), both favouring the open-book exam (and out of a total of 25 marks). Table A1 in Appendix A shows the differences in outcomes in more detail. Note that the regions from different time zones receive equivalent versions of the exam papers (i.e., the exam questions are different). All the regions in Table 5 were in time zone 2 except for Malaysia and Canada which were in time zone 3.

Table 5: Summary of differences in essay marks between open-book and closed-book versions in the regions and essay options included. This includes the differences in means from the descriptive statistics, as well as the parameter estimates for the effect of the open / closed book variable obtained from the regression analyses.

Region (option)	Time zone	Year (June)	N	N Open	N Closed	Which mean was higher?	By how much?	Parameter estimates for the open / closed book variable (absolute values)
Thailand (2b)	2	2023	85	36	49	Closed	0.61	0.65
UK (2b)	2	2023	197	80	117	Closed	3.09	1.07
India (4b)	2	2023	174	67	107	Open	2.85	1.85
France (4b)	2	2023	84	56	28	Closed	0.36	0.53
Kenya (4b)	2	2023	76	26	50	Closed	0.38	0.66
Kenya (5b)	2	2023	271	141	130	Open	1.09	1.01
India (5b)	2	2023	96	43	53	Open	2.97	2.10
Malaysia (2b)	3	2023	99	23	76	Closed	1.50	1.25
Canada (5b)	3	2023	67	32	35	Open	6.05	3.32
India (3b)	2	2024	203	53	150	Open	1.68	0.21
Thailand (3b)	2	2024	131	46	85	Closed	2.06	1.96
India (4b)	2	2024	104	24	80	Open	1.82	2.01
Kenya (2b)	2	2024	156	30	126	Open	0.23	0.22
Kenya (3b)	2	2024	101	63	38	Open	1.43	0.61
Kenya (5b)	2	2024	100	31	69	Open	0.51	0.87
Spain (3b)	2	2024	74	18	56	Closed	2.06	1.12
UK (3b)	2	2024	80	18	62	Closed	3.25	0.43

Note: the maximum mark available for each essay was 25. The numbers and letters after each region (e.g., 2b) are the question options¹¹. Where the regression modelling found a significant effect for the open / closed book variable, the parameter estimate is shown in bold. The shading in the table highlights which of the statistically significant findings were in favour of open-book exams (lighter grey) versus closed-book exams (darker grey). The parameter estimates are presented as absolute values, as in all cases the effects were in the same direction as the differences in means.

As described earlier, hierarchical regression was used to model the results data for each region and exam question. The regression analyses aimed to explore the effect of open- and closed-book exams on essay performance, after controlling for prior literature in English attainment, age, gender and school. The key results of the regression analyses (parameter estimates for the open / closed book variable) are shown in the final column of Table 5. They are presented as absolute values, as in all cases, the parameter estimates were in the same direction as the differences in means. Parameter estimates in bold indicate results that were

¹¹ 2 023: 2b was Sherriff's, *Journey's End*; 4b was Shakespeare's *Twelfth Night*; 5b was Shakespeare's *Othello*. 2024: 2b was Soyinka's *Death and the King's Horseman*; 3b was Williams' *A Streetcar Named Desire*; 4b was Shakespeare's *A Midsummer Night's Dream*; 5b was Shakespeare's *Othello*. Note that the same option number in different time zones within the same year will have been a different essay question but about the same text.

statistically significant (at an alpha of 0.05). Note the relatively small sample sizes in several cases in Table 5, which limit statistical power¹². Full regression results for the statistically significant findings can be found in Appendix B.

Table 5 shows that, in most cases, there were small and non-significant differences in performance between students who wrote their essays during open-book versus closed-book exams, after controlling for age, gender, prior attainment and school. Overall, the results suggest that there may not be any major differences between the two versions (open-book versus closed-book) in terms of student results. However, there were some exceptions, particularly in India, in which there were small but statistically significant differences between the two versions for three Shakespeare essay questions in June 2023 and 2024. The difference was approximately two marks (favouring the open-book version). Students in Thailand, in contrast, did approximately two marks better (out of 25 marks) in the closed-book version for one essay question in 2024. These initial insights suggest that there may be regional differences in the impact of open- and closed-book exams, although quite small.

Discussion

Gathering data from multiple methods allowed exploration of different aspects of the research aim. Firstly, for research question I, exploring the essay characteristics in different conditions allowed initial insights into the use of textual references (although based on a sample). These exploratory descriptive insights show that students in open-book conditions were able to select and apply a number of references well, with access to the book. In general, it does not seem to be the case that students writing open-book exams in this sample were repeating long references without selecting and applying them well. Compared to the closed-book sample, essays written with access to the text had slightly more references to the text, which were longer and tended to be direct references rather than paraphrases. They were also better quality in terms of their selection and application in general. However, this may not lead to significantly higher marks (as the mean marks for the open-book sample of essays was only one mark higher) and sample bias cannot be ruled out.

The data from research question I also highlights the different routes and strategies that students may use to gain marks, depending on whether they have access to the book or not. For example, whether students focus on gaining marks from textual references or other areas of the assessment objectives (as marking is holistic). This indicates, not a change in the overall assessment constructs, but perhaps a slightly different focus in different conditions. In the UK context, it has been questioned whether open- and closed-book literature in English exams may have differential effects depending on students' abilities (Marsh, 2017; UK Parliament Debate, 2017). The numbers for different ability groups in this research were not large enough to make definitive claims about this, but this would be a useful avenue for further research.

¹² Online sample size calculators consulted suggested a minimum total sample size of 84, if we expect a medium effect size, power of 0.8 and significance level of 0.05 ([Free A-priori Sample Size Calculator for Multiple Regression - Free Statistics Calculators; Sample size calculator](#)).

For research question 2, gathering perspectives from a small group of expert assessors facilitated insights into any potential impact on marking, which could inform understanding of how constructs might differ in the different conditions. The assessors drew on their assessment as well as their teaching experience when answering the interview questions. The findings indicated that marking was not perceived by expert assessors to be influenced by format. The marking is holistic, and it is what students do with the references to the text that was perceived to be most important. The assessors stated that exact verbatim quotations were not necessary. This provides evidence for the argument that the constructs being assessed are broadly the same in the open- versus the closed-book exams, as intended.

Mainly disadvantages were raised by assessors in relation to open-book exams in this context, although a potential benefit of easing student anxiety was noted. The latter echoed the literature review, where research in other settings found that students perceive open-book exams as less anxiety-provoking (Abu-Snoubat et al., 2022; Ben-Chaim & Zoller, 1997; Eilertsen & Valdermo, 2000; Feller, 1994; Francis, 1982). The disadvantages raised included the potential impact on exam preparation and exam taking, the requirement for clean texts and the financial burden associated with this. Assessors raised a point found in the literature review, which is the common perception that students waste time in open-book exams (Francis, 1982; Marsh, 2017; Pena, 2023; Quentin, 2022; UK Parliament Debate, 2017). This was not directly explored in this research, but the analyses indicate that this may not necessarily be the case; if students ran out of time we would expect to see lower marks in open-book conditions, which was not a clear trend according to the results data in research question 3. As the literature review highlights, the effects of open-book and closed-book exams may depend on teaching and exam preparation. Questions were raised about fairness, for example, if some students have access to texts that have publisher notes, summaries or annotations which could inappropriately help them in their exam (although the usefulness of this was questioned). These themes were also found in the literature review (Al Malki et al., 2023; Didau, 2017; UK Parliament Debate, 2017). In terms of a possible impact on teaching and learning, the topic of memorisation arose, and strategies to help students that move away from rote learning approaches were discussed. Different views were expressed on the value of memorising quotations and these perspectives also echo the tensions found in the literature review. Assessors discussed the role of memorising quotations, noting its potential value while also questioning its place in our modern world.

Finally, in research question 3, exploring the impact on results helped to quantify any differences that may be present and how they may influence students' exam results. Although not all potentially confounding variables could be controlled for, and sample sizes were limited, the results suggest possible regional differences in the effects of open-book and closed-book exams on student results. These findings could have some significance in practice for student outcomes, although the differences were relatively small and the essay is just one part of a wider set of assessments. Possible reasons for the differences in performance include regional teaching, learning and exam preparation or administration practices. Future research would be useful to add more evidence from different regions and years, and different questions and texts.

Limitations

Before concluding, the main limitations of this research are acknowledged. This research was small scale and exploratory in nature, providing insights into a highly complex and under-researched area.

For research question 1, the research was conducted using essays written about a Shakespeare text. Such texts available to schools reportedly have publisher chapter notes, summaries and annotations that add to the features of the open text environment beyond the drama text itself. Research with different texts without these publisher notes would be useful. Another limitation is the limited samples for the comparison of textual references, although the samples were selected to cover a range of regions and similar attainments. There is a possibility of sample bias, and replication with more data or another sample is important. The analysis was limited to descriptive statistics for the most part. As students were not observed during their essay writing as part of this research, we also lack details about how students may have used or consulted their books during their exams.

For research question 2, the perceptions of the assessment were gathered from four assessors in the UK context, so the findings should be considered as insights and points for discussion rather than generalisations. The assessors' insights are likely to be based on their experiences of teaching in the UK context, and research with teachers and learners in different regions internationally would be useful.

When comparing outcomes (research question 3), there was a limited set of regions and options explored. In several cases there were small sample sizes for some of the groups, which lowers the statistical power in these cases. Additionally, only a subset of variables was controlled for due to data availability. Marks on the poetry and prose paper were used as a proxy for prior attainment, albeit an imperfect one. More data over a longer time period would be useful to analyse.

Finally, applicable to both research questions 1 and 3, the essay responses written by students in open-book versus closed-book conditions were written as part of different exam papers. In the open-book exam version, students wrote one essay which constituted the entire paper, and in the closed-book exam version, students wrote the essay as well as one other extended response exam question worth the same amount of marks. However, any effects of this would likely be very minor, as the number of marks and the amount of time available per question were the same.

Conclusion and implications

This research was a first step into exploring differences in open-book and closed-book exams in a particular context – using an example of international literature in English essays – to explore the impact of having the text available, with the same assessment questions, marking criteria and time constraints. Overall, it was found that there are merits and disadvantages of both modes, and the potential for regional differences. The latter could perhaps be due to the influences of cultural differences, perspectives, or different teaching and exam preparation and administration strategies. This research helped to reveal the nuances related

to the context of literature in English, and findings may apply to other similar contexts internationally (with similar definitions and implementation methods for open-book exams). The evidence about differences in textual references was limited by the sample but highlighted that students can use different strategies to do well. The data provided some initial insights into how students in open-book conditions select and apply textual references in comparison with students in closed-book conditions.

The research highlights the debate that exists around open- and closed-book exams and whether they assess different constructs. The data supported the notion that, in this case, the two exams assess broadly the same constructs, with potentially slight variation in the focus on constructs depending on student strategies. As noted, the impact of open-book and closed-book exams will vary depending on how they are implemented and administered, within an overall assessment design.

To conclude, this is a complex issue, and this exploratory small-scale research provides initial insights about the differences in open-book and closed-book exams in this context to inform syllabus development and identify future research needed. Assessment designers can use the themes explored in this study to investigate how open-book and closed-book exams may differ in their particular context. In the case of literature in English, the findings can help inform an assessment strategy that considers the advantages and disadvantages of different modes of assessment, the link with teaching, learning and administration practices, and considerations of how the impact may vary regionally.

This study invites future research in various areas. Research on the differences in the nature of essays with larger samples of student essays would be useful. Additionally, exploring the differences in essays for a variety of different texts, beyond Shakespeare, would add to the literature. Observational research exploring how the books are used by students during their exams and engaging with student voice would also be useful avenues for further work. Research into different teaching methods and their effectiveness in preparing students for open-book and closed-book exams is recommended. Finally, future research is needed to explore the differences as applied to different contexts and types of implementations (for example, students being allowed different materials). Only in this way can we avoid generalisations about open- and closed-book exams and understand the nuances.

References

Abu-Snoubar, T. K., Aldowkat, I., Al-Shboul, Y., Atiyat, M. A., & Al-Hyari, H. (2022). *The attitudes of Jordanian English language and literature undergraduate students toward open-book exams*. *Frontiers in Education*, 7, 1050587.

Al Malki, M., Al'Adawi, S., Al-Azzawi, N., & Al-Abri, K. (2023). *Open-book-based assessment during COVID-19: Challenges and opportunities*. *International Journal of Learning, Teaching and Educational Research*, 22(8), 146–168.

Ben-Chaim, D., & Zoller, U. (1997). *Examination-type preferences of secondary school students and their teachers in the science disciplines*. *Instructional Science*, 25(5), 347–367.

Cambridge International Education. (2024). *Cambridge IGCSE Literature in English (0475): Syllabus overview*.

Didau, D. (2017). *Are 'closed book' examinations a bad idea?*

Eilertsen, T. V., & Valdermo, O. (2000). *Open-book assessment: A contribution to improved learning?* *Studies in Educational Evaluation*, 26(2), 91–103.

Feller, M. (1994). *Open-book testing and education for the future*. *Studies in Educational Evaluation*, 20(2), 235–238.

Francis, J. (1982). *A case for open-book examinations*. *Educational Review*, 34(1), 13–26.

Hunter, A., & Brewer, J. D. (2016). *Designing multimethod research*. In S. N. Hesse-Biber & R. B. Johnson (Eds.), *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry* (pp. 185–205): Oxford Library of Psychology.

Marsh, C. (2017). *Poetry and assessment: An investigation into teachers' perceptions of the impact of closed book examinations on teaching and learning at GCSE*. *English in Education*, 51(3), 275–293.

Pena, S. R. M. (2023). *Open and closed-book assessments in secondary school mathematics*. [Master's thesis, University of Toronto (Canada)].

Permzadian, V., & Cho, K. W. (2023). *Does the format of an assessment (closed book or open book) affect learning? A systematic review of the literature*. *Teaching in Higher Education*, 30(4), 880–897.

Pravini, H., Hansakie, D., & Amunugama, A. (2019). *A Study on attitudes of teachers and students towards implementing an open-book examination for GCE O/L English Literature*. [Paper presentation]. The Undergraduate Research Symposium (HUG 2019), Department of English Language Teaching, Faculty of Humanities, University of Kelaniya, Sri Lanka.

Quentin, A. (2022, March 25). *Open-book exams: What does the evidence say?* *AQi Blogs*.

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

Stock, N. (2017). [Deconstructing the divergence: Unravelling the 2013-2015 reforms in GCSE English Language and Literature](#). *English in Education*, 51(2), 143–156 .

UK Parliament Debate. (2017). [GCSE English Literature exams: Volume 624](#).

University of Bristol Centre for Multilevel Modelling. (2024). [What are multilevel models and why should I use them?](#)

Appendices

Appendix A. Descriptive statistics – student outcomes (research question 3)

Table A1: Mean marks and standard deviations (SD) for each option for each version. In the mean marks columns, the mean marks are in bold for the version (open-book or closed-book) that had a higher mean mark. In the comparison column, mean differences between versions that are greater than one mark are shown in bold.

Region (option)	Time zone	Year (June)	Open book			Closed book			Comparison	
			Mean marks	SD	N	Mean marks	SD	N	Which mean was higher?	By how much?
Thailand (2b)	2	2023	15.17	3.29	36	15.78	3.37	49	Closed	0.61
UK (2b)	2	2023	12.18	4.11	80	15.27	3.88	117	Closed	3.09
India (4b)	2	2023	16.25	4.42	67	13.40	3.94	107	Open	2.85
France (4b)	2	2023	13.39	3.57	56	13.75	3.34	28	Closed	0.36
Kenya (4b)	2	2023	12.12	3.78	26	12.50	3.33	50	Closed	0.38
Kenya (5b)	2	2023	14.29	3.76	141	13.20	4.18	130	Open	1.09
India (5b)	2	2023	17.40	3.99	43	14.43	3.54	53	Open	2.97
Malaysia (2b)	3	2023	15.09	3.98	23	16.59	3.29	76	Closed	1.50
Canada (5b)	3	2023	14.19	3.64	32	8.14	4.10	35	Open	6.05
India (3b)	2	2024	15.89	3.51	53	14.21	3.99	150	Open	1.68
Thailand (3b)	2	2024	13.89	3.17	46	15.95	4.18	85	Closed	2.06
India (4b)	2	2024	15.38	3.33	24	13.56	4.71	80	Open	1.82
Kenya (2b)	2	2024	12.17	2.55	30	11.94	3.66	126	Open	0.23
Kenya (3b)	2	2024	14.67	4.23	63	13.24	5.03	38	Open	1.43
Kenya (5b)	2	2024	11.90	3.96	31	11.39	3.98	69	Open	0.51
Spain (3b)	2	2024	11.39	3.43	18	13.45	3.12	56	Closed	2.06
UK (3b)	2	2024	12.56	4.48	18	15.81	4.01	62	Closed	3.25

Appendix B. Statistically significant regression results (research question 3)

The below tables show the statistically significant results from the regression analyses (fixed effects followed by random effects) by region. This analysis explored the effect of open-book and closed-book exams (O or C) on essay performance, after controlling for prior literature in English attainment (total mark on the poetry and prose paper), age, gender (male or female, M or F) and school (included as a random effect).

Table B1: Fixed effects solution for India question 4b in 2023

Effect	Gender	O or C	Parameter estimate	Standard error	Degrees of freedom	t-statistic	P value	Lower	Upper
Intercept			-0.73	7.96	13.00	-0.09	0.93	-17.92	16.47
Total mark poetry and prose			0.29	0.03	156.00	8.93	<.0001	0.23	0.36
Age			0.48	0.48	156.00	1.00	0.32	-0.47	1.44
Gender	F		0.19	0.52	156.00	0.36	0.72	-0.85	1.22
Gender	M		0.00	.	.	.	.	.	.
Open-book or closed-book		C	-1.85	0.74	156.00	-2.52	0.01	-3.30	-0.40
Open-book or closed-book		O	0.00	.	.	.	.	.	.

Table B2: Fixed effects solution for India question 5b in 2023

Effect	Gender	O or C	Parameter estimate	Standard error	Degrees of freedom	t-statistic	P value	Lower	Upper
Intercept			8.88	7.96	10.00	1.12	0.29	-8.85	26.61
Total mark poetry and prose			0.22	0.05	81.00	4.16	<.0001	0.12	0.33
Age			0.08	0.46	81.00	0.16	0.87	-0.84	1.00
Gender	F		0.82	0.72	81.00	1.14	0.26	-0.61	2.26
Gender	M		0.00	.	.	.	.	.	.
Open-book or closed-book		C	-2.10	0.75	81.00	-2.79	0.01	-3.60	-0.60
Open-book or closed-book		O	0.00	.	.	.	.	.	.

Table B3: Fixed effects solution for India question 4b in 2024

Effect	Gender	O or C	Parameter estimate	Standard error	Degrees of freedom	t-statistic	P value	Lower	Upper
Intercept			21.94	9.92	10.00	2.21	0.05	-0.17	44.04
Total mark poetry and prose			0.29	0.05	89.00	6.22	<.0001	0.20	0.38
Age			-0.99	0.60	89.00	-1.65	0.10	-2.19	0.20
Gender	F		1.58	0.73	89.00	2.17	0.03	0.14	3.02
Gender	M		0.00	.	.	.	.	.	.
Open-book or closed-book		C	-2.01	0.85	89.00	-2.36	0.02	-3.71	-0.32
Open-book or closed-book		O	0.00	.	.	.	.	.	.

Table B4: Variance components for the random effects for India

Region	Year	Variance components (for school random effect)	Residual	Percentage variance explained
India 4b	2023	0.57	10.71	5.50
India 5b	2023	0	11.84	0
India 4b	2024	0.04	12.86	0.31

For India, the variance components for the random effect (school) and the percentage of variance explained by the difference between each school are shown in Table B4. Differences in school accounted for a small proportion of the variance for question 4b in 2023, but not for questions 5b in 2023 or 4b in 2024.

Table B5: Fixed effects solution for Thailand question 3b in 2024

Effect	Gender	O or C	Parameter estimate	Standard error	Degrees of freedom	t-statistic	P value	Lower	Upper
Intercept			11.45	10.25	10.00	1.12	0.29	-11.39	34.29
Total mark poetry and prose			0.11	0.06	110.00	1.90	0.06	0.00	0.22
Age			-0.07	0.62	110.00	-0.11	0.91	-1.30	1.16
Gender	F		1.07	0.74	110.00	1.45	0.15	-0.39	2.53
Gender	M		0.00	.	.	.	.	.	.
Open-book or closed-book		C	1.96	0.75	110.00	2.63	0.01	0.48	3.44
Open-book or closed-book		O	0.00	.	.	.	.	.	.

The variance components for the random effects were zero (residual of 14.32) for Thailand 3b, thus differences in school did not explain the variance.

Table B6: Fixed effects solution for Canada question 5b in 2023

Effect	Gender	O or C	Parameter estimate	Standard error	Degrees of freedom	t-statistic	P value	Lower	Upper
Intercept			28.90	10.53	8.00	2.74	0.03	4.62	53.18
Total mark poetry and prose			0.29	0.05	54.00	5.95	<.0001	0.19	0.38
Age			-1.42	0.62	54.00	-2.28	0.03	-2.66	-0.17
Gender	F		1.40	0.63	54.00	2.24	0.03	0.15	2.66
Gender	M		0.00	.	.	.	.	.	.
Open- book or closed- book		C	-3.32	1.26	54.00	-2.64	0.01	-5.84	-0.80
Open- book or closed- book		O	0.00	.	.	.	.	.	.

The variance components for the random effects were 1.76 (residual of 6.20) for Canada 5b. Thus, 22 per cent of the variance was explained by differences in the school students attended.