Research in Engineering Education Symposium (REES)
Regional Conference in Engineering Education (RCEE)
& Research in Higher Education (RHEd)
September 29 – October 3, 2025
Johor, Malaysia

Final Exam Score Characteristic Following a Sudden Shift in Testing Format from Open-Book to Closed-Book

Norhuda Abdul Manaf, Tomoya Tsuji, Hafizah Farhah Saipan Saipol, Shahirah Abu Bakar, Aizul Nahar Harun and Peng Yen Liew
Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia
Corresponding Author's Email: norhuda.kl@utm.my

ABSTRACT

BACKGROUND & CONTEXT

The impact of switching from an open-book to a closed-book format for midterm examinations was evaluated among second-year undergraduate chemical engineering students.

PURPOSE OR GOAL

This study aims to assess the impact of the testing format shift in the midterm on the characteristics of final exam scores and to explore correlations between these scores.

APPROACH OR METHODOLOGY

An exploratory ex-post-facto observational design was employed to compare midterm and final exam results from two consecutive years. In 2023, the midterm exam was conducted as an in-person, open-book test, while in 2024, it was conducted as an in-person, closed-book test. Final exams in both years remained in an in-person, open-book format. The mean and standard deviation (SD) of the examination scores were calculated. The average ratios of final to midterm exam scores and the correlations between midterm and final exam scores were evaluated. Cronbach's alpha internal consistency reliabilities were also calculated for the final exams for descriptive purposes but no formal statistical tests were run with these.

OUTCOMES

The correlation between midterm and final exam scores remained consistent across both years, regardless of the midterm testing format. The data suggest that the open-book format retained its capacity to differentiate students based on their understanding of course material, like the closed-book format.

CONCLUSIONS

The analysis underscores that open-book exams can effectively evaluate higher-order cognitive skills and comprehension, offering flexibility in exam administration without compromising assessment integrity.

KEYWORDS

Open book, closed-book, chemical engineering.

https://doi.org/10.52202/086293-0001

Introduction

Open-book exams are often perceived as easier because students can directly refer to resources to find answers. However, they may sometimes prove more challenging than closed-book exams when assessing deeper comprehension and application skills. The closed-book exam has remained a consistent tool in academic settings (Johanns et al., 2017). Yet, as the goal of education is to transfer knowledge and foster critical thinking, there is increasing interest in how different exam formats support this objective.

The comparative effectiveness of open-book versus closed-book examinations remains a widely debated topic within educational research. Open-book assessments are often perceived as less challenging due to the availability of reference materials during the examination process. According to (Williams & Wong, 2009), students generally express a preference for open-book formats, as they believed access to notes, textbooks, and other resources can enhance performance and reduce exam-related anxiety. Supporting this, (Gharib et al., 2012) reported that students tend to achieve higher average scores on open-book exams compared to closed-book formats. Although open-book exams may alleviate cognitive load by reducing the need for memorization during testing, this format may also result in more superficial initial encoding of information (Sweller, 2010). Research by (Eilertsen & Valdermo, 2000) suggested that open-book formats can promote deeper learning by encouraging higher-order thinking, synthesis, and problem-solving rather than simple recall. This is because students anticipating an open-book exam typically focus on organizing materials for quick retrieval. Whereas those preparing for closed-book assessments tend to engage in active recall and memorization practices. In addition, weaker students may struggle under time constraints in open-book exams due to challenges in locating and synthesizing relevant information quickly (Boniface, 1985).

(Boniface, 1985) observed that closed-book exams often yield greater score variability, potentially serving as a more accurate indicator of student mastery. Closed-book formats necessitate a thorough internalization of course content to perform well, and as a result, students in such settings may outperform their counterparts in open-book formats when no external aids are available (Agarwal & Roediger III, 2011). Underpinning with the study conducted by (Rummer et al., 2019) revealed a better learning outcomes for closed-book compared to open-book practice tests which align with the theoretical expectations that retrieval practice drives the testing effect (Rummer et al., 2019). While few studies have explored students' performance across different examination formats (open-book versus closed-book), the present work addresses this gap. A transition from an open-book to a closed-book examination format can substantially impact students' preparation behaviors and may elevate test anxiety levels (Gharib et al., 2012), particularly among students who are not adequately prepared for the more restrictive format. Previous research prevailed with mixed findings regarding format transitions. (Cor & Brocks, 2021) found that shifting from open-book to closed-book formats can significantly affect academic outcomes, including lower mean scores, greater score dispersion, and varied impacts across different student proficiency levels. Thus, present work is intended to evaluate the impact of switching from an open-book to a closed-book format for midterm examinations among second-year undergraduate chemical engineering students. While maintaining the final examination format in open-book format. The focus is on how this transition affects the characteristics of midterm exam scores and subsequent performance on the final examination.

Methods

An exploratory ex-post-facto observational design was employed to analyze how a change in the midterm exam format from open-book to closed-book impacted the final exam score characteristics in a particular course at one of public university in Malaysia. Data were collected and analyzed based on two consecutive cohorts of second-year chemical engineering students. In 2023, the midterm exam was conducted as an in-person, open-book test where

students were permitted to use only the assigned textbook, with no additional external resources, notes, or supplementary materials allowed. While in 2024, it was administered as an in-person, closed-book test. The final exams in both years remained in-person and open-book. This study aims to explore the patterns and relationships resulting from the change in examination format without interfering with the examination environment, and without students being aware that such an analysis was being conducted.

A common practice for the course is to conduct both midterm and final examinations in an open-book format. To ensure the consistency and reliability of the analysis, both cohorts were instructed using a similar teaching approach and applied identical course learning outcomes and assessment rubrics over 14 weeks of instruction, ensuring that the exam questions targeted similar levels of cognitive taxonomy. By maintaining these conditions, the midterm exams in 2023 and 2024 are assumed to be minimally different, so that any observed effects are attributable to the change in exam format rather than to differences in question types or content coverage. In this context, the open-book format is defined as students were allowed to bring and view a standard textbook that was used in the regular lecture. Whilst, opposite implementation for the closed-book format. In this case, students are not allow to refer to any supplementary notes/resources during the examination.

Data including midterm and final examinations grades were compiled and converted to percentage to allow for comparability. In both the open- and closed-book format versions of the exams in 2023 and 2024, midterm exams were timed and constructed of primarily 2 questions involved calculations while for the final exam questions it comprised of 5 questions. Similar weightage of marks and marking rubrics were used for both cohorts to avoid biases and were proctored in-person. The enrolment class size for each cohort of the course is 49 (2023) and 54 (2024) and not included students who were repeating the course.

Within each cohort, the mean and standard deviation (SD) of the examination scores were calculated. The average ratios of final to midterm exam scores and the correlations between midterm and final exam scores were evaluated. The distribution of examination scores was examined using the Shapiro-Wilk test of normality, which indicated significant deviations from normal distribution ($p < 0.05$). Based on this violation, the non-parametric Mann-Whitney U test was applied to compare the midterm and final exam score distribution between 2023 and 2024 cohort. This test was appropriate because it does not rely on normality assumptions and is robust to skewed distributions or unequal variances (Fagerland & Sandvik, 2009).

Cronbach's alpha internal consistency reliabilities were calculated for both the midterm and final examinations for descriptive purposes, though no formal statistical tests were conducted using these measures. Cronbach's alpha assesses the internal consistency of a test by measuring the extent to which all items within the test evaluate the same underlying concept or construct. This reliability measure reflects the inter-relatedness of test items and can be interpreted as the correlation of the test with itself, providing insight into whether the examination items collectively measure a unified domain of knowledge (Tavakol & Dennick, 2011).

In addition to within-cohort comparisons, correlations between midterm and final exam scores were evaluated using Pearson correlation coefficients. This analysis was selected to examine the degree of consistency in student performance across the two assessment points rather than to imply causal relationships (Liu et al., 2016). Where, a higher correlation indicates that students' performance on the midterm more strongly aligns with their final exam performance, suggesting a more stable measurement of student knowledge within the given examination format (closed-book or open-book).

Results

The statistical analyses were performed using IBM SPSS Statistics. Data presented in Table 1 reveals several noteworthy patterns when comparing open-book (2023) and closed-book (2024) examination formats. Students performed significantly better on the closed-book

https://doi.org/10.52202/086293-0001

midterm (58.1 ± 17) compared to the open-book midterm (38.0 ± 14), representing a substantial difference of 20.1 percentage points. This is underpinned with the previous studies which shown that students often perform better on closed-book tests because the absence of reference materials encourages deeper preparation, whereas open-book formats may foster less rigorous study habits (Boniface, 1985; Weber et al., 1983). Similarly, final exam scores following the closed-book midterm were higher (56.5 ± 0.12) than those following the open-book midterm (48.0 ± 0.01), suggesting a potential carryover effect from the assessment approach. While earlier research found that open-book tests may improve immediate performance but lead to similar delayed performance compared to closed-book tests (Agarwal et al., 2008), the present results suggest that closed-book formats may provide a stronger foundation for subsequent assessments.

Table 1: Midterm and final exam mean percent score for two-cohorts, 2023 and 2024

Assessment	2023 (open-book)			2024 (closed-book)		
	Midterm (mean ± SD)	Final score (mean ± SD)	Correlation (r)	Midterm (mean ± SD)	Final score (mean ± SD)	Correlation (r)
Course X	38.0 ±0.14	48.0 ±0.01	0.38	58.1 ±0.17	56.5 ±0.12	0.64

Given that the Shapiro-Wilk test indicated non-normal distributions for both midterm and final exam scores, with the closed-book cohort showing particularly pronounced non-normality ($W = 0.89$, $p = .03$), the Mann-Whitney U test was employed for all comparisons. Table 2 presents the Mann-Whitney U test results for both final exam scores and midterm marks. For final exam scores, the difference between cohorts was not statistically significant ($U = 1039.5$, $z = -1.59$, $p = .111$). This suggests that although the closed-book midterm cohort scored higher on average, the change in midterm format had limited statistical impact on subsequent final exam performance. Notably, the 2024 cohort displayed greater score dispersion, possibly reflecting varied adaptation strategies to the closed-book format. In contrast, midterm marks showed a statistically significant difference between cohorts ($U = 668.0$, $z = -4.12$, $p < .001$). This finding indicates that the closed-book midterm format may have promoted stronger student engagement and more consistent study habits throughout the semester. These results support existing literature linking closed-book assessments with deeper learning processes and sustained academic preparation.

Table 2: Mann-Whitney U Test results for final exam scores and midterm marks

Assessments	U	Z	p-value
Midterm	668.0	-4.119	<0.001
Final	1039.5	-1.594	0.111

In addition, correlation analysis is provided to show internal consistency of student performance across the two assessment points. The final exam is defined as the dependent variable and the midterm scores as the independent variable. The association between midterm and final exam scores was examined using Pearson correlation coefficient which measure the strength and direction of the monotonic relationship between the two assessments. The closed-book format demonstrated a moderately strong correlation ($r = 0.64$) compared to the weaker correlation in the open-book format ($r = 0.38$). Those stronger relationship suggests that the closed-book assessment may better predict final exam performance and potentially indicates more consistent measurement of student knowledge across both assessment points.

Reliability analysis further support the benefits of the closed-book format. The analysis was conducted using Cronbach's alpha, with each exam question treated as an item and student responses forming the basis of the analysis. Table 3 presents reliability metrics for midterm and final exams across two cohorts with different. The data reveals a consistent improvement in reliability measures when comparing the open-book to the closed-book format. For midterm examinations, both reliability indicators show substantial increases where Cronbach's alpha improved from 0.36 to 0.52. The 0.36 implies that 36% of the variance in the scores is reliable variance and 64% is error variance in 2023. While, in 2024, 0.52 implies that 52% of the variance in the scores is reliable variance and 48% is error variance. These reliability differences suggest that closed-book examinations may generally produce more consistent measurement of student knowledge (Lichtenstein & Fischhoff, 1977).

Table 3: Midterm and final exam reliabilities for two-cohorts, 2023 and 2024

Assessment	2023 (open-book)		2024 (closed-book)	
	Pearson correlation	Cronbach's alpha	Pearson correlation	Cronbach's alpha
Midterm	0.23	0.36	0.40	0.52
Final exam	-	0.64	-	0.57

Figure 1 shows the correlations between midterm and examination scores for both open- an closed- book formats in 2023 and 2024. In 2023, it can be observed that a moderate positive correlation with considerable scatter around the trend. In contrast, the right panel demonstrates the relationship between closed-book midterm scores and open-book final exam performance (2024 cohort), exhibiting a noticeably stronger correlation with less variability around the regression line.

The tighter clustering of data points in the closed-book midterm suggests that this format provides more consistent and predictable measurement of student achievement (Block, 2012). This enhanced predictive validity indicates that closed-book midterm performance serves as a more reliable indicator of subsequent academic success on the final examination. The reduced scatter in the closed-book format may reflect several underlying mechanisms: students develop more systematic study approaches when they cannot rely on external resources(Agarwal & Roediger III, 2011); subsequently the closed-book test stimulated a deep learning approach more than did open-book format (Heijne-Penninga et al., 2008).

Furthermore, the stronger correlation in the closed-book condition supports constructivist learning theory, which emphasizes the importance of deep learning through conceptual understanding of material rather than surface-level information processing (Vogel-Walcutt et al., 2011; Vygotsky & Cole, 1978). The enhanced measurement consistency observed in closed-book assessments suggests that this format may better capture students' internalized conceptual frameworks and their ability to apply knowledge independently, which aligns with constructivist principles of meaningful learning.

For open-book, the high variance in score possibly due to more competent students receive help from using an open book, while the less competent students use the book inefficiently waste time and find no answers (Lachman, 1963). This can be translated that the open-book format provides benefit to the smarter students in both midterm and final examination format.

https://doi.org/10.52202/086293-0001

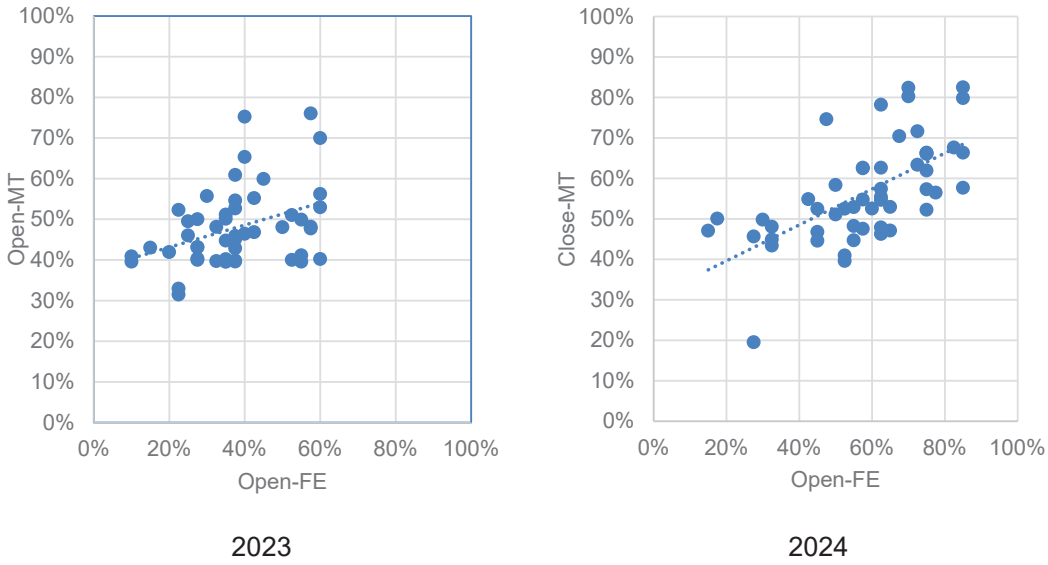


Figure 1: Correlations between open-book and closed-book midterm with the open-book final exam format in 2023 and 2024.

Aforementioned results hold key educational implications. Closed-book exams appear to encourage more thorough preparation and engagement, supporting long-term retention and transfer of knowledge (Johanns et al., 2017). On the contrary, in the open-book format, students might have over-relied on their materials, spending excessive time searching for information rather than demonstrating applied understanding (Rummer et al., 2019). The assessment design might also have played a role, as open-book exam questions may have been intentionally designed to be more challenging to compensate for material access. In the context of this course, the implementation of a closed-book exam may be particularly suitable for achieving learning outcomes with the aim to promote mastery of fundamental concepts or testing at the lower Bloom's Taxonomy level (Forehand, 2010). Conversely, open-book exams can still play a valuable role, particularly in assessing higher-order thinking skills such as synthesis, application, and critical evaluation (Johanns et al., 2017; Maity, 2017).

Conclusion

The sudden shift from open-book to closed-book examination format revealed a general impact on student marking and grading performance for this course. To avoid potential biases, learning outcomes and teaching approaches remained consistent across both cohorts throughout the assessment period. The data suggest that both examination formats retained their capacity to differentiate students based on their understanding of course material, though with varying degrees of measurement precision. However, this work is limited to quantitative assessment measures. Future research incorporating student feedback on the examination format transition would add significant value to the existing body of knowledge on assessment strategies in engineering education, providing a more comprehensive understanding of how format changes affect both learning outcomes and student experiences.

References

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., III, & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology: T*he *Official Journal of the Society for Applied Research in Memory and Cognition*, 22 (7), 861–876. https://doi.org/10.1002/acp.1400

Agarwal, P. K., & Roediger, H. L., III. (2011). Expectancy of an open-book test decreases performance on a delayed closed-book test. *Memory*, 19 (8), 836–852. https://doi.org/10.1080/09658211.2011.613840

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/10.1080/10511970.2010.523801

Boniface, D. (1985). Candidates' use of notes and textbooks during an open-book examination. *Educational Research*, 27 (3), 201–209. https://doi.org/10.1080/0013188850270305Cor,

Eilertsen, T. V., & Valdermo, O. (2000). Open-book assessment: A contribution to improved learning? *Studies in Educational Evaluation*, 26 (2), 91–103. https://doi.org/10.1016/S0191-491X(00)00010-9

Fagerland, M. W., & Sandvik, L. (2009). Performance of five two-sample location tests for skewed distributions with unequal variances. *Contemporary Clinical Trials*, 30 (5), 490–496. https://doi.org/10.1016/j.cct.2009.06.007

Forehand, M. (2010). Bloom's taxonomy. *Emerging Perspectives on Learning, Teaching, and Technology*, 41 (4), 47–56.

Gharib, A., Phillips, W., & Mathew, N. (2012). Cheat sheet or open-book? A comparison of the effects of exam types on performance, retention, and anxiety. *Online Submission*, 2 (8), 469–478.

Heijne-Penninga, M., Kuks, J. B., Hofman, W. A., & Cohen-Schotanus, J. (2008). Influence of open-and closed-book tests on medical students' learning approaches. *Medical Education*, 42 10), 967–974. https://doi.org/10.1111/j.1365-2923.2008.03125.x

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, 89–94. https://doi.org/10.1016/j.nepr.2017.08.018

Lachman, S. (1963). Objective achievement-test performance under open book and closed book conditions. *American Psychologist*, 18 (7), 450.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance*, 20 (2), 159–183. https://doi.org/10.1016/0030-5073(77)90001-0

Liu, J., Tang, W., Chen, G., Lu, Y., Feng, C., & Tu, X. M. (2016). Correlation and agreement: Overview and clarification of competing concepts and measures. *Shanghai Archives of Psychiatry*, 28 (2), 115–120. https://doi.org/10.11919/j.issn.1002-0829.216045

M. K., & Brocks, D. R. (2021). Does a sudden shift of testing format from closed-book to open-book change the characteristics of test scores on summative final exams? *Currents in Pharmacy Teaching and Learning*, 13 (9), 1174–1179. https://doi.org/10.1016/j.cptl.2021.06.039

Maity, S. (2017). *Effectiveness of open book examination in English for standard VIII in different environments*. [Doctoral dissertation, Maharaja Sayajirao University of Baroda].

Rummer, R., Schweppe, J., & Schwede, A. (2019). Open-book versus closed-book tests in university classes: A field experiment. *Frontiers in Psychology*, 10, Article 463. https://doi.org/10.3389/fpsyg.2019.00463

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 29–47). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.004

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Theophilides, C., & Koutselini, M. (2000). Study behavior in the closed-book and the open-book examination: A comparative analysis. *Educational Research and Evaluation*, 6(4), 379–393. https://doi.org/10.1076/edre.6.4.379.6933

Vogel-Walcutt, J. J., Gebrim, J., Bowers, C., Carper, T., & Nicholson, D. (2011). Cognitive load theory vs. constructivist approaches: Which best leads to efficient, deep learning? *Journal of Computer Assisted Learning*, 27 (2), 133–145. https://doi.org/10.1111/j.1365-2729.2010.00381.x

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press

Weber, L. J., McBee, J. K., & Krebs, J. E. (1983). Take home tests: An experimental study. *Research in Higher Education*, 18, 473–483. https://doi.org/10.1007/BF00974805

https://doi.org/10.52202/086293-0001

Williams, J. B., & Wong, A. (2009). The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book, open-web exams. *British Journal of Educational Technology*, 40(2), 227–236. https://doi.org/10.1111/j.1467-8535.2008.00929.x

Copyright Statement

Copyright © 2025 Norhuda Abdul Manaf, Tomoya Tsuji, Hafizah Farhah Saipan Saipol, Shahirah Abu Bakar, Aizul Nahar Harun, Peng Yen Liew: The authors assign to the Research in Engineering Education Netork (REEN) and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full, and this copyright statement is reproduced. The authors also grant a non-exclusive licence to REEN to publish this document in full on the World Wide Web (prime sites and mirrors), on portable media and in printed form within the REES and RCEE & RHEd 2025 proceedings. Any other usage is prohibited without the express permission of the authors.