Influence of open- and closed-book tests on medical students' learning approaches

Heijne-Penninga, Marjolein; Kuks, Jan B. M.; Hofman, W. H. Adriaan; Cohen-Schotanus, Janke
Published in:
Medical Education

DOI:
10.1111/j.1365-2923.2008.03125.x

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2008

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Influence of open- and closed-book tests on medical students’ learning approaches

Marjolein Heijne-Penninga,1 Jan B M Kuks,1 W H Adriaan Hofman2 & Janke Cohen-Schotanus3

CONTEXT Two learning approaches are consistently distinguished in the literature: deep and surface learning. The deep learning approach is considered preferable. Open-book tests are expected to stimulate deep learning and to offer a possible way of handling the substantial growth in medical knowledge. In this study we test the hypothesis that open-book tests stimulate deep learning more than closed-book tests.

METHODS Medical students in Years 2 (n = 423) and 3 (n = 306) participated in this study. They evaluated their preparation for open- and closed-book tests using the test for Deep Information Processing (DIP). This questionnaire consists of 24 items divided into three subscales: Critical Reading, Broaden One’s Context, and Structuring. A paired t-test was used to analyse the data.

RESULTS Both cohorts scored significantly higher when preparing for closed-book tests for the overall DIP score and on the Broaden One’s Context and Structuring scales. Year 3 students also scored significantly higher on the Critical Reading scale when preparing for closed-book tests. Gender differences were found: women used deeper learning approaches than men.

CONCLUSIONS Our hypothesis was not supported. In fact, the opposite was found: closed-book tests stimulated a deep learning approach more than open-book tests. Three possible explanations are: deep learning is particularly necessary for remembering and recalling knowledge; students feel more confident when preparing for closed-book tests, and students are more motivated to study for closed-book tests. The debate on the concept of deep learning in higher education should probably be renewed.

KEYWORDS educational measurement/methods; education, medical, undergraduate/methods; cohort studies; clinical medicine/education.

INTRODUCTION

The body of medical knowledge is expanding rapidly.1 It is neither possible nor even desirable for medical students to remember this growing volume of factual information without the support of different reference sources. Adding open-book tests to the assessment programme assists in handling this growth of knowledge in medical education.2 Additionally, it is assumed that open-book tests stimulate a deep learning approach,3–5 which is considered preferable to surface learning.6–9 However, the notion that open-book tests stimulate deep learning is mainly based on opinion and logical conjecture and not on empirical data. The aim of this study is to provide empirical evidence to support the notion that medical students use a deeper learning approach when preparing for open-book tests than they do for closed-book tests.

Learning approaches

Two kinds of learning approaches are consistently distinguished in the literature: deep and surface learning.6–8 Students who focus on rote learning with the intention of reproducing the learning material are using a surface approach.6,8 By contrast, deep learning can be described as focusing on understanding by applying and comparing ideas.6,8 A deep
learning approach is considered preferable because the intellectual processes valuable to medical students seem to be more closely linked to this approach. Different labels are used to describe or measure deep learning in the literature. The most important aspects of deep learning are the intention to form a personal understanding of the learning material and an active engagement and interest in the subject being learned. In summary, deep learning seems to be characterised by three dimensions:

1 understanding: trying to understand the learning material by gaining an overview and creating outlines and structure;
2 elaboration: relating the learning material to other sources and personal ideas, and questioning and using evidence critically, and
3 analysis: trying to clarify the learning material by searching for its major points, finding reasons for what is being said and arriving at conclusions.

Research has shown that learning approaches are highly sensitive to the learning context, including the assessment programme. Students’ perceptions of the learning environment influence their learning approaches, for example, the way they perceive quality of education. It has been shown that clear learning objectives, structure and the stimulation of independent learning have a positive relationship with the deep learning approach and student perceptions of the relevance of a subject. Poor assessment methods and a heavy workload correlate with a surface learning approach. The impact on student learning seems to be truly complex and can vary between individuals and groups of students.

Adding open-book tests to the assessment programme modifies the learning context and may subsequently lead to a change in students’ learning approaches. As open-book tests stimulate teachers to ask questions at higher cognitive levels, students may be encouraged to use deeper learning approaches, especially when the need for recall is limited. When studying for open-book tests, students are able to read and think rather than read and memorise. Therefore, it is expected that open-book tests stimulate deep learning.

However, we found only one study that confirmed this expectation. In this study students were asked about their activities before and during open- and closed-book tests. According to the results, students preparing for open-book tests tended to apply higher-order thinking and studied the course material in greater depth. When preparing for closed-book tests, however, these students postponed their study activities and memorised information. The participants in this study were undergraduates whose assessment programmes did not regularly contain open-book tests. The results are difficult to generalise to a medical setting, especially in conditions where open-book tests are a regular part of the assessment programme.

In summary, open-book tests can be of use in handling the expanding body of knowledge. However, their influence on learning approaches has rarely been studied. This study examines the impact of preparing for open- and closed-book tests on the depth of learning in a medical curriculum. The hypothesis tested in this study is that open-book tests stimulate deep learning more than closed-book tests.

METHODS

Context and participants

At the University of Groningen, every written knowledge examination in the Bachelor of Medicine
programme concerns theory delivered in a 10-week module. Each module is examined in three sessions. Until 2003 it was common to test students using closed-book tests only. Since 2003 all written examinations have consisted of an open-book component that covers the entire amount of knowledge to be studied preceded by a closed-book component concerning essential facts to be learned by heart (which represent only a limited part of the material to be studied).

To avoid confusion about the distinction between these essential facts and the entirety of the theoretical knowledge set for the module, the theory of a module is divided into core and back-up knowledge. Core knowledge is defined as knowledge that every medical professional should know without needing to consult reference sources. Back-up knowledge is defined as knowledge that students need to understand and apply properly, with the use of reference sources, if desired. Students are instructed to study the entirety of the literature set for the module (in order to be able to work with it efficiently during the open-book examination) and then to see whether they are able to reproduce the core knowledge (in order to be able to reproduce this by heart during the closed-book examination).

The first part of each written examination comprises closed-book questions. Once the students’ closed-book answer sheets are collected, reference books can be opened for the second half of the examination, which contains open-book questions. The examination lasts a maximum of 3 hours, with 1 hour available for the closed-book questions. On average, examinations contain 48 open-book questions and 80 closed-book questions. All examinations are multiple-choice in format and are taken in class. During a previous research study the reliability and difficulty of these types of tests were examined using generalisability theory. The results of that study showed that the reliability of open- and closed-book tests varied between 0.71 and 0.85. Open-book test reliabilities were slightly lower than closed-book test reliabilities, but still sufficient. The level of difficulty, defined as an average percentage of correct answers, did not differ between the two types of tests.

Medical students in Years 2 (n = 423) and 3 (n = 306) participated in this study. They had been exposed to open- and closed-book questions from their first year of medical training. Students were informed that they were participating in a research study, although the hypothesis tested in this study was not mentioned.
Critical Reading (understanding); Broaden One’s Context (elaboration), and Structuring (analysis). The Critical Reading scale consists of nine items (e.g. ‘When reading a text in preparation for this open/closed-book examination, I understand the meaning of the text very quickly’). The Broaden One’s Context scale consists of eight items (e.g. ‘When reading a text in preparation for this open/closed-book examination, I compare what I read with things I already know’). The last scale, Structuring, consists of six items (e.g. ‘When reading a text in preparation for this open/closed-book examination, I make notes of the most important issues’). All items were rated on a 5-point Likert scale (1 = never, 5 = always). A total DIP score and a score for each scale were calculated.

Procedure

The questionnaires were presented to the students before the start of the examination. They were given extra time to complete the two questionnaires. To avoid order influences, half the students from each year-group started with the questionnaire about preparing for open-book questions, and the other half started with the questionnaire about preparing for closed-book questions.

Analysis

The internal consistencies of the DIP and the three scales of the DIP were calculated to indicate reliability. A paired t-test was used to distinguish differences between open- and closed-book test preparation. Differences between Year 2 and 3 students and between men and women were also analysed using the independent-samples t-test. To indicate the importance of the observed differences, effect sizes (ES) were calculated using the formula of Rosnow and Rosenthal.21 The following interpretation was applied: ES = 0.10 (small effect); ES = 0.30 (medium effect), and ES = 0.50 (large effect).22

RESULTS

Questionnaires were returned by 405 (96%) Year 2 students and 271 (89%) Year 3 students. About two-thirds of the respondents from both years were female. The average age of students in Years 2 and 3 was 20.5 years and 21.3 years, respectively. Some respondents did not complete both questionnaires or skipped some of the items. To handle the missing data, paired deletion was employed. Therefore, n differs between the scales in Tables 2 and 3.

The internal reliability of the DIP scales and the total DIP score per measurement varied from \( \alpha = 0.68 \) to \( \alpha = 0.84 \) (Table 1), indicating generally sufficient measurement precision.

Table 4 shows the DIP results for preparation for both test formats. Students scored significantly higher on all scales and on the overall DIP score when reporting on their preparation for closed-book questions. Only Critical Reading did not differ significantly in Year 2 and had a low effect size in Year 3.

Year 3 students used a higher level of deep learning than Year 2 students. These differences were significant for the overall DIP score, for all the closed-book...
test preparation scales, and for Critical Reading in open-book test preparation (ES = low).

The results show significant gender differences (Table 5). Female students scored higher on Structuring than males, regardless of the assessment type. Year 2 women also scored significantly higher than Year 2 men on the total DIP score when preparing for open-book questions. Year 3 women scored significantly higher than Year 3 men on the overall DIP score when preparing for closed-book questions. Both the effect sizes are low.

**DISCUSSION**

The hypothesis that open-book tests stimulate a deep learning approach more than closed-book tests was not confirmed. On the contrary, the opposite was found: closed-book tests are more strongly related to a deep learning approach than open-book tests.

A first possible explanation is that students preferentially apply a deep learning approach to structuring their learning material in such a way that they can recall it faster, which is especially necessary for closed-book tests. They probably do not realise that structuring and recalling are also important for open-book tests that require answers to be given to a sufficient number of questions in a limited period of time. Secondly, students’ perceptions of their ability to fulfil a task has a positive relationship with deep learning. Perhaps students feel more confident when preparing for closed-book tests than they do for open-book tests. Preparing for open-book tests and using references during the tests seem to be difficult skills. Thirdly, as well as confidence, motivation has a positive influence on the level of deep learning. Students are probably more motivated to study for closed-book tests, which is connected to the levels of priority students accord to preparing for both assessment types and their perceived relevance. They possibly consider closed-book tests to be more important and thus accord less priority to preparation for open-book tests because they know they can

### Table 3 Internal reliability (α) for the Deep Information Processing factors per measurement*

<table>
<thead>
<tr>
<th>Year</th>
<th>Critical Reading</th>
<th>Broaden One's Context</th>
<th>Structuring</th>
<th>Total DIP score</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 CB</td>
<td>0.80</td>
<td>0.83</td>
<td>0.73</td>
<td>0.84</td>
</tr>
<tr>
<td>2 OB</td>
<td>0.77</td>
<td>0.83</td>
<td>0.68</td>
<td>0.83</td>
</tr>
<tr>
<td>3 CB</td>
<td>0.77</td>
<td>0.78</td>
<td>0.71</td>
<td>0.81</td>
</tr>
<tr>
<td>3 OB</td>
<td>0.73</td>
<td>0.82</td>
<td>0.72</td>
<td>0.81</td>
</tr>
</tbody>
</table>

* CB = closed-book; OB = open-book; DIP = Deep Information Processing

### Table 4 Paired t-test, closed-book versus open-book test preparation and effect size for Year 2 and 3 students

<table>
<thead>
<tr>
<th></th>
<th>Closed-book</th>
<th></th>
<th>Open-book</th>
<th></th>
<th>T</th>
<th>P</th>
<th><strong>ES</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean SE</td>
<td>Mean SE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 2</td>
<td>Critical Reading</td>
<td>377 29.8 0.25</td>
<td>30.0 0.25</td>
<td>0.885 0.377 –</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Broaden One’s Context</td>
<td>372 25.8 0.25</td>
<td>24.8 0.25</td>
<td>5.255 0.000 0.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Structuring</td>
<td>378 23.5 0.23</td>
<td>22.1 0.22</td>
<td>7.722 0.000 0.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total DIP score</td>
<td>338 79.1 0.57</td>
<td>76.7 0.57</td>
<td>5.378 0.000 0.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 3</td>
<td>Critical Reading</td>
<td>251 31.8 0.28</td>
<td>31.2 0.28</td>
<td>2.820 0.005 0.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Broaden One’s Context</td>
<td>251 26.6 0.28</td>
<td>25.4 0.30</td>
<td>5.407 0.000 0.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Structuring</td>
<td>255 24.2 0.29</td>
<td>22.3 0.30</td>
<td>7.504 0.000 0.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total DIP score</td>
<td>232 82.5 0.64</td>
<td>78.8 0.65</td>
<td>6.444 0.000 0.39</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Effect size: low = 0.10; medium = 0.30, and large = 0.50. SE = standard error; ES = effect size; DIP = Deep Information Processing

always consult their references if necessary. Moreover, in the present context, results for open-book questions can be compensated for by results for closed-book questions, and the tests contained more closed- than open-book questions.

In the light of these possible explanations, it seems that if open-book tests are to stimulate deep learning, special attention to student confidence and motivation is required. A solution might be to train students in preparing for open-book tests and using references effectively. Equalising the numbers of questions in both tests may represent a second course of possible action. However, it should be realised that answering an open-book test question probably takes more time than answering a closed-book question. A third possibility concerns using open-book questions exclusively instead of a combination of open- and closed-book questions. Open-book test preparation would thus have sole priority and would not have to compete with closed-book test preparation. A fourth solution involves shortening the time available for sitting an open-book test or increasing the number of open-book questions in a test. This could stimulate students to use a deep learning approach because more efficiency in reference use would be required, which concords with a deeper learning strategy.

The results also show a gender difference: women use deeper learning approaches than men. Most studies of learning approaches do not report gender differences. The results of studies that have examined gender differences are either inconsistent or the differences are hard to explain. In higher education, women use a deeper learning approach than men, which is consistent with our results. Female students are often reported as studying in a more organised manner and as better able to manage their studies effectively. This could explain why they score especially highly on Structuring for open- and closed-book tests. Structuring was measured using items such

<p>| Table 5 Paired t-test, men versus women for closed-book and open-book, DIP scores and effect size for Year 2 and 3 students |
|--------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|</p>
<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th></th>
<th>Women</th>
<th></th>
<th>t</th>
<th>p</th>
<th>ES*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 2 (CB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Critical Reading</td>
<td>121</td>
<td>30.1</td>
<td>265</td>
<td>29.6</td>
<td>0.861</td>
<td>0.390</td>
<td>–</td>
</tr>
<tr>
<td>Broaden One’s Context</td>
<td>124</td>
<td>25.8</td>
<td>264</td>
<td>25.8</td>
<td>–0.019</td>
<td>0.985</td>
<td>–</td>
</tr>
<tr>
<td>Structuring</td>
<td>123</td>
<td>21.4</td>
<td>267</td>
<td>24.3</td>
<td>–6.093</td>
<td>0.000</td>
<td>0.30</td>
</tr>
<tr>
<td>Total DIP score</td>
<td>118</td>
<td>77.5</td>
<td>250</td>
<td>79.7</td>
<td>–1.899</td>
<td>0.058</td>
<td>–</td>
</tr>
<tr>
<td>Year 2 (OB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Critical Reading</td>
<td>121</td>
<td>30.2</td>
<td>268</td>
<td>29.9</td>
<td>0.471</td>
<td>0.638</td>
<td>–</td>
</tr>
<tr>
<td>Broaden One’s Context</td>
<td>120</td>
<td>24.5</td>
<td>263</td>
<td>24.9</td>
<td>–0.727</td>
<td>0.468</td>
<td>–</td>
</tr>
<tr>
<td>Structuring</td>
<td>119</td>
<td>20.7</td>
<td>267</td>
<td>22.9</td>
<td>–4.754</td>
<td>0.000</td>
<td>0.24</td>
</tr>
<tr>
<td>Total DIP score</td>
<td>114</td>
<td>74.8</td>
<td>250</td>
<td>77.7</td>
<td>–2.491</td>
<td>0.013</td>
<td>0.13</td>
</tr>
<tr>
<td>Year 3 (CB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Critical Reading</td>
<td>75</td>
<td>31.6</td>
<td>182</td>
<td>31.7</td>
<td>–0.205</td>
<td>0.838</td>
<td>–</td>
</tr>
<tr>
<td>Broaden One’s Context</td>
<td>76</td>
<td>26.3</td>
<td>183</td>
<td>26.8</td>
<td>–0.869</td>
<td>0.386</td>
<td>–</td>
</tr>
<tr>
<td>Structuring</td>
<td>78</td>
<td>22.0</td>
<td>185</td>
<td>25.2</td>
<td>–5.299</td>
<td>0.000</td>
<td>0.31</td>
</tr>
<tr>
<td>Total DIP score</td>
<td>72</td>
<td>80.1</td>
<td>176</td>
<td>83.4</td>
<td>–2.462</td>
<td>0.015</td>
<td>0.16</td>
</tr>
<tr>
<td>Year 3 (OB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Critical Reading</td>
<td>76</td>
<td>31.4</td>
<td>181</td>
<td>31.0</td>
<td>0.638</td>
<td>0.525</td>
<td>–</td>
</tr>
<tr>
<td>Broaden One’s Context</td>
<td>76</td>
<td>25.2</td>
<td>178</td>
<td>25.3</td>
<td>–0.126</td>
<td>0.900</td>
<td>–</td>
</tr>
<tr>
<td>Structuring</td>
<td>73</td>
<td>20.7</td>
<td>183</td>
<td>23.0</td>
<td>–3.494</td>
<td>0.001</td>
<td>0.21</td>
</tr>
<tr>
<td>Total DIP score</td>
<td>71</td>
<td>77.7</td>
<td>171</td>
<td>79.0</td>
<td>–0.937</td>
<td>0.350</td>
<td>–</td>
</tr>
</tbody>
</table>

*Effect size: low = 0.10; medium = 0.30, and large = 0.50
CB = closed-book; OB = open-book; DIP = Deep Information Processing; ES = effect size

as ‘I try to find structure in a text by looking at the title and headlines’, ‘I make notes on the most important issues’ and ‘I write down my conclusions on a text’. However, the differences may also result from men and women answering these questions differently. Byrne et al.²⁰ suggest that men fail to report their actual approaches to learning effectively.

An interesting third outcome is that Year 3 students scored significantly higher on deep learning than Year 2 students. Although the effect sizes were low, these results confirm other outcomes.¹⁴,³⁰ McCune and Entwistle indicated that Year 1 students were less able to read learning material critically and develop personal ideas than more experienced students. They were less able to move beyond a basic understanding of the learning material.¹⁵ These differences can probably also be found between students in Years 2 and 3. Furthermore, Year 3 students have more experience than Year 2 students in preparing for open- and closed-book questions. Their confidence is possibly greater and, therefore, also their level of deep learning.²⁴ It is possible students develop their learning over time and increasingly learn to use a deep learning approach. An equally possible explanation is that Year 3 students are more motivated to study because they are closer to clerkships and to working as medical practitioners. As previously discussed, motivation has a positive relationship with the level of deep learning achieved.²⁴

Strong points of this study include the fact that the hypothesis was tested by using a validated and sufficiently reliable questionnaire. Furthermore, two large groups of students participated in this study, the response rate was high and a crossover design was used to control for possible order influences. A possible weak point is that students were concurrently preparing for open- and closed-book tests. This could have biased their responses to the questionnaires; however, the systematic differences oppose this hypothesis. A second possible weak point is that this study was performed at one university and within one faculty, although two cohorts of students were used. Replication of this study in other disciplines may be necessary. Thirdly, a questionnaire from 1996 was used. However, the way deep learning is referred to in this questionnaire is comparable with its definition in the current literature.

In this study, deep learning was described as the act of attempting to understand the learning material throughout and to link the learning material to other knowledge and personal experiences.¹² However, research has shown that students in different disciplines displayed differences in deep learning. For science students, a deep approach appeared to require an initial concentration on details. For humanities students, the deep approach involved working from the outset to develop a personal understanding of the material.¹³ Perhaps the definition of deep learning is not universal for students in different fields of study. Moreover, the distinction between deep and surface learning was made some time in the 1970s. As noted in the introduction, context plays an important role in influencing student learning approaches, and today’s learning context differs substantially from that of the 1970s. Nowadays, the body of knowledge is expanding rapidly and knowledge is easily accessible. The casual observer is inclined to wonder whether approaches such as those that involve selecting knowledge and finding relevant knowledge at a specific moment are becoming increasingly important.

This transformed context also leads to different kinds of students. Present-day students have grown up with technology – they are members of the so-called Net-generation.³¹ The characteristics and qualities of this generation seem to differ from those of previous generations. For example, the Net-generation is accustomed to multi-tasking: they surf the Internet while chatting with friends, listening to the radio and composing a text. Students who have grown up with computers deal with information differently compared with those of the past.³¹ Perhaps the traditional definition of deep learning is no longer appropriate or complete for today’s students.

In summary, students following today’s medical curricula have to cope with a growing volume of knowledge in a context that differs substantially from the learning environment of the 1970s. Open-book tests enable students and medical curricula to handle this growing volume of knowledge. Although we did not find that open-book tests stimulate deep learning in medical students more than closed-book tests, they may stimulate other important approaches and skills and could nonetheless result in the outcome level that society currently requires.

Contributors: MHP contributed to the conception and design of the study, and the collection, analysis and interpretation of data. She also drafted the manuscript and revised it in concordance with suggestions from the other authors. JBMK contributed to the conception and design of the study, and the collection and interpretation of data, and commented on several drafts of the manuscript. WHAH and JCS both contributed to the conception and design of
the study and the interpretation of the data and commented on several drafts of the manuscript. All authors approved the final version of the article.

Acknowledgements: the authors thank the students who participated in this study, as well as J Bouwkamp-Timmer for her support with the literature search and her critical and constructive comments on the manuscript.

Funding: none.

Conflicts of interest: none.

Ethical approval: this study was approved by the Faculty Examination Board, University of Groningen, Groningen, The Netherlands.

REFERENCES


Received 19 September 2007; editorial comments to authors 5 December 2007; accepted for publication 27 March 2008.