Comparing Open-Book and Closed-Book Examinations: A Systematic Review

Steven J. Durning, MD, PhD, Ting Dong, PhD, Temple Ratcliffe, MD, Lambert Schuwirth, MD, PhD, Anthony R. Artino Jr, PhD, John R. Boulet, PhD, and Kevin Eva, PhD

Abstract

Purpose
To compare the relative utility of open-book examinations (OBEs) and closed-book examinations (CBEs) given the rapid expansion and accessibility of knowledge.

Method
A systematic review of peer-reviewed articles retrieved from MEDLINE, ERIC, Embase, and PsycINFO (through June 2013). In 2013–2014, articles that met inclusion criteria were reviewed by at least two investigators and coded for six outcome categories: (1) examination preparation, (2) test anxiety, (3) exam performance, (4) psychometrics and logistics, (5) testing effects, and (6) public perception.

Results
From 4,192 identified studies, 37 were included. The level of learner and subject studied varied. The frequency of each outcome category was as follows: (1) exam preparation (n = 20; 54%); (2) test anxiety (n = 14; 38%); (3) exam performance (n = 30; 81%); (4) psychometrics and logistics (n = 5; 14%); (5) testing effects (n = 24; 65%); and (6) public perception (n = 5; 14%). Preexamination outcome findings were equivocal, but students may prepare more extensively for CBEs. For during-examination outcomes, examinees appear to take longer to complete OBEs. Studies addressing examination performance favored CBE, particularly when preparation for CBE was greater than for OBE. Postexamination outcomes suggest little difference in testing effects or public perception.

Conclusions
Given the data available, there does not appear to be sufficient evidence for exclusively using CBE or OBE. As such, a combined approach could become a more significant part of testing protocols as licensing bodies seek ways to assess competencies other than the maintenance of medical knowledge.

Downloaded from https://journals.lww.com/academicmedicine by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3n4uRkX3k2dqrk+49qfzPWHbXDL36covdKeHaeSzaE= on 07/15/2020

Today's health care professions students and trainees have access to an unprecedented amount of information thanks to the rapid expansion of knowledge and the emergence of information technology. This easy access to information raises fundamental questions about the adequacy of closed-book examination (CBE) practices commonly used by the health professions. Some scholars argue that any examination of relevance must assess the examinee's ability to find, understand, evaluate, and use external resources. Such proponents of the open-book examination (OBE) argue that OBEs are more authentic to real-world practice and that success is not about "rote memorization."[1–3]

Please see the end of this article for information about the authors.

Correspondence should be addressed to Steven J. Durning, Department of Medicine, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Rd., Bethesda, MD 20814; telephone: (301) 295-3603; e-mail: steven.durning@usuhs.edu.

Acad Med. 2016;91:583–599.
First published online November 2, 2015
doi: 10.1097/ACM.0000000000000977

Supplemental digital content for this article is available at http://links.lww.com/ACADMED/A310.

Because professionals of the future will not be able to "know" all the information needed for competent performance,[4] meaningful assessment of medical practice, the argument goes, should allow individuals to look up information when taking an exam.

Scholars defending CBEs cite literature that has consistently found expert performance to be closely tied to rich, well-organized content knowledge of a subject. For example, studies have found that high performance on CBEs is associated with better practice outcomes.[5,6] In many situations a physician's ability to look up unknown information is restricted by time constraints and Internet access, and well-organized, content-specific knowledge remains paramount for expert performance. Merely putting more information at a physician's fingertips is, therefore, not likely to result in improved care because the physician needs knowledge to guide his or her search and to integrate new information with previous experience. Thus, reliance on information technology could detrimentally increase cognitive load (i.e., mental effort), decrease learning and critical appraisal of information, and ultimately harm patient care.[7]

Views on what defines a competent health care professional are changing. Where formerly the focus lay almost entirely on the possession of knowledge, currently physicians are expected to effectively use external point-of-care knowledge. For modern assessment to be aligned with this changing notion of competence, educators require better understanding of the various pros and cons of OBE and CBE assessment approaches. This is true both in terms of promoting assessment-for-learning and in contexts such as credentialing and licensing assessment.

To inform this issue, which affects the examination of physicians across the continuum of their careers, we conducted a systematic review of the literature comparing the two assessment strategies. Our questions were (1) What is the evidence regarding the comparative effectiveness of OBEs and CBEs? and (2) How might these findings inform current examination practices and future research in health professional education? We broadly defined OBEs as tests or assessments that allow the use of any resource such as the Internet, a textbook, course notes, or

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

journals, and we searched for studies in all educational fields.

Method

Scoping search

We were aware of no prior systematic reviews on the topic, so two authors (S.J.D. and T.D.) conducted a scoping search to better understand the breadth and depth of the relevant literature. This initial search of MEDLINE and ERIC (Education Research Information Center) was conducted in the spring of 2013. A third investigator (a research librarian) conducted a separate scoping search using the same data sources. The scoping search identified 488 articles. We excluded articles if they were deemed to be unrelated to our review, available only in abstract form, not available in English, or representing textbooks; this resulted in 78 citations that were discussed and underwent further review. During this further review, we iteratively generated themes that could be used as preliminary outcome categories for the systematic review and also used this step to further refine our inclusion and exclusion criteria and our search strategy and terms (Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A310).

Systematic review

We followed PRISMA Guidelines[8] and guidelines provided in the medical education literature.[9] We limited our search to full-length, published, peer-reviewed, English-language journal articles involving learners in either descriptive reports or educational interventions, using any study design related to our research questions. We further limited the papers reviewed to those that empirically compared (either directly or indirectly) OBEs and CBEs.

Relevant studies were identified by searching three databases during the summer of 2013 and included no date restrictions (i.e., we searched everything available through the date searched): (1) MEDLINE via Ovid (June 2013), (2) Embase via Ovid (July 2013), and (3) ERIC (June 2013). To identify additional studies, we searched the bibliographies of articles found by our electronic search, contacted experts in the field, and conducted a Web search using Google Scholar and PsycINFO. Supplemental Digital Appendix 1, http://links.lww.com/ACADMED/A310, displays the terms used for the systematic search.

We used a data collection form (Supplemental Digital Appendix 2, http://links.lww.com/ACADMED/A310) to rate each article. This form was constructed based on the findings of our scoping review and refined through conference calls among the authors. The form was pilot tested and revised by having each member of the investigative team use the form to review two articles. We discussed additional articles until consensus on the form was achieved.

Three authors (S.J.D., T.D., T.R.) independently reviewed the titles and abstracts of the retrieved publications. Each was initially categorized as *include*, *exclude*, or *uncertain*. All *include* and *uncertain* titles and abstracts were reviewed in the subsequent stage (i.e., review of the full-text version of the papers; see Figure 1). Authors disagreed regarding inclusion for 44 of the 4,192 titles and abstracts (see Figure 1), all of which were subsequently included in the full paper review. After the full paper review, 299 articles remained. The same three study authors then reviewed the full text of all 299 articles (see Figure 1) using the same categorization framework (*include*, *exclude*, *uncertain*). In doing so, 193 were deemed beyond the scope of this review. The remaining 106 full-text papers underwent a more detailed review and coding by the larger study team with each paper having at least two reviewers. Sixty-nine articles were excluded following this additional round of review, which included a series of conference calls and detailed coding using the data extraction form. Ultimately, 37 papers were included in our review.

We structured the outcome categories according to the themes that were generated from our scoping review. We report them here in the sequence in which they would occur in the testing process: (1) examination preparation, (2) test anxiety, (3) exam performance, (4) psychometrics and logistics, (5) testing effects, and (6) public perception. Any article could have multiple outcomes and was reviewed for relevant themes by two of the study authors. Following review and coding, conference calls were held among all coders until complete agreement was achieved for the coding of every article. A third coder was needed to resolve conflicts for 3 of the 37 papers.

The quality of each manuscript was examined by addressing the extent to which the research found was fit for purpose. This was done by having each reviewer code the manuscript for the presence of explicit research questions, hypotheses, conceptual and/or theoretical frameworks, and by recording additional quality judgments. Reviewers used a five-point rating scale (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree) to assess four domains: trustworthiness of findings, study rigor, implementation of study findings, and appropriateness of data analysis. These latter judgments were made in relation to the degree to which each study effectively addressed a research question comparing the relative benefits of OBEs versus CBEs.

Results

Our search identified 4,192 articles, 37 of which were included in our review[1–3,10–43] (see Figure 1 and Appendixes 1 and 2). The frequency with which outcomes were identified was as follows: (1) exam preparation (n = 20; 54%); (2) test anxiety (n = 14; 38%); (3) exam performance (n = 30; 81%); (4) psychometrics and logistics (n = 5; 14%); (5) testing effects (n = 13; 35%); and (6) public perception (n = 5; 14%).

Study quality

Overall, the quality of the articles included in our review was deemed to be adequate for our purpose. Explicit research questions were presented in 31 articles (84%), hypotheses were stated in 14 (38%), and hypotheses were justified in 10 (27%). Conceptual and/or theoretical frameworks were described in 7 articles (19%).

Study context

Thirty-four investigations (92%) were single-institution studies. Nearly half were performed in the United States (n = 18; 49%). Other locations included the Netherlands (n = 5; 14%), the United Kingdom (n = 4; 11%), Greece (n = 3; 8%), and Australia (n = 2; 5%), and 1 study (3%) was included from each of the following locations: Canada, Denmark, Norway, Africa, and Israel.

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

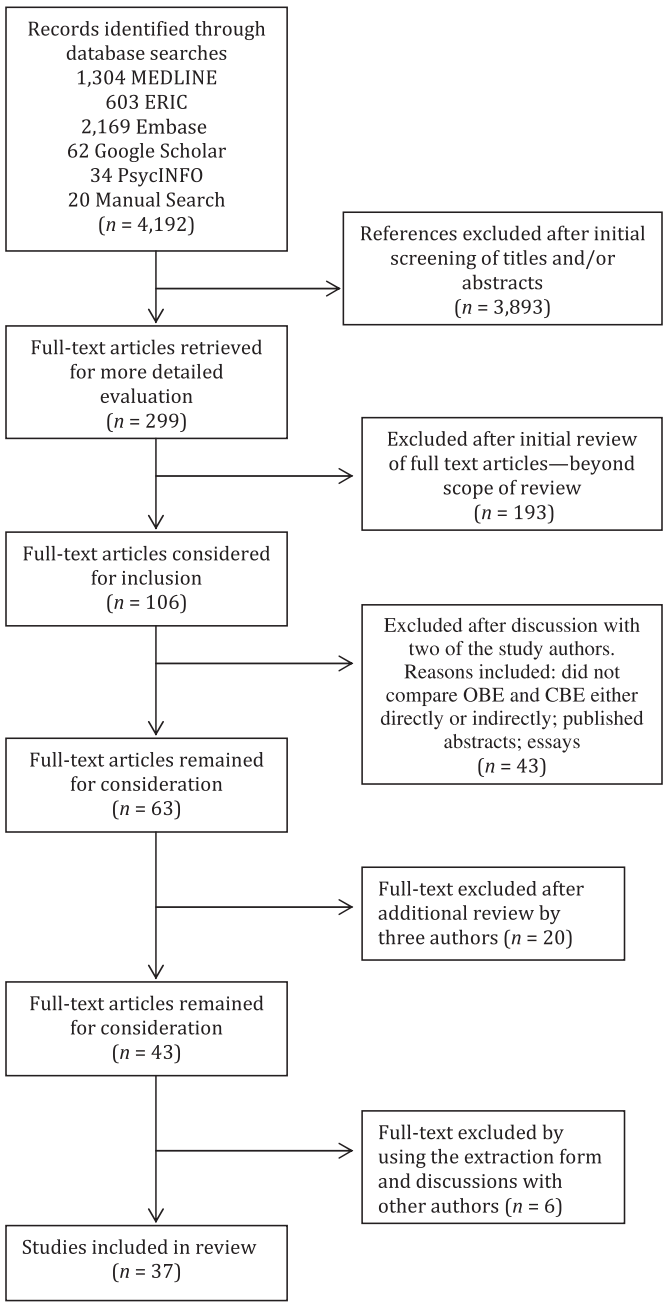

Figure 1 Flowchart of article selection for a systematic review comparing open- and closed-book examinations. The review was conducted in 2013–2014 and included all literature published as of the search dates. Abbreviations: ERIC indicates Education Resources Information Center; OBE, open-book examination; CBE, closed-book examination.

The majority of studies pertained to college-level students (n = 24; 65%); 2 studies investigated high school students (5%); 8 investigated medical students (22%; 2 of these were multi-institutional); 2 investigated other postcollege instructional settings (5%); and 1 study (3%) included practicing physicians. For the majority, the stakes of the examination were rated as medium (n = 21; 57%) in that the assessments were generally end-of-course examinations. Two (5%) were considered high-stakes, being equivalent to national licensing examinations. Few studies included a formal incentive (e.g., extra credit or a small payment) (n = 6; 16%) to participants beyond earning a course grade.

Few studies reported enrolling participants with significant prior experience with OBEs (n = 7; 19%) or some experience with OBEs (n = 4; 11%); most articles either reported that participants had no prior experience or did not mention prior experience (n = 26; 70%). Because the findings did not appear to differ according to type of learner (e.g., high school, undergraduate, or practicing physicians), we describe the findings in each theme as a whole, unless otherwise stated. Appendix 1 provides detailed results for each paper by theme. Some papers included more than one theme.

Exam preparation

Exam format could potentially influence test preparation (and, hence, learning). Some argue that CBEs promote superficial learning by requiring students to memorize large amounts of material, whereas OBEs focus learners on the application of what they have learned. Others argue that CBEs, compared with OBEs, prompt students to study more because they will not be able to look things up during the exam.

In terms of preparation time, findings were inconsistent across studies, but in sum appear to favor CBEs. Some showed that students reported more preparation time for CBEs than OBEs[10–12] (Appendix 1) or attended class less often if the test was an OBE.[12] Others reported that students prepared for OBEs and CBEs similarly[13,14]; no studies reported more preparation time for OBEs than CBEs. Of note, an increase in preparation time could indicate insufficient prior engagement with the material rather than being a proxy for improved learning and performance.[15]

Reviewing the articles examining preparation strategy revealed that students did not change study tactics for OBEs versus CBEs,[16,17] and no correlation between test format and deep versus surface learning approaches was found.[17]

Thus, research exploring exam preparation was equivocal with respect to whether students prepare differently (or at greater length) for CBEs or OBEs. When differences did exist, they tended to show that participants studied more when they expected a CBE.

Test anxiety

Emotions affect cognitive performance.[44] Although negative emotions were once thought to have exclusively deleterious

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

effects on performance, contemporary theories of emotion suggest that such an assumption is overly simplistic.[45] For example, a negative emotion like anxiety might actually motivate a student to study for a CBE, which could result in superior performance when compared with an unstressed student preparing for an OBE. Regardless, reducing test anxiety is often reported to be a primary motivation for considering OBEs. Our findings indicate, however, that anxiety effects were typically examined as a secondary issue relative to a study's primary purpose (see Appendix 1), and all studies that assessed emotions lacked a theoretical grounding. In particular, of the 14 studies with emotion-related outcomes, none employed a theory of emotion to help frame the study or explain the findings.

Evidence suggests that students may overestimate the effect that OBEs or partial OBEs (i.e., exams in which students can bring some prepared material like a "cheat sheet" rather than having access to any desired material) have on reducing their anxiety. Several studies suggest that students associate OBEs with less anxiety,[16,27,28] but only a minority of students actually report lower anxiety.[24,28] For example, Baillie and Toohey[24] found that anxiety associated with taking OBEs instead of CBEs was not reduced as much as expected, with 45% of students reporting being just as stressed with OBEs as with CBEs. It has been suggested that certain aspects of OBEs, such as the belief that examiners will choose questions of greater difficulty, can be anxiety provoking for students.[19] It remains to be seen whether students overestimate the impact OBEs have on reducing their anxiety because they lack familiarity with the test format.

On balance, these findings suggest that students may overestimate the impact that OBEs have on reducing their anxiety and, by extension, on potentially improving their performance. Not only was the reporting of methods and analyses for examining anxiety effects incomplete, but these effects are often explored as an afterthought in extant studies, and they lacked theoretical grounding.

Exam performance

The most common outcome explored was examination performance, defined as comparing learners' achievement on OBEs versus achievement on CBEs (Appendix 1). Intuitively, one might expect that examinees would perform better on OBEs because they have the capacity to look up answers. Opponents suggest that the OBE format does not inherently lessen difficulty but, instead, frees the examiner to focus questions on the test taker's ability to apply knowledge (i.e., testing what cannot simply be "looked up"), and the time required to look up information can increase difficulty by creating pressures requiring learners to retrieve and communicate answers efficiently. Two caveats are noteworthy when considering exam performance as an outcome: (1) In most studies, students had little to no experience with OBEs—only one study[21] that addressed examination performance reported that students had prior OBE experience; and (2) exam performance is a challenging outcome to study because the difficulty of an exam depends on the questions asked, and some proponents of OBE argue that its main advantage is enabling instructors to pose questions with a different style or focus. Different questions across different examination formats may, therefore, be required to enable the advantages of OBEs to be recognized.

The majority of the examinations were MCQ format, but some were also essay and/or short answer (Appendix 1). Typically, no significant difference in examinee performance was found,[30,34,38] or performance was better on CBEs (Appendix 1). In investigations demonstrating better performance on CBEs, when reasons for this finding were explored, the authors generally suggested that the difference in performance related to examination preparation. Some studies did show better performance on OBEs immediately after learning, but even those differences did not persist over time (i.e., OBE and CBE performance were equivalent, or CBE performance was superior on a subsequent delayed test; Appendix 1).

An investigation by Block[25] is useful for understanding the relationship between test preparation and exam performance. In the first experiment, learners who were expecting a CBE performed 10% better on a subsequent test over those who were expecting an OBE. In a second experiment, which again demonstrated improved performance when learners expected CBEs, participants reported spending less time studying (i.e., less preparation) when they expected an OBE than when they expected a CBE. In a different study by Carrier,[18] students scored significantly lower when expecting an OBE than when expecting a CBE for their final examination. The author suggested that this finding may be due to examinees' deeper approach to learning (defined as studying lecture notes, making chapter notes, highlighting text, and coming to office hours—all activities that correlated with higher exam scores) when preparing for a CBE. In another study,[17] students commented that they were less prepared for a final examination that they knew would be an OBE because they expected to be able to find the answers in the book during the exam. To counter the notion that lower performance is due to examinees' inability to find material in a resource during an OBE, three studies reported that the preparation of OBE materials (e.g., note cards) was not sufficient to improve performance on a CBE.[23,25,26] Finally, in an investigation[25] comparing performance on OBE and CBE tests earlier in the term with performance on a CBE final examination, students in the experimental section scored lower on their final exam and recalled significantly less about topics that were covered on preceding OBEs than those covered by CBEs.

In sum, studies comparing exam performance appear to favor CBEs. However, the combination of relatively little experience with OBEs and the differences in exam preparation noted in several investigations highlighted in this section leave open the possibility that OBE performance could be improved through instructing students about OBEs or providing practice tests. On this point, three sets of authors indicated that students need to have the right expectation for what it takes to do well on OBEs.[19,21,24]

Psychometrics and logistics

Research has generally shown that the validity of a test is determined more by the content of the questions included than by the examination format.[46–48]

However, two studies directly examined the impact of the exam format on the psychometric utility of the assessment. One comparison was limited because test content and number of questions were

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

confounded with assessment format,[3] whereas the second study concluded that a suitably constructed set of questions could be used to discriminate student abilities in either an OBE or CBE environment.[32]

In practice, it may not be realistic to compare reliability across test formats while keeping the number of items constant. Three studies that compared CBEs with OBEs with respect to their influence on the time required to take the test found that students took 10% to 60% longer to complete OBEs.[10,30,32] Thus, if one controls for amount of testing time, it is likely that fewer questions would be asked in OBE format, and, hence, the reliability of the equivalent CBE-formatted exam can be anticipated to be higher.

Testing effects

Testing effects occur when taking an exam improves subsequent performance. Such benefits can arise in indirect ways (e.g., being prompted to study) or from direct effects of the material becoming more memorable when participants are tested on it than when they simply study for a test.[49] Most commonly, direct testing effects are demonstrated by separating research participants into two groups, one of which is asked to study material and then take an intervention test, while the other is asked only to study (multiple times to equate the time participants are exposed to the material across groups). The testing effect is demonstrated when the tested group outperforms the study group on a subsequent outcome exam. This testing effect (test-enhanced learning) has been well documented in multiple fields.[50]

Proponents of CBE argue that learning requires active construction of memory that is less likely to occur when one relies on external resources to answer test questions. OBE proponents argue that OBEs may enhance the ability to apply knowledge because rote memorization is not emphasized.

Both OBE and CBE demonstrate testing effects (Appendix 2). Four studies comparing OBEs and CBEs demonstrated testing effects that were roughly equivalent[10,13,31,37] (Appendix 2). The testing effect of CBEs was superior in one study.[12] These researchers demonstrated that during a summative CBE participants performed worse on material covered by an OBE intervention relative to a CBE intervention.[12] Testing effects were observed regardless of examination format. Consistent with prior studies, students' collective self-perceptions ran counter to the empirical finding that testing effects occur regardless of test format; students felt that studying alone was more effective preparation than taking either an OBE or CBE.[31]

Public perception

Public perception (i.e., different groups' opinions about OBEs and CBEs) was examined from the learner's and the teacher's perspective. Studies suggest that learners have a more positive perception of OBEs over CBEs.[2,17,19,22] On the other hand, students also commented that OBE questions were more difficult and that they desired additional practice or training for the OBE format.[17]

Teachers' views often challenged the implementation of OBEs.[17,23] Teachers expressed concerns over the increased resources associated with preparing OBEs, as well as the perceived additional time required for learners to take OBEs.[2,22]

Discussion

Overall, the empirical literature comparing OBEs and CBEs is fairly limited. Among the studies that do exist, there is a fair amount of diversity, both in terms of learner level and the subjects studied (see Appendix 1). Although it can be challenging to generalize these findings from diverse learner groups and academic subjects to the field of medicine, this diversity is potentially beneficial when attempting to gain a general understanding of the influence of exam format on learning outcomes.

The studies we reviewed were generally of adequate quality for the questions addressed, and we did not identify any systematic differences in the use of OBE versus CBE by the field studied (e.g., medical education versus education versus other) or level of content (e.g., graduate versus undergraduate student). Prior to the examination, findings were equivocal; if test format does affect outcomes, it favors the argument that people prepare more for CBEs. This may be driven by the finding that students anticipate lessened anxiety with OBEs even though this does not appear to translate to actual experiences of lessened anxiety. During the examination, examinees appear to take longer to complete OBEs, which could either influence the test's reliability, if testing time is kept constant, or influence the length of time that must be offered to candidates to complete an equally reliable exam. Studies addressing examination performance favored CBEs, particularly when learners reported spending more time preparing for CBEs than for OBEs. With respect to postexamination outcomes of CBEs and OBEs, we did not find robust evidence for differences in testing effects or public perception. That said, one might imagine concerned patients who wonder, "How can you be an expert if you need to look things up on the Internet?"[51]

The type of examination used might need to be based less on learning and performance outcomes and more on logistical limitations, as well as the desire to authentically represent what individuals do in practice. Given that we found evidence of the testing effect under both OBE and CBE conditions, and that participants' perceptions of testing effects run counter to empirical findings, a related question is how often an individual should be examined to maximize testing effects. A further exploration of contemporary learning theories might provide a useful lens for understanding and interpreting how environmental factors and personal factors interact in dynamic ways to influence examination performance[52] and the pedagogical value of testing.

It is challenging for high-stakes testing organizations that value test security to allow learners to have unrestricted access to the Internet during an exam.[53] At the same time, choosing a limited number of Web-based external resources erodes authenticity, could disadvantage examinees who are less familiar with the chosen tools, and potentially affects fairness if technical difficulties arise during an examination. Additional feasibility questions include the cost of allowing Web-based resource access and the additional time required to achieve the same reliability with OBE relative to CBE. Issues such as cost and fairness have not been addressed in prior investigations.

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

In terms of authenticity, the studies conducted to date have rarely looked at "high-stakes" assessment. Although there is good reason to argue that a physician's ability to find information is an important skill to maintain, there can be a perception that OBEs are easier than CBEs. Although studies are lacking, an excerpt from the American Board of Ophthalmology regarding changes to their recertification examination captures the sentiment of many:

> The decision to change from an open-book, take-home examination to a closed-book, computerized proctored examination was based primarily on the recognition of the value of the certificate within the public domain … state medical licensing boards are increasingly asking for a proctored examination.[54]

We believe this preference is indicative of the perception that OBEs are perhaps less rigorous and/or less valid than a proctored examination.

The findings of our review are subject to several limitations in the existing literature. Only a minority of studies reported that learners had significant prior experience with OBEs. Providing learner training and making OBEs more prevalent could greatly alter perceptions of OBEs. Second, very few of the investigations reviewed included electronic resources (e.g., the Internet) as a parameter for OBEs because most were conducted before the Internet was widely used. Third, few investigations have involved practicing physicians. Fourth, the majority of studies were conducted within a single institution, which limits their generalizability, and only a minority of papers included a conceptual and/or theoretical framework, which can make interpretation difficult.

As the volume of medical knowledge continues to expand rapidly, education and assessment will have to instill within trainees the motivation and learning strategies needed to become lifelong, self-regulated learners. We wish to point out that the outcomes used in the studies reviewed here did not capture elements deemed to be essential by the current assessment-for-learning discourse. For example, no study looked at whether the incorporation of CBEs or OBEs yielded differences in reflection-on-action or receptivity to feedback when examinees formulated learning goals or were presented with external data.

OBEs and CBEs can contribute to an assessment program in part because of their complementary pros and cons. OBEs should not be thought of as an alternative to CBEs, but their value may be in expanding beyond what is measured by CBEs. For example, exploring the "skill" of looking up information on the Internet seems unlikely to be accomplished through CBE. A strategy, therefore, could be coupling OBEs with CBEs to explore these different "skills" without compromising reliability. Furthermore, testing effects are not currently being optimized given the infrequency of examinations. A series of mandatory but ungraded OBEs might help to improve aspects of these processes, such as capitalizing on the testing effect without dramatically increasing learner anxiety. One examination each decade, as is practiced by many certifying bodies, is unlikely to maximize the educational impact of testing or induce habits of continuous professional development. Further, by including some OBE items, the opportunity for improving authenticity and reducing the stigma with the need to look things up could be leveraged. Any such benefits, however, may only be realized by recognizing the need identified by several authors that OBE training is necessary for both students and examiners. Expectations need to be established regarding the types of questions used, the need for preparation, and how much time examinees can use to search for information.

Conclusions

Given the data collected to date, there does not appear to be sufficient evidence for relying solely on OBE or CBE formats. Therefore, we believe that a combined approach could become a more significant part of testing programs, including physician certification or recertification.

Acknowledgments: The authors would like to thank Rhonda J. Allard, MILS, reference librarian at the Uniformed Services University of the Health Sciences, for her assistance with our search strategies.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

Disclaimer: The views expressed herein are those of the authors and do not necessarily reflect the views of the Uniformed Services University of the Health Sciences, the Department of Defense, or the U.S. government.

S.J. Durning is professor of medicine and pathology, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

T. Dong is assistant professor of medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

T. Ratcliffe is assistant professor of medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

L. Schuwirth is professor of medical education, Flinders University, Bedford Park, South Australia, Australia, and professor for innovative assessment, Maastricht University, Maastricht, the Netherlands.

A.R. Artino Jr is professor of medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

J.R. Boulet is vice president of research and evaluation, Foundation for Advancement of International Medical Education and Research, Philadelphia, Pennsylvania.

K. Eva is professor of medicine, University of British Columbia, Vancouver, British Columbia, Canada.

References

1 Feldhusen JF. An evaluation of college students' reactions to open book examinations. Educ Psychol Meas. 1961;XXI:637–646.
2 Theophilides C, Dionysiou O. The major functions of the open-book examination at the university level: A factor analytic study. Stud Educ Eval. 1996;22:157–170.
3 Heijne-Penninga M, Kuks JB, Hofman WH, Cohen-Schotanus J. Influence of open- and closed-book tests on medical students' learning approaches. Med Educ. 2008;42:967–974.
4 Adair JG, Vohra N. The explosion of knowledge, references, and citations. Psychology's unique response to a crisis. Am Psychol. 2003;58:15–23.
5 Ramsey PG, Carline JD, Inui TS, Larson EB, LoGerfo JP, Wenrich MD. Predictive validity of certification by the American Board of Internal Medicine. Ann Intern Med. 1989;110:719–726.
6 Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. JAMA. 2007;298:993–1001.
7 Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive load theory: Implications for medical education: AMEE guide no. 86. Med Teach. 2014;36:371–384.
8 Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Ann Intern Med. 2009;151:264–269, W64.
9 Cook DA, West CP. Conducting systematic reviews in medical education: A stepwise approach. Med Educ. 2012;46:943–952.
10 Agarwal PK, Roediger HL 3rd. Expectancy of an open-book test decreases performance on a delayed closed-book test. Memory. 2011;19:836–852.
11 Boniface D. Candidates' use of notes and textbooks during an open-book examination. Educ Res. 1985;27:201–209.
12 Moore R, Jensen PA. Do open-book exams impede long-term learning in introductory

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

biology course? J Coll Sci Teach. 2007; 36:46–49.
13 Gharib A, Phillips W, Mathew N. Cheat sheet or open-book? A comparison of the effects of exam types on performance, retentions, and anxiety. Psychol Res. 2012;2:469–478.
14 Betts LR, Elder TJ, Hartley J, Trueman M. Does correction for guessing reduce students' performance on multiple-choice examinations? Yes? No? Sometimes? Assess Eval High Educ. 2009;34:1–15.
15 Heijne-Penninga M, Kuks JB, Hofman WH, Cohen-Schotanus J. Influences of deep learning, need for cognition and preparation time on open- and closed-book test performance. Med Educ. 2010;44:884–891.
16 Broyles IL, Cyr PR, Korsen N. Open book tests: Assessment of academic learning in clerkships. Med Teach. 2005;27:456–462.
17 Dale VH, Wieland B, Pirkelbauer B, Nevel A. Value and benefits of open-book examinations as assessment for deep learning in a post-graduate animal health course. J Vet Med Educ. 2009;36:403–410.
18 Carrier LM. College students' choices of study strategies. Percept Mot Skills. 2003;96:54–56.
19 Eilertsen TV, Valdermo O. Open-book assessment: A contribution to improved learning? Stud Educ Eval. 2000;26:91–103.
20 Heijne-Penninga M, Kuks JB, Hofman WH, Cohen-Schotanus J. Directing students to profound open-book test preparation: The relationship between deep learning and open-book test time. Med Teach. 2011;33:e16–e21.
21 Rakes GC. The effects of open book testing on student performance in online learning environments. https://oit.utk.edu/instructional/development/rite/projects/Documents/rakes_rite_06.pdf. Accessed September 3, 2015.
22 Theophilides C, Koutselini M. Study behavior in the closed-book and the open-book examination: a comparative analysis. Educ Res Eval. 2000;6:379–393.
23 Wachsman Y. Should cheat sheets be used as study aides in economics tests? Econ Bull. 2002;1:1–11.
24 Baillie C, Toohey S. The "power test": Its impact on student learning in a materials science course for engineering. Assess Eval Higher Educ. 1997;22:33–48.
25 Block RM. A discussion of the effect of open-book and closed-book exams on student achievement in an introductory statistics course. PRIMUS. 2012;22:228–238.
26 Dickson KL, Bauer JJ. Do students learn course material during crib sheet construction? Teach Psychol. 2008;35:117–120.
27 Ben-Chaim D, Zoller U. Examination-type preferences of secondary school students and their teachers in the science disciplines. Instr Sci. 1997;25:347–367.
28 Dickson KL, Miller MD. Authorized crib cards do not improve exam performance. Teach Psychol. 2005;32:230–233.
29 Jehu D, Picton CJ, Futcher S. The use of notes in examinations. Br J Educ Psychol. 1970;40:335–337.
30 Weber LJ, McBee JK, Krebs JE. Take home tests: An experimental study. Res High Educ. 1983;18:473–483.
31 Agarwal PK, Karpicke JD, Kang SK, Roediger HL, McDermott KB. Examining the testing effect with open- and closed-book test. Appl Cogn Psychol. 2008;22:861–876.
32 Brightwell R, Daniel J, Stewart A. Evaluation: Is an open book examination easier? Biosci Educ. 2004;3. doi: 10.3108/beej.2004.03000004.
33 Heijne-Penninga M, Kuks JB, Hofman WH, Muijtjens AM, Cohen-Schotanus J. Influence of PBL with open-book tests on knowledge retention measured with progress tests. Adv Health Sci Educ Theory Pract. 2013;18:485–495.
34 Ioannidou MK. Testing and life-long learning: Open-book and closed-book examination in a university course. Stud Educ Eval. 1997;23:131–139.
35 Kalish RA. An experimental evaluation of the open book examination. J Educ Psychol. 1958;49:200–204.
36 Krarup N, Naeraa N, Olsen C. Open-book tests in a university course. High Educ. 1974;3:157–164.
37 Pauker JD. Effect of open book examinations on test performance in an undergraduate child psychology course. Teach Psychol. 1974;1:71–73.
38 Schumacher CF, Butzin DW, Finberg L, Burg FD. The effect of open- vs. closed-book testing on performance on a multiple-choice examination in pediatrics. Pediatrics. 1978;61:256–261.
39 Shine S, Kiravu C, Astley J. In defence of open-book engineering degree examinations. Int J Mech Eng Educ. 2004;32:197–211.
40 Whitley BE. Does "cheating" help? The effect of using authorized crib notes during examinations. Coll Stud J. 1996;30:489–493.
41 Heijne-Penninga M, Kuks JB, Schönrock-Adema J, Snijders TA, Cohen-Schotanus J. Open-book tests to complement assessment-programmes: Analysis of open and closed-book tests. Adv Health Sci Educ Theory Pract. 2008;13:263–273.
42 Phillips G. Using open-book tests to strengthen the study skills of community-college biology students. J Adolesc Adult Lit. 2011;49:574–582.
43 Skidmore RL, Aagaard L. The relationship between testing condition and student test scores. J Instr Psychol. 2004;31:304–314.
44 Schutz PA, Pekrun R, eds. Emotion in Education. Burlington, Mass: Academic Press; 2007.
45 McConnell MM, Eva KW. The role of emotion in the learning and transfer of clinical skills and knowledge. Acad Med. 2012;87:1316–1322.
46 Norman GR, Smith EK, Powles AC, Rooney PJ, Henry NL, Dodd PE. Factors underlying performance on written tests of knowledge. Med Educ. 1987;21:297–304.
47 Schuwirth LW, van der Vleuten CP, Donkers HH. A closer look at cueing effects in multiple-choice questions. Med Educ. 1996;30:44–49.
48 Ward WC. A comparison of free-response and multiple-choice forms of verbal aptitude tests. Appl Psychol Meas. 1982;6:1–11.
49 Roediger HL 3rd, Butler AC. The critical role of retrieval practice in long-term retention. Trends Cogn Sci. 2011;15:20–27.
50 Larsen DP, Butler AC, Roediger HL 3rd. Test-enhanced learning in medical education. Med Educ. 2008;42:959–966.
51 Schuman J. Declining board exam pass rates: Blame millennial doctors? KevinMD.com. July 8, 2013. http://www.kevinmd.com/blog/2013/07/declining-board-exam-pass-rates-blame-millennial-doctors.html. Accessed September 3, 2015.
52 Pekrun R. The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. Educ Psychol Rev. 2006;18:315–341.
53 Lipner RS, Lucey CR. Putting the secure examination to the test. JAMA. 2010;304:1379–1380.
54 American Board of Ophthalmology. Maintenance of certification. Why did the MOC examination change from an open-book, take-home exam to a proctored, computerized exam? http://abop.org/faqs/maintenance-of-certification/#exam. Accessed August 26, 2015.

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

Appendix 1

Coding Results for the 37 Articles Selected in a Systematic Review to Compare Open- and Closed-Book Examinations, 2013–2014

Articles by category[a]	No. of participants	Participants' prior experience with OBE	Level of the participants' investigated	Country	Outcomes	How measured?	Key findings
Exam preparation							
Feldhusen (1961)[1]	90	Not stated	Undergraduate students	United States	Preparation tactics	Self-reported questionnaire	The students felt that it is useless to "cram" for exams and that the open-book exam reduces memorization of factual material in preparation for OBE.
Theophilides and Dionysiou (1996)[2]	173	Not stated	Undergraduate students	Greece	Preparation tactics	Self-reported questionnaire	The perceived functions of OBE include a factor of exam preparation. The items within this factor are: When preparing for exam, Compares and contrasts information obtained; studies various resources; interrelates information acquired and conclusions drawn; reconstructs course content and integrates knowledge gained; practices study skills (note taking, textbook studying).
Heijne-Penninga et al (2008)[3]	570	Yes	Medical students (second- and third-year)	Netherlands	Exam preparation and testing effects	Deep information processing questionnaire	Counter to the hypothesis and prevailing wisdom, CBE preparation and not OBE preparation was associated with deep learning.
Agarwal and Roediger (2011)[10]	108	Not stated	Undergraduate students	United States	Study time	Timed by computer	Participants studied for less time when expecting an OBE.
Boniface (1985)[11]	30	Not stated	Undergraduate students	United Kingdom	Preparation efforts	Self-reported questionnaire	Students thought they would do more preparation for CBE than OBE.
Moore and Jensen (2007)[12]	351	Not stated	Undergraduate students	United States	Preparation efforts; class attendance	Informal discussion and class observation	Students seemed to prepare for OBE in the same way that they prepared for CBE. However, some students' class attendance dropped significantly when the upcoming exam was an OBE.
Gharib et al (2012)[13]	387	Not stated	Undergraduate students	United States	Preparation time	Self-reported questionnaire	The actual reported study time for OBE versus CBE did not differ, although the students believed that they would study most for CBE.
Betts et al (2009)[14]	116	Implied—OBE seems to be part of the curriculum	Undergraduate students	United Kingdom	Preparation efforts	Self-reported questionnaire	Students reported preparing for the OBE and CBE to a similar extent. However, there was a marginally significant interaction between examination condition and gender. Females prepared more for CBE than for the OBE. There was no significant difference in the amount that males prepared for the OBE and CBE.
Heijne-Penninga et al (2010)[15]	239	Yes	Medical students (second-year)	Netherlands	Preparation time	Self-reported questionnaire	The students reported more preparation time for CBE than OBE.

(Appendix 1 continues)

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

Appendix 1
(Continued)

Articles by category[a]	No. of participants	Participants' prior experience with OBE	Level of the participants' investigated	Country	Outcomes	How measured?	Key findings
Broyles et al (2005)[16]	174 (only 18 were interviewed after the exams)	Not stated	Medical students (third-year)	United States	Preparation tactics	Interview	More than 60% of the students did not change study tactics for OBE (those who did change studied less); 25% waited until last week of 4-week rotation to study.
Dale et al (2009)[17]	14	2/14 had prior experience with OBE	Graduate, professional education	United Kingdom	Preparation tactics	Self-reported questionnaire and interview	No statistically significant correlation was found between perceptions of different assessment methods (OBEs, essays, SAQs, MCQs) and deep versus surface learning approach scores. Most students felt prepared for the OBEs.
Carrier (2003)[18]	58	Not stated	College students	United States	Preparation tactics	Self-reported questionnaire	For CBE, surface studying was mostly done, but deep approach (studying lecture notes, making chapter notes, highlighting and/or underlining, and coming to office hours) correlated with high scores. For OBE, similar proportions of students used surface and deep approach; nothing correlated with examination performance.
Eilertsen and Valdermo (2000)[19]	350	Not stated	High school students	Norway	Training for OBE	Action research: survey, interview, class observation	The problem of not preparing thoroughly for OBE declines when OBE is applied over a period of time and as part of a broader approach aimed at strengthening students' understanding of learning and knowledge. Many students need to learn at an early stage that they have to be equally well if not better prepared for OBE versus CBE.
Heijne-Penninga et al (2011)[20]	663	Yes	Undergraduate students	Netherlands	Preparation time and tactics	Self-reported questionnaire	Second- and third-year college students differed significantly in OBE prep time ($t(662) = 2.25$, $P < .01$). Third-year students spent less prep time and prepared more deeply.
Rakes (accessed online 2015)[21]	49	Participants took a practice test to familiarize with OBE.	Graduate students	United States	Training for OBE	Test performance; training intervention	In an online learning environment, the administration of OBE may adversely affect students' exam performance because they do not necessarily understand the requirements of OBE. Training may mitigate the inclination not to study for OBE.
Theophilides and Koutselini (2000)[22]	201/276 respondents to survey	Not stated	Undergraduate students	Greece	Preparation tactics	Self-reported questionnaire	When students expect an OBE, they are more attentive throughout the semester and engage in more study activities that promote deep learning of the course subject matter.
Wachsman (2002)[23]	299	Not stated	Undergraduate students	United States	Preparation for cheat-sheet exam (partial OBE)	Test performance	The combined effect of preparing and using cheat sheets is positively associated with students' test performance, even when controlling for preparation time.

(Appendix 1 continues)

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

Appendix 1
(Continued)

Articles by category[a]	No. of participants	Participants' prior experience with OBE	Level of the participants' investigated	Country	Outcomes	How measured?	Key findings
Baillie and Toohey (1997)[24]	Not clear	Had orientation in classroom	Undergraduate students	Australia	The effect of teaching preparation for OBE	Interview	The change of the teaching method of the course has helped the students to be prepared for OBE.
Block (2012)[25]	938	Not stated	Undergraduate students	United States	Preparation efforts	Instructors' observation	The students came to the OBE not fully prepared and expected to find the needed answers in the book. CBE with notecards led to better preparation.
Dickson and Bauer (2008)[26]	53	Not stated	Undergraduate students	United States	Preparation for cheat-sheet exam (partial OBE)	Interview	Preparing crib sheets does not enhance learning, but use of crib sheets enhanced test performance.
Test anxiety							
Feldhusen (1961)[1]	90	Not stated	Undergraduate students	United States	Worry, tension	Self-reported questionnaire	When comparing OBE and CBE, students reported less worry and tension with OBE.
Theophilides and Dionysiou (1996)[2]	173	Not stated	Undergraduate students	Greece	Anxiety	Self-reported questionnaire	Of the 5 factors that constituted a questionnaire about the characteristics of OBEs, none of the factors were related to exam anxiety. Further, the 5 factors did not vary by students' exam anxiety level or expected graduation grade.
Gharib et al (2012)[13]	396	Not stated	Undergraduate students	United States	Anxiety	Self-reported questionnaire	Students reported higher anxiety in a cheat-sheet exam relative to OBE (CBE anxiety not examined). Test anxiety measured right before the exam (cheat-sheet and OBE) was negatively correlated with scores on the exams.
Betts et al (2009)[14]	116	Implied—OBE seems to be part of the curriculum	Undergraduate students	United Kingdom	Anxiety	Self-reported questionnaire	Students reported feeling more anxious when a correction for guess was included in a CBE than when the correction was used in an OBE.
Broyles et al (2005)[16]	18	Not stated	Medical students (third-year)	United States	Anxiety, stress reduction	Interviews	Most students (80%) noted that they were less anxious and less stressed when taking OBEs.
Dale et al (2009)[17]	14	2/14 had prior experience with OBE	Graduate, professional education	United Kingdom	Enjoyment, stress	Self-reported questionnaire, interviews	Limited qualitative results suggest that OBEs were thought to be less stressful than traditional CBEs.
Eilertsen and Valdermo (2000)[19]	350	Not stated	High school students	Norway	Anxiety	Interviews, self-reported questionnaires	OBE may reduce anxiety for some, but at the same time, some aspects of OBE, such as unfamiliar assignments or shortage of time to make use of the available materials, can also be anxiety provoking.

(Appendix 1 continues)

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

Appendix 1
(Continued)

Articles by category[a]	No. of participants	Participants' prior experience with OBE	Level of the participants' investigated	Country	Outcomes	How measured?	Key findings
Theophilides and Koutselini (2000)[22]	181	Not stated	Undergraduate students	Greece	Perceived differences in exam types	Self-reported questionnaire	Limited qualitative results suggest that the OBE alternative reduced exam tension and stress. Participants stated that they approached OBEs with greater optimism and worked out their answers in a more relaxed way.
Baillie and Toohey (1997)[24]	55	Had orientation in classroom	Undergraduate students	Australia	Anxiety	Focus groups, interviews	Anxiety associated with taking OBE was not reduced as much as investigators expected. A large portion of students (45%) were just as stressed with OBE as CBE.
Dickson and Bauer (2008)[26]	53	Not stated	Undergraduate students	United States	Anxiety	Self-reported questionnaire	The vast majority of students (80%) reported that making a crib sheet reduced their stress during the exam.
Ben-Chaim and Zoller (1997)[27]	236	Not stated	High school students	Israel	Preferred exam types, anxiety, stress	Self-reported questionnaire, interviews	OBEs, particularly in the form of take-home exams without strict time limits, reduce the level of student anxiety.
Dickson and Miller (2005)[28]	52	Most students had experience with OBE.	Undergraduate students	United States	Anxiety	Self-reported questionnaire	Despite high expectations by 79% of students that using a "cheat sheet" would lower anxiety during a test, only 41% reported that using the cheat sheet during the test actually lowered anxiety.
Jehu et al (1970)[29]	29	Not stated	Undergraduate students	United Kingdom	Anxiety	Self-reported questionnaire	The availability of notes (cheat sheet) reduced anxiety *during* an exam but did not reduce anxiety *before* the exam.
Weber et al (1983)[30]	64	Not stated	Undergraduate students	United States	Anxiety	Self-reported questionnaire	Students believed that OBEs and take-home exams caused much less stress than CBEs.
Exam performance							
Heijne-Penninga et al (2008)[3]	570	Yes	Medical students (second- and third-year)	Netherlands	Immediate performance (I) + learning approach survey	Deep information processing survey + MCQ	Students scored significantly higher when preparing for CBE. Counter to the hypothesis and prevailing wisdom, CBE preparation and not OBE preparation was associated with deep learning.
Agarwal and Roediger (2011)[10]	108	Not stated	Undergraduate students	United States	I and delayed performance (D)	Short answer	OBE led to better initial performance but similar results after 2-day delay (delayed test). Participants studied for less time when expecting an OBE and correspondingly performed worse (10% improved performance with expecting CBE over OBE).

(Appendix 1 continues)

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

Appendix 1
(Continued)

Articles by category[a]	No. of participants	Participants' prior experience with OBE	Level of the participants' investigated	Country	Outcomes	How measured?	Key findings
Boniface (1985)[11]	30	Not stated	Undergraduate students	United Kingdom	D	MCQ + short answer	There were large negative correlations between the amount of time devoted to consulting notes and texts and exam score (R = –0.44 for both, –0.39 for notes, –0.13 for texts). The correlations between the amount of time devoted looking things up and scores on previous assessments were also negative (–0.33 for CBE).
Moore and Jensen (2007)[12]	351	Not stated	Undergraduate students	United States	I and D	MCQ	Average grades on Exam 1 in the control and experimental sections were not significantly different. However, grades on Exams 2 and 3 were significantly higher for OBE. On the final exam, grades were significantly higher for CBE. Some students' class attendance dropped significantly when the upcoming exam was an OBE.
Gharib et al (2012)[13]	387	Not stated	Undergraduate students	United States	OBE, CBE, partial OBE (cheat-sheet), I	MCQ	OBE and partial OBE > CBE score
Betts et al (2009)[14]	116	Implied—OBE seems to be part of the curriculum	Undergraduate students	United Kingdom	I and correction for guessing	MCQ	Scored higher and left fewer questions unanswered when no correction for guessing. Students favor the use of correction for guessing for OBE but not CBE.
Heijne-Penninga et al (2010)[15]	239	Yes	Medical students (second-year)	Netherlands	D and preparation time and need for cognition	MCQ	Need for cognition or the tendency of an individual to engage in effortful cognitive activities and to enjoy thinking positively influenced both OBE and CBE performance; deep learning and time for preparation did not correlate with performance on either OBE or CBE.
Broyles et al (2005)[16]	174	Not stated	Medical students (third-year)	United States	I	MCQ	OBE 88.2 versus CBE 84 (ANOVA P = .03). Small statistically significant improvement with OBE versus CBE.
Carrier (2003)[18]	58	Not stated	College students	United States	D and study preparation approach (surface or deep)	Self-reported questionnaire + MCQ, exam	For CBE, deep approach (studying lecture notes, making chapter notes, highlighting and/or underlining, and coming to office hours) correlated with higher performance. For OBE, deep or surface approach did not correlate with exam performance. Similar proportions of students used surface and deep approach for OBE and CBE.

(Appendix 1 continues)

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

Appendix 1
(Continued)

Articles by category[a]	No. of participants	Participants' prior experience with OBE	Level of the participants' investigated	Country	Outcomes	How measured?	Key findings
Heijne-Penninga et al (2011)[20]	663	Yes	Medical students (second- and third-year)	Netherlands	D	Self-reported questionnaire + MCQ	No difference in general; performed worse on CBE when tested on material previously received in OBE (or assumed that it would be tested by OBE)
Rakes (accessed online 2015)[21]	49	Participants took a practice test to familiarize with OBE.	Graduate students	United States	D and additional arm of OBE training	MCQ	CBE performance superior to OBE; OBE scores improved with explicit training. In online learning environment, the administration of OBE may adversely affect students' exam performance because they do not necessarily understand the requirements of OBE. Training may mitigate the inclination not to study for OBE. Partial OBE helped students who received C and D grades on prior CBE.
Wachsman (2002)[23]	299	Not stated	Undergraduate students	United States	I and D (cheat-sheet versus CBE)	MCQ	The combined effect of preparing and using cheat sheets is positively associated with students' test performance, even when controlling for preparation time.
Block (2012)[25]	938	Not stated	Undergraduate students	United States	D	MCQ + short answer	Better delayed performance when prior exam CBE versus OBE format
Dickson and Bauer (2008)[26]	53	Not stated	Undergraduate students	United States	Cheat-sheet exam (partial OBE) versus CBE, I	Interview + MCQ	Preparing crib sheets does not enhance performance, but use of crib sheets enhanced test performance.
Dickson and Miller (2005)[28]	54	Most students had experience with OBE	Undergraduate students	United States	Cheat-sheet exam (partial OBE) versus CBE, I	Questionnaire + MCQ	Using crib card did not lead to higher exam scores on either low- or high-order MCQs despite students' belief that it would.
Jehu et al (1970)[29]	29	Not stated	Undergraduate students	United Kingdom	Partial OBE (note sheet) versus CBE	Essay	No significant difference in scores
Weber e al (1983)[30]	64	Not stated	Undergraduate students	United States	I (take-home test versus OBE versus CBE)	MCQ	No significant differences existed between take-home exams (64.9%) and OBEs (61.5%) or between OBEs and CBEs (57.9%).
Agarwal et al (2008)[31]	36 (Experiment 1); 48 (Experiment 2)	Not stated	Undergraduate students	United States	I and D	MCQs	Initial: OBE superior performance than CBE; both superior to studying alone. Delayed: OBE and CBE effect equivalent. CBE with feedback resulted in greater final performance than without it.
Brightwell et al (2004)[32]	196	Not stated	Undergraduate students	Australia	D	Online MCQ examination	No differences in mean, minimum, or maximum scores
Heijne-Penninga et al (2013)[33]	1,648	Students at the school with a PBL curriculum and both CBE and OBE had prior experience	Medical students (fifth- and sixth-year)	Netherlands	D and PBL versus traditional learning (TL) approach	Not stated	OBE and CBE for PBL students higher than TL students

(Appendix 1 continues)

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

Appendix 1
(Continued)

Articles by category[a]	No. of participants	Participants' prior experience with OBE	Level of the participants' investigated	Country	Outcomes	How measured?	Key findings
Ioannidou (1997)[34]	72	Not stated	Undergraduate students	Cyprus	D	MCQ + essay	No significant difference in scores on OBE versus CBE
Kalish (1958)[35]	158	Not stated	Undergraduate students	United States	I and D	MCQ	The results have indicated that, although under the conditions of this experiment the group average scores are not affected by OBE versus CBE, the 2 types of examinations appear to measure different abilities (based on correlation between Test 1 and Test 2 being higher when both tests were CBE relative to when the second one was OBE).
Krarup et al (1974)[36]	108	Yes	Medical students (sixth-term)	Denmark	I and D	MCQ	No differences in scores overall; recall items (15%) showed higher OBE performance
Pauker (1974)[37]	96	Not stated	Undergraduate students	Canada	I and D	MCQ	No difference in scores by examination format
Schumacher et al (1978)[38]	196 students, 96 practicing pediatricians	Not stated	Practice group (practicing pediatricians) and student group (third- and fourth-year medical students)	United States	I	MCQ	For both tests, practicing MDs performed better. Medical students did better on OBE than CBE; practicing MDs' performance on OBE and CBE not different.
Shine et al (2004)[39]	131	Not stated	Undergraduate students	Botswana, New Zealand, and United Kingdom	I and D	MCQ + essay + short answer	Averages of test performance same or lower on OBE than CBE. OBE takes examiners more thought and skill in designing tests.
Whitley (1996)[40]	136	Not stated	Undergraduate students	United States	Partial OBE (crib sheet) versus CBE, I	MCQ and short answer	Equivocal effect: Using notes improved performance in 1 of 2 examination sessions.
Heijne-Penninga et al (2008)[41]	934	Not stated	Undergraduate students	Netherlands	I	MCQ	In 9 out of 14 examinations, the closed-book exam scores were significantly higher. In the others, the open-book exam scores were significantly higher.
Phillips (2011)[42]	1,080	Not stated	Undergraduate students	United States	I and D	Not stated	Mean score on the third OBE was significantly higher than the first. Weaker students' improvement was the greatest.
Skidmore and Aagaard (2004)[43]	141	Not stated	Undergraduate students	United States	I	MCQ	Students got higher mean test scores in OBE than CBEs. The lowest-achieving students benefited most from the group discussion that helped them fill in the gaps for each other.
Testing effects[b]							
Feldhusen (1961)[1]	90	Not stated	Undergraduate students	United States	Self-perceived judgment of learning	13-item questionnaire administered at end of semester	Students felt that OBEs were superior to CBEs with promoting learning during testing.

(Appendix 1 continues)

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

Appendix 1
(Continued)

Articles by category[a]	No. of participants	Participants' prior experience with OBE	Level of the participants' investigated	Country	Outcomes	How measured?	Key findings
Theophilides and Dionysiou (1996)[2]	173	Not stated	Undergraduate students	Greece	Deep learning	38 statements with 5-point Likert scale	OBEs perceived by students to have 5 advantages including allowing students to creatively use the knowledge they gained in the course and encouraging students to apply deep approaches to studying.
Agarwal and Roediger (2011)[10]	108	Not stated	Undergraduate students	United States	Self-perceived judgment of learning	0%–100% self-perception of how well a passage would be remembered in 1 week	Students erroneously predicted that studying (rather than taking a test) would lead to better retention.
Dale et al (2009)[17]	14	2/14 had prior experience with OBE	Graduate, professional education	United Kingdom	Deep versus surface learning	Questionnaire including Approaches and Study Skills Inventory for Students, interviews	A small association was observed between a reported preference for a deep learning approach and favorable responses regarding OBE.
Eilertsen and Valdermo (2000)[19]	350	Not stated	High school students	Norway	Attitudes toward OBE	Mixed methods with questionnaires and focus group interviews	OBE format was helpful in getting the students and teachers to understand the nature of knowledge and process of learning.
Dickson and Bauer (2008)[26]	53	Not stated	Undergraduate students	United States	Self-report measure of helpfulness of crib sheet with respect to perceived learning	End-of-semester questionnaire	91.8% students felt using a "crib sheet" was helpful for learning. However, crib sheet preparation was not associated with improved exam performance (when crib sheet was not used).
Dickson and Miller (2005)[28]	54	Most students had experience with OBE.	Undergraduate students	United States	Self-perceived judgment of learning	Pre- and postsemester questionnaire assessing perceived effect on exam performance and learning (post only)	Perceived learning benefits from crib sheets mixed. Students' perceived effect on exam performance was more positive before the semester than at the end.
Kalish (1958)[35]	158	Not stated	Undergraduate students	United States	Self-perceived judgment of learning	Single-item question dealing with "degree of help" with OBE format	No relationship between perceived degree of assistance with OBE format and performance

Abbreviations: OBE indicates open-book examination; CBE, closed-book examination; MCQ, multiple-choice question; SAQ, short-answer question; PBL, problem-based learning.

[a]Psychometrics and logistics and public perception categories are not listed in this appendix because there were no unique articles for these categories (all articles included in the review are listed in this appendix). See the Results section for summaries of the findings in these categories.

[b]Also see Appendix 2, in which the authors explicitly used testing effects as a theoretical framework for the investigation.

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

Appendix 2

From a 2013–2014 Systematic Review of 37 Articles Comparing Open- and Closed-Book Examination, Selected Articles That Explicitly Stated the Use of the Testing Effect as a Theoretical Framework[a]

Reference no.[b]	Author(s)	Theoretical framework(s)	Participants	Test format	Quantitative results	Key finding(s)
31A	Agarwal et al (2008)	Testing effect	36 undergraduate psychology students	1-week delayed CBE (short-answer responses to passages)	CBE mean = 0.59 CBE with feedback mean = 0.68[c] OBE mean = 0.65[c]	Testing effect demonstrated in all exam types. Taking OBE or CBE with receiving feedback prior to final exam was superior to taking CBE without receiving feedback.
31B	Agarwal et al (2008)	Testing effect	48 undergraduate psychology students	1-week delayed CBE (short-answer responses to passages)	CBE mean = 0.55 CBE with feedback mean = 0.66[c] OBE mean = 0.66[c]	OBE or CBE with receiving feedback prior to final exam was superior to CBE without receiving feedback prior to final exam.
10A	Agarwal and Roediger (2011)	Testing effect, Barnett and Ceci's taxonomy of transfer, Bloom's taxonomy	72 undergraduate students (Department of Psychology subject pool)	2-day delay GRE-style passages and MCQs	Study-only mean = 0.49 OBE mean = 0.63[c] CBE with feedback mean = 0.61[c]	Equivalent testing effect demonstrated for OBE and CBE (with feedback) compared with study-only condition
10B	Agarwal and Roediger (2011)	Testing effect	108 undergraduate students (Department of Psychology subject pool)	2-day delay GRE-style passages and MCQs	Delayed fact CBE mean (SD) = 0.22 (0.01) OBE mean (SD) = 0.20 (0.01) Delayed comprehension CBE mean (SD) = 0.63 (0.01) OBE mean (SD) = 0.66 (0.02) Delayed transfer CBE mean (SD) = 0.40 (0.02) OBE with feedback mean (SD) = 0.42 (0.02)	Similar to 2A, replicated similar performance on 2-day delayed CBE regardless of prior OBE or CBE (this time, practice CBE without feedback)
26	Dickson and Bauer (2008)	Coding hypothesis (improved learning with crib sheet) and dependency hypothesis (crib sheet changes study habits)	53 undergraduate psychology students	15-item CBE pretest MCQ exams prior to partial OBE	First closed-book pretest mean = 55.7 (SD 15.8)[c] First partial OBE mean (SD) = 74.7 (13.9)[c] Second closed-book pretest mean (SD) = 62.6 (16.8)[c] Second partial OBE mean (SD) = 69.1 (17.8)[c]	Improvements in performance seen in partial OBE were not seen on closed-book pretests of identical questions calling improved learning through crib sheet creation into question.
13A	Gharib et al (2012)	Testing effect, partial OBE	297 undergraduate, Intro to Psychology (only 2 of 5 sections took retention quizzes—191 quizzes)	2-week, unannounced closed-book 10-item retention quiz, MCQ	OBE mean (SD) = 6.41 (1.94) Partial OBE mean (SD) = 6.44 (1.88) CBE mean (SD) = 6.38 (1.94)	No differences in testing effects demonstrated between OBE, partial OBE, or CBE formats in an undergraduate Introduction to Psychology class
13B	Gharib et al (2012)	Testing effect, partial OBE	99 undergraduate, Intro to Statistics (343 retention quizzes analyzed)	2-week, unannounced closed-book 10-item retention quiz, MCQ	OBE mean (SD) = 6.18 (1.75) Partial OBE mean (SD) = 6.31 (1.83)	No differences in testing effects noted in undergraduate statistics class
33	Heijne-Penninga et al (2013)	Testing effect, information-processing theory, knowledge organization, PBL curriculum	1,648 medical students (Denmark)	200 CBE MCQ progress tests (4 tests annually, core and backup knowledge)	OBE curriculum core knowledge was significantly higher versus CBE curriculum on 4 of 8 tests ($P < .0021$, 1-sided t test) (no difference with backup knowledge).	A small benefit in "core knowledge" retention was observed in a cohort of medical students whose PBL curriculum included OBE assessment.

(Appendix 2 continues)

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

Appendix 2
(Continued)

Reference no.[b]	Author(s)	Theoretical framework(s)	Participants	Test format	Quantitative results	Key finding(s)
12	Moore and Jensen (2007)	Testing effect	351 undergraduate Intro to Biology students (United States)	70 MCQ closed-book final exam	Overall CBE mean (SD) = 74 (4) OBE mean (SD) = 63 (5)[c] Matched content (second and third exams) CBE mean (SD) = 75 (4) OBE mean (SD) = 57 (7)[c] CBE mean (SD) = 74 (30) OBE mean (SD) = 61 (6)[c]	Testing effect of CBE superior to OBE. Further, the content they scored lower on matched the content covered by the OBEs.
37	Pauker (1974)	Testing effect	96 undergraduate students (Canada)	75 MCQ CBE final	OBE mean (SD) = 52.8 (8.3) CBE mean (SD) = 54.8 (5.8)	Testing effect identical except for lowest tertile of students who performed significantly better if all prior tests CBE

Abbreviations: CBE indicates closed-book examination; OBE, open-book examination; GRE, Graduate Record Examination; MCQ, multiple-choice question; PBL, problem-based learning.
[a]Use of additional explicit theoretical frameworks is also shown.
[b]A and B refer to articles where there was more than one investigation reported.
[c]Statistically significant difference.

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.