

Supplementary Material*

Lipner RS, Brossman BG, Samonte KM, Durning SJ. Effect of Access to an Electronic Medical Resource on Performance Characteristics of a Certification Examination. A Randomized Controlled Trial. *Ann Intern Med*. doi:10.7326/M16-2843

Supplement. **Study Protocol**

* This supplementary material was provided by the authors to give readers further details on their article. The material was reviewed but not copyedited.

Supplemental Material: Invitation Letter

Dear Dr. [REDACTED]

As the American Board of Internal Medicine (ABIM) partners with physicians to enhance the Maintenance of Certification (MOC) program, we are seeking your input in an upcoming study **to examine the possibility of ABIM providing access to online resources, such as UpToDate, during an assessment of medical knowledge.**

ABIM has heard from many physicians that they would like to have access to online resources during MOC assessments since they use these tools in practice. This study will provide crucial insights into how the use of online resources might impact both diplomate experience and performance.

Details about the study:

- Participating physicians will complete a five-hour pilot exam (including a break and tutorial) in one of Pearson VUE's testing centers across the country between **April 17 and September 1, 2016.**
- **This is not the Internal Medicine MOC exam.** Your decision to participate and your performance on the pilot exam will have no effect on your certification status. Individual scores will not be available to participants. ABIM will share feedback with participants about their performance on the pilot test and will report collective results of the study to the community.
- **Those who participate will be randomly assigned to one of two groups.** In one group, participants will have access to UpToDate, an online information resource, for a part of the exam. In the control group, participants will not have access to online materials. UpToDate is the only online resource in this pilot, but other resources will likely be considered for use in the future.
- We realize that you are busy and know that your time and input is important. In appreciation, we are offering everyone who fully participates in the study:
 - An honorarium of [REDACTED]
 - [REDACTED] MOC points
 - Complimentary access to UpToDate during the research period

How to sign up:

We hope you will be willing to help us enhance our program. To participate in this study, please contact **xxxx** to schedule your pilot exam appointment and choose your test center:

- Schedule online at [REDACTED] and sign up for the "Research Study on Open Book Assessments"
- OR
- Call [REDACTED].

When you schedule your exam appointment, be sure to use **[ABIM ID]** as your identification number with Pearson VUE.

If you have questions regarding signing up, please call [REDACTED].

Supplemental Material. Acknowledgment form

You have signed up to participate in an American Board of Internal Medicine's (ABIM) research study on the potential of providing access to open-book materials on portions of the Maintenance of Certification (MOC) exam. This research is important in understanding the feasibility of this kind of change and any impact it has on diplomate experience and performance.

PARTICIPATION: By agreeing to participate in this study, you have scheduled a five-hour pilot exam (including a break and tutorial) in one of Pearson VUE's testing centers. This pilot exam may or may not consist of sections that will allow access to UptoDate as a web resource. Participants will also be asked to complete a survey following the completion of the pilot exam.

It is important that you approach the pilot exam as if it was a real Maintenance of Certification (MOC) exam, answering all questions to the best of your ability. This will allow ABIM to determine the true effects of accessing web resources on both diplomate experience and performance on the exam.

Participation in this study is completely voluntary and will not affect your certification status. If you choose to withdraw from the study, or have any questions related to your participation, please contact [REDACTED]

You must complete this acknowledgement and consent form, attesting that you will take the pilot exam in good faith, answering all questions to the best of your ability, in order to receive the incentives listed below. If you voluntarily withdraw from the study, or do not take the pilot exam in good faith, you will not receive the incentives listed below.

INCENTIVES: Participants will receive a \$250 honorarium and 20 MOC points upon completion of the pilot exam. All participants will also receive complementary access to UptoDate for 6 months.

CONFIDENTIALITY: All information collected in this study will be strictly confidential. The results of the study may be published or presented, but all data will be aggregated and will not identify any individual participant.

PARTICIPANT STATEMENT

My participation in this research project is voluntary and I may withdraw from the study at any time without harm. The investigators involved in this study may decide at any time that I will not be eligible for the incentives if I do not take the pilot exam in good faith, answering all the questions to the best of my ability.

By signing this participant statement, I have not given up any of my legal rights. I have read and understand the above information. I agree to participate in this study, and to approach this pilot exam as if it was a real Maintenance of Certification (MOC) exam, answering all questions with the single best answer I would select given all the information. An electronic copy of this statement will be maintained.

Let us know!

Out of the following, what were the **top two** incentives you considered when deciding to participate in this research study (check two responses only)?

Honorarium

MOC points

Access to UpToDate

Practice for the exam

Being a part of transforming

MOC Other (please specify):

Please let us know any additional information about your decision to participate in this research study.

Supplemental Material: Protocol

A Comparison of Open Book and Closed Book Exam Formats for a Large-Scale Medical Certification Exam

Protocol #: 16-01

Initial Submission: 4/1/2016

Principal Investigator:

Rebecca S. Lipner, Ph.D.

American Board of Internal Medicine

510 Walnut Street, Suite 1700

Philadelphia, PA 19106

215-446-3541

rlipner@abim.org

TABLE OF CONTENTS

I. Background and Introduction.....	3
II. Study Rationale.....	4
III. Objective(s).....	4
IV. Data Sample.....	5
V. Study Assessments-Plan and Methods.....	5
VI. Confidentiality.....	7
VII. Intended Use of the Data.....	7
VIII. Plans for Dissemination of Findings.....	7
IX. Evaluation of Adverse Events.....	7
X. Ethical Considerations.....	7
XI. Study Monitoring and Oversight.....	7
XII. Data Analysis.....	8
XIII. Investigator Statement.....	9
XIV. References.....	9

I. Background and Introduction

The widespread use of and perpetual advancements in technology provide physicians access to a wealth of readily-available information. This influx of information may allow physicians to approach clinical scenarios differently than in the past, when information was less accessible. With the current state of technology, physicians can look up necessary information in everyday practice. This approach to practice is not currently reflected by the secure exams produced by the ABIM. Currently, the ABIM secure exams are closed-book exams (CBE). Proponents of open-book exams (OBE) posit that permitting examinees to access external material better simulates real-world clinical scenarios and allows examinees to focus on applying skills rather than recalling knowledge. The current study aims to explore the OBE format to determine the impact of implementing a partially OBE for the ABIM Maintenance of Certification (MOC) secure exam.

Currently, there is limited research comparing closed- and open-book exams in high-stakes medical certification exams. One study reviewed the existing literature surrounding comparisons between closed- and open-book exams and noted that only 5% ($n = 2$) of the studies included were considered high-stakes and only 22% ($n = 8$) studied medical students¹. The authors note, however, that there were minimal differences in results across learner types (e.g., undergraduate, physicians in practice) and felt comfortable discussing the aggregate trends in the literature. The summary of the literature examined several outcomes comparing CBEs and OBEs: examination preparation, test anxiety, exam performance, psychometrics and logistics, testing effects, and public perception. Overall, the results found minimal differences between CBEs and OBEs. While findings were inconsistent across studies, results generally suggested that students tend to prepare more for CBEs over OBEs. The authors note that test anxiety was typically examined as a secondary research question. Results of the studies suggested that individuals tend to overestimate the impact OBEs have on reducing test anxiety. The authors found that the research tends to favor CBEs with respect to exam performance, but note that lack of familiarity with OBEs may have been a confounding factor. There were only two studies that examined the impact of exam format on the psychometric utility of an assessment which found that both CBEs and OBEs could be used to discriminate reasonably well between student abilities. Additionally, the authors also examined differences in testing effects, which are improvements in subsequent performance after taking an exam, between CBEs and OBEs. Ultimately, authors found that testing effects are observed for both CBEs and OBEs. Finally, the authors summarized results related to public perception, from the teacher and the learner's perspectives, of CBEs and OBEs. For the learners, results were mixed such that some studies found that OBEs were seen to have several advantages over CBEs, whereas other studies found that OBEs were thought to have more difficult items and required learners to have more practice or training with the OBE format. Teachers' perceptions were often concerned with the difficulty of implementing OBEs. The authors conclude that there is not sufficient evidence to exclusively support OBEs or CBEs, and suggest a combined approach that allows both formats to appear on the same exam.

The literature review also noted that the decision to use an OBE or a CBE format largely depends on the logistical feasibility of implementing the exam with sufficient security,

reliability, and validity¹. Lipner and Zhang conducted a proof of concept study to examine the potential for ABIM to incorporate OBE formats into the secure exam². Specifically, the project aimed to examine whether access to web resources could be done securely, effectively, and authentically. The authors worked with Pearson VUE to set up a short exam made up of 60 items no longer used in the ABIM self-evaluation modules. After the exam, a survey was conducted to gather information about the examinee's experience and the authenticity of the external resource. The proof of concept found that Pearson's system was able to access the web content, but did not stop users from accessing other external content (e.g., email, Facebook). Pearson's system also had trouble managing scripting errors and the navigation between multiple windows was not user-friendly. However, given that the proof of concept was in its early stages, Pearson's effort to address all the errors was limited and acknowledged the potential for addressing the errors more aggressively in the future. Finally, the results suggested that the content used, Isabel's medical content, was not a suitable resource for an ABIM secure exam.

Overall, there is insufficient evidence examining the advantages and disadvantages of CBE and OBE formats. Much of the current research suggests minimal differences between the formats or is slightly in favor of the CBE format. The proof of concept study suggested that implementing external resources with Pearson's system is possible, but may still need work before an OBE format could be administered securely and with the appropriate information. The current study aims to explore the impacts and feasibility of implementing a partially OBE format for the ABIM's secure exam. The study will consider the psychometric impact of an OBE format, logistical concerns (i.e., timing, security), and the appropriateness of potential external resources.

II. Study Rationale

The ABIM plans to evaluate how external resources could be used on the secure exams. These exams are currently "closed book" in that candidates and diplomates are not allowed to access information via online medical resources (e.g., "UpToDate") or any other print materials. The purpose of this project is to determine the feasibility of moving toward an "open book" format for all or some portion of the ABIM secure exams.

III. Objective(s)

This research study will address questions related to how the use of external resources may impact what is being measured including any psychometrics and test development issues. The output of this project will be an evaluation of the use of external resources on ABIM exams. This evaluation will consider how the use of resources impacts perceptions of the exam, time allowances for the exam (and therefore costs), and the impact of having external resources on the psychometric properties of the exam.

Specifically, the research questions that we plan to address through the research study are as follows:

1. Would allowing external resources lengthen the time required to complete the exam?
2. How does the measurement construct change with the addition of external resources, and how does this impact:
 - a. Classical item statistics

- b. IRT parameters
- c. Test equating
- d. Standard setting

IV. Data Sample

ABIM diplomates who recently completed the Internal Medicine MOC exam will be contacted and asked to participate in a research study. The sample of participants will be representative of most of the population of exam takers and include an appropriate composition of ability levels: Diplomates who passed the exam on their first attempt, diplomates who failed the exam on their first attempt and passed on their second attempt, and diplomates who failed the exam on two attempts and who have not yet taken their third attempt.

V. Study Assessments-Plan and Methods

In order to address these questions, diplomates who recently completed the Internal Medicine MOC exam will be contacted and asked to participate in a research study. The sample of participants will be representative of most of the population of exam takers and include an appropriate composition of ability levels. Prior research has indicated that repeated attempts on medical certification exams does not significantly improve performance on subsequent exams, so there is little reason to believe that participating in this study will unfairly advantage diplomates who need to retake the secure exam for the third time.

Each participant will be asked to report to a Pearson testing center on a specific date to complete a modified version of the ABIM secure exam. An incentive of \$250, 20 MOC points, and complementary access of UpToDate for a period of time will be provided to ensure that motivation does not impact the results of the study. Each exam will consist of 120 IM MOC questions, or two modules of 60 questions each. Whereas Module 1 will be “closed book” for every participant (i.e., no external resources permitted), Module 2 will either be closed book or “open book” (i.e., external resources permitted). This allows the first module to serve as a control between the groups for test equating, while the second module serves as a point of comparison between open book and closed book formats.

In total, 800 diplomates who agree to participate will be randomly assigned to one of four equal-sized groups (200 diplomates per group):

1. Closed-book exam group using the current ABIM time constraints (CBE)
2. Closed-book exam group with 15 additional minutes for the second module (CBE+15)
3. Partial open-book exam group using the current ABIM time constraints (OBE)
4. Partial open-book exam group with 15 additional minutes for the second module (OBE+15)

The additional time allotment for the second module will allow the investigators to determine whether adding additional time per question impacts exam performance. Durning et al. indicated that including external resources increased testing time between 10%-60%.¹ It should be noted

that the additional 15 minutes per 60-item module is solely for research purposes and will likely not translate to how we would proceed operationally. We would likely reduce the number of exam questions to fit within the current time constraints. For example, modules that allow for the use of external resources may consist of 55 items rather than 60 items. This would give the diplomates more time per question while at the same time still fitting within the current ABIM time constraints.

Each of the four groups included in this study are important, as each serves a specific purpose. The CBE group represents the current conditions for the ABIM secure exams and can be used as the baseline for comparison. The CBE vs. OBE comparison identifies whether including external resources impacts IRT calibration, scoring (e.g., diplomate performance), equating, and standard setting. The OBE vs. OBE+15 comparison identifies whether diplomates are unfairly disadvantaged when external resources are available yet less time is allotted to adequately use the resources. However, if the OBE vs. OBE+15 comparison demonstrates that more time is required we cannot completely attribute this to the inclusion of external resources: it may be possible that diplomates naturally take more time when more time is available. The CBE vs. CBE+15 comparison will help to isolate whether additional testing time is actually required when external resources are available or whether diplomates just tend to use more time when more time is available.

Given that item response theory (IRT) will be used to compare performance amongst these groups, a power analysis was conducted that used average difference in IRT difficulty as the primary outcome. Incorporating a mean difference in item difficulty of 0.50 (SD = 1.5) between the open book and closed book formats, 516 total diplomates (129 per group) were required to achieve a power level of 0.90. An additional requirement for the study is that we meet appropriate levels of IRT measurement precision. Incorporating IRT measurement precision into this power calculation requires roughly 800 diplomates (or 200 per group).

Finally, the participants will complete a survey at the end of the administration which will collect information about their perceptions of the experience and use of OBEs for future ABIM MOC exams.

VI. Confidentiality

Data collected will be entered in a master file of physician demographic information (e.g. gender, age, years in practice, specialty). These data will not be linked to the name or any other identifying information of the physician. All data generated from the test groups will be de-identified and analyzed at a group level. Only researchers from the ABIM will have access to the study data which will be stored electronically in a secure, locked site. Any publications generated from the data will present data in the aggregate and will not be linked to any individual physician.

VII. Intended Use of the Data

The intent of the current study is to explore the impact and feasibility of implementing an OBE or partially OBE format for the ABIM secure exam. The study will consider the psychometric

impact of an OBE format, logistical concerns (i.e., timing, security), and the appropriateness of potential external resources.

VIII. Plans for Dissemination of Findings

We expect to publish the findings in peer-reviewed academic journals as well as national conferences. The expected end date for this study is December 31st 2016.

IX. Evaluation of Adverse Events

There are no anticipated adverse events. Any expressed concerns about the test will be tracked but they were not solicited.

X. Ethical Considerations

The proposed research study presents no more than minimal risk to participants. Participation in this study will have no effect on the subject's certification status.

The protocol will require IRB review and exemption. All IRB correspondence should be directed to the ABIM.

XI. Study Monitoring and Oversight

In publications, data will be anonymous, focusing on aggregate data. The results are descriptive, and will not make definitive statements about the value or lack of value of the study, rather offering support for value or lack of value and areas where it could be strengthened. Authorship has been determined by the investigators in advance of initiating writing or abstract preparation.

XII. Data Analysis

All statistical analyses will be completed using the programs SAS and MS-Excel. As the purpose of this study is to determine how response times, classical statistics, and IRT parameters are impacted by external resources, an initial analysis will be conducted that compares response times for each Module 2 item across each of the four groups (CBE, CBE+15, OBE, OBE+15). One-way ANOVAs will be used for this purpose, followed by pairwise comparisons in the event of significant differences. Preplanned comparisons will also be conducted – comparing response times from the combined CBE/OBE group with response times from the combined CBE+15/OBE+15 group – to determine whether allowing additional testing time impacts response times. Due to the skewed nature of response times, log normal transformations will first be applied prior to conducting the ANOVAs. Furthermore, due to the large number of hypothesis tests that will be run, a conservative alpha level of 0.01 will be used for all tests. Effect sizes will be used to determine whether statistically significant differences are of practical importance.

Depending on the outcome of this first analysis, the IRT calibration will then proceed in one of two ways. A Rasch calibration will be conducted in which Module 1 serves as the anchor set for

equating. For Module 2, each item will receive either two or four sets of Rasch parameter estimates, depending on the results of the timing data. If lengthening the exam does not have a statistically significant impact on response times (CBE/OBE vs. CBE+15/OBE+15 in the preplanned comparisons above), then each Module 2 item will receive two sets of Rasch parameter estimates: an open book set of parameters, and a closed book set of parameters. If lengthening the exam does have a statistically significant impact on response times, then each Module 2 item will receive four sets of Rasch parameter estimates, corresponding to each of the four study groups. The parameter estimates will then be used for the purpose of comparison.

Either a t-test or an ANOVA (depending on whether or not we can combine the CBE and OBE groups with the CBE+15 and OBE+15 groups, respectively) will be conducted to statistically compare the Rasch difficulty parameters for Module 2 items. This will indicate whether there is a systematic difference between the difficulty of open book and closed book items. Likewise, a t-test or ANOVA will be conducted that compares the classical discrimination (biserial correlation) across groups, indicating whether there is a systematic difference between the discrimination ability of open book and closed book items.

Whereas the prior analyses will be used to determine whether there are any *systematic differences* between open book and closed book items (e.g., “on average open book items are easier than closed book items”), differential item functioning (DIF) will be used to determine if external resources introduces multidimensionality beyond what is already present in the closed book exam. The DIF results indicate whether physicians of equal ability could be expected to perform differently on open book and closed book items, and therefore indicate whether closed book exams are dimensionally different from open book exams. To do this, ability estimates (“theta”) will be generated from the Rasch parameters, and – for each Module 2 item – DIF will be used to determine whether physicians of equal ability levels perform different on the open book and closed book version of the item.

XIII. Investigator Statement

The Principal Investigator has reviewed the clinical protocol thoroughly and agrees that it contains all the necessary information about the study.

XIV. References

1. Durning SJ, Dong T, Ratcliffe T, et al. Comparing Open- and Closed-Book Examinations: A Systematic Review. *Academic Medicine*. Available Online Ahead of Print November 2, 2015.
2. Lipner R, Zhang J. Report on External Resource Proof of Concept Research Study. American Board of Internal Medicine. 2013.