

## CATALYST: Piloting a Longitudinal Assessment and Learning Program for Board Recertification and Continuous Professional Development

Dorothy T. Horber, PhD; Joseph Flamini, MBA; John R. Gimpel, DO, MEd; Tsung-Hsun (Edward) Tsai, PhD; Kayse Shrum, DO; Kimberly Hudson, PhD, MEd

From the National Board of Osteopathic Medical Examiners in Conshohocken, Pennsylvania (Drs Horber, Gimpel, Tsai, and Hudson, and Mr Flamini), and the Oklahoma State University Center for Health Sciences in Tulsa, Oklahoma (Dr Shrum).

Financial Disclosures:  
None reported.

Support: None reported.

Address correspondence to Dorothy T. Horber, PhD, National Board of Osteopathic Medical Examiners, Continuous Professional Development, 101 W Elm St, Ste 100, Conshohocken, PA 19428-2003.

Email: [dhorber@nbome.org](mailto:dhorber@nbome.org)

Submitted July 25, 2018;  
revision received February 2, 2019;  
accepted March 15, 2019.

**Context:** As a proposed alternative to the traditional recertification examination, CATALYST is a longitudinal formative assessment platform created on cognitive learning principles. CATALYST was designed by the National Board of Osteopathic Medical Examiners to encourage more complex and durable practice-relevant learning and demonstration of ongoing competencies.

**Objective:** To investigate the value of the CATALYST platform using board diplomates' subjective feedback and comparison of performance on CATALYST questions with performance on board examinations.

**Methods:** Diplomates from 3 osteopathic specialty boards (the American Osteopathic Board of Internal Medicine, the American Osteopathic Board of Pediatrics, and the American Osteopathic Board of Obstetricians and Gynecologists) participated in this pilot study. Over the course of 16 weeks, participants were provided 2 questions per week via the CATALYST platform. An evaluation survey was emailed at the end of the study period to collect participants' feedback. Survey results and correlations of CATALYST performance with past or upcoming board examination scores were analyzed.

**Results:** A total of 196 diplomates completed the surveys, with 95% reporting that participation in the platform would help them stay current in their specialties and 91% reporting that participation would help them provide better care to their patients. For the AOBIM, a significant correlation was found between the number of CATALYST questions answered correctly and performance on the board examination ( $r=0.51$ ,  $P<.001$ ). The correlations found for the AOBP and AOBOG were not significant ( $r=0.197$ ,  $P=.296$ , and  $r=0.370$ ,  $P=.075$ , respectively).

**Conclusion:** The CATALYST platform could offer valuable contributions to the board recertification process and to patient safety. Further investigations are being conducted on a new user-friendly platform.

*J Am Osteopath Assoc.* 2020;120(3):190-200. Published online February 13, 2020.  
doi:10.7556/jaoa.2019.131

**Keywords:** board certification, maintenance of certification, osteopathic continuous certification

**I**n 2013, the American Osteopathic Association (AOA) implemented Osteopathic Continuous Certification (OCC), making osteopathic board certification an ongoing, lifelong process.<sup>1</sup> Through OCC, board-certified physicians demonstrate competency and current knowledge in their specialty areas. One of the components of

OCC, cognitive assessment, requires that time-limited certificate holders pass at least 1 psychometrically valid examination per certification cycle. Traditionally, such examinations have been proctored, high-stakes, secure, multiple-choice examinations, which involve considerable expense,<sup>2</sup> can require significant time away from practice, and have been criticized as lacking relevance to clinical practice. Thus, physicians have increasingly shown interest in alternatives to these traditional examinations.

Research has demonstrated that last-minute studying (cramming) immediately before testing creates the “illusion of mastery,”<sup>3(p16)</sup> the mistaken belief that something has been learned. For example, repeated re-reading can provide the illusion of mastery of underlying concepts and a consequent overestimation of one’s competence. This factor may contribute to the findings that physicians often are unaware of their knowledge gaps.<sup>4</sup>

Studies in cognitive learning have shown that more complex and durable learning can be accelerated in a number of ways, such as by asking individuals to retrieve what they learned previously, by providing immediate feedback, by spaced repetition of information on key concepts, and by waiting to restudy new material.<sup>3</sup> Learning and retention can also be enhanced by the practice of mixing one skill or topic with another and by varying practice to generalize the learning. For example, Moulton et al<sup>5</sup> demonstrated that residents retain and transfer surgical skills when they are taught in a distributed manner rather than in a single multihour event. Thus, the current requirement of a single examination every several years may not be the most effective way to assess current knowledge, demonstrate competency, or enhance continuous learning.

Considering this research alongside the opportunities presented by recent innovations in technology, such as mobile devices and web-based interactive platforms, a number of medical specialty certifying boards have implemented an alternative format for their recertification examinations. In 2014, the American Board of Anesthesiology (ABA) piloted the Maintenance of

Certification Assessment (MOCA) Minute,<sup>6,7</sup> an interactive learning and assessment tool to replace their traditional maintenance of certification (MOC) examination. The MOCA Minute, which is now the core of MOCA 2.0, the ABA’s redesigned MOC program, incorporates multiple-choice questions with immediate feedback, including the correct answer, a rationale for that answer, and links to additional learning resources. Diplomates are required to answer 30 questions per calendar quarter, and they may receive similar questions over time on some topics to gauge retention.<sup>6</sup> Diplomates are encouraged to answer questions over time (eg, a few questions per week), rather than in a single session each quarter, to promote continuous learning and knowledge retention.<sup>6</sup>

Research conducted by the ABA suggests that voluntary active participation in a program that incorporates frequent knowledge assessments in conjunction with targeted learning resources is associated with improved performance on a high-stakes certification examination and increased specialty knowledge.<sup>8</sup> Furthermore, those who participated in the ABA study<sup>8</sup> perceived the program to be valuable: 91% agreed or strongly agreed that participation in the program helped them provide better care for their patients, and 97% agreed or strongly agreed that the questions were useful learning tools.

In January 2019, the American Board of Pediatrics (ABP) initiated MOCA for Pediatrics (MOCA-Peds), an online, nonproctored assessment platform.<sup>9</sup> Launched as a pilot in January 2017, MOCA-Peds is an online platform that offers diplomates the option of participating in a continuous formative assessment program or taking the traditional proctored examination at a secure test center. In the continuous formative assessment program, 20 multiple-choice questions based on predetermined learning objectives are delivered quarterly and can be answered individually or as a group using a computer, tablet, or smartphone.<sup>10</sup> As with MOCA Minute, after answering a question, the correct answer, rationale, and supporting references are provided. The 4 lowest-scored quarters will be dropped

in each 5-year MOC cycle without negative consequences. Confidentiality is required, but diplomates are permitted to access resources to help answer questions.

Based on these studies, the National Board of Osteopathic Medical Examiners (NBOME), in conjunction with the AOA, the American Osteopathic Board of Internal Medicine (AOBIM), the American Osteopathic Board of Pediatrics (AOBP), and the American Osteopathic Board of Obstetricians and Gynecologists (AOBOG), began a 2-stage pilot study of a formative assessment program, CATALYST, for possible consideration as a replacement for these boards' periodic recertification examinations.

The goals of this initial pilot study were to:

1. Gather subjective data on participants' perceptions of CATALYST as an assessment platform.
2. Collect diplomates' feedback on the assessment process (convenience, relevance, and usefulness compared with traditional examinations).
- 3a. Stage 1: Evaluate the correlation between CATALYST data and performance on past traditional AOBP and AOBIM certification examinations.
- 3b. Stage 2: Evaluate the correlation between CATALYST data and performance on the March 2018 OB-GYN recertification examination (concurrent and predictive validity evidence).

## Methods

This 2-stage pilot study used a nonexperimental design with convenience samples from the AOBIM, AOBP, and AOBOG diplomate groups. Both stages received approval from the Institutional Review Board of the Center for the Advancement of Healthcare Education and Delivery.

## Participants

Participants were recruited from the 3 board groups, with AOBIM and AOBP participants contributing to the first stage of the study and AOBOG participants

contributing to the the second stage. Those invited to participate in stage 1 included 514 AOBIM diplomates who had passed the AOBIM initial certification examination in 2016 and 131 AOBP diplomates who had passed the AOBP initial certification examination within the previous 3 years. The stage 1 study was conducted concurrently beginning in April 2017.

Stage 2 invitees included 334 AOBOG diplomates from 3 categories: those who were eligible to take the recertification examination in 2018 (n=107) or in 2019 (n=130) or who were physician examiners (n=97). Stage 2, which began in October 2017, was organized around the AOBOG recertification examination scheduled for March 2018.

## Procedures

In both stages 1 and 2, those who expressed interest in participating in the pilot study were provided with access to the NBOME Learning Center, course registration information, and a sample item. Each Monday for 16 weeks, participants were notified electronically that the current week's questions were available. Questions were available 24 hours per day, 7 days per week, and could be answered using a smartphone, tablet, or computer.

## Stage 1

Stage 1 was described as an assessment for learning, which required participants to log on to the NBOME Learning Center's secure online portal to answer 2 specialty-specific questions each week for 16 weeks. After answering each question, the participant received feedback, including whether the selected answer was correct or incorrect, the correct answer, the rationale for it, and relevant reference material.

Participants were permitted 3 minutes to respond to each question. The AOBIM and AOBP did not require participants to answer questions on a strictly weekly basis; instead, they allowed participants to "catch up," provided that all questions through week 8 were answered by that time, and that questions for weeks 9 through 16 were answered by the end of week 16.

In addition, participants were required to complete a brief evaluation survey at project completion. As an incentive for completing all pilot requirements, participants received 20 category 1-B specialty-specific continuing medical education (CME) credits awarded by the AOA. No partial credit was provided if some of the 32 items or the survey were not completed.

### **Stage 2**

The AOBOG selected different parameters than the AOBIM and AOBP. Participants were required to answer each week's questions within 14 days or be discontinued from the pilot study and not receive the 20 specialty-specific AOA CME credits. Also, participants were permitted 4 minutes' response time. The rest of the processes for stage 2 were the same as those in stage 1.

### **Content**

#### *Stage 1*

The AOBIM and AOBP each contributed 32 specialty-specific questions as well as (the sample question). The content, written and approved by subject matter experts, was selected for alignment with the AOBIM and AOBP certification examination blueprints. The AOBP's questions were retired items selected from previous certification examinations; AOBIM's questions were newly written by AOBIM Board members. The NBOME developed the accompanying rationales and references for these questions.

#### *Stage 2*

The items used in stage 2 were selected by the AOBOG Board members, who also provided rationales and resources for the correct answer in the form of American College of Obstetrics and Gynecology Practice Bulletins. These items were previously used on recertification examinations and were aligned with the examination blueprint.

### **Evaluation Survey**

An evaluation survey was developed by NBOME staff and reviewed by the 3 boards. The survey contained

12 items, of which 9 were Likert type, 1 was forced choice, and 2 were open ended. The goal of the evaluation survey was to obtain participants' opinions on the CATALYST assessment format, process, and user experience. After week 16, the evaluation survey was distributed via email to all participants.

### **Analysis**

To satisfy goals 1 and 2, demographic and evaluation survey data were collected during both stages and summarized from everyone who completed the evaluation survey. To satisfy the third pilot goal, data from both stages were merged from the "engaged" participants (defined as those who completed at least 28 of the 32 items, 87.5%), matched by name to their respective board examination scores, and correlated with the board examination scores.

For stage 1, the AOBIM and AOBP provided certification data in the form of percentage correct on the examination; for stage 2, the AOBOG provided board recertification data in the form of raw scores. For each cohort, a Pearson product-moment correlation analysis was performed using the statistical package SAS 9.4 to determine if there was a statistically significant difference between participants' performance on CATALYST and on the traditional examinations.

Statistical significance was set at  $\alpha=.05$ . The correlational analyses were performed using data analysis from engaged participants to maximize the number of participants. The criterion for inclusion in survey data analysis was completion of the survey.

### **Results**

Of the 514 internal medicine diplomates invited to participate, 69 enrolled (13% response rate) and 42 completed all required aspects (61% completion rate). Of the 131 invited pediatric diplomates, 38 enrolled (29% response rate), and 32 completed all required aspects (84% response rate). Of the 334 obstetrician-gynecologist diplomates, 177 enrolled (53% response

rate), and 98 completed all required aspects (55% completion rate). A total of 52, 36, and 108 participants from the AOBIM, AOBP, and AOBOG, respectively, were included in the final analysis (**Table 1**).

The AOBIM participants were predominantly male and practiced in a full-time inpatient setting, whereas the AOBP participants were predominantly female in a full-time outpatient setting. Approximately two-thirds of the AOBOG participants were women, and the same proportion stated that their practice setting was a mix of inpatient and outpatient (**Table 1**).

Results of the evaluation survey for the Likert-type and forced-choice items are presented in **Table 2**. The responses to the survey were extremely positive from all 3 boards' participants. Among AOBIM and AOBP respondents, 100% agreed or strongly agreed that participation in CATALYST would help them stay current in their respective specialty and that they would recommend CATALYST to others in their specialty. Among AOBOG respondents, the percentages were 92% and

96%, respectively. Eighty-seven percent of the AOBIM respondents, 95% of the AOBP respondents, and 92% of AOBOG respondents agreed or strongly agreed that participation in CATALYST would help them take better care of their patients. In addition, almost all of the respondents (96% of internists, 100% of pediatricians, and 99% of obstetrician/gynecologists) agreed or strongly agreed that they would rather answer a fixed number of CATALYST questions periodically to help keep them current in their specialty rather than take the traditional high-stakes recertification examination administered in a test center.

The first open-ended question asked respondents, "What did you like about CATALYST?" Answers focused on the ease and convenience of responding, the opportunity to learn, and the relevance of the content to practicing physicians (**Figure 1**). The second open-ended question asked respondents, "What suggestions do you have for improving CATALYST?" A substantial number of the respondents in each

**Table 1.**  
**Gender, Practice Setting, and Participation in CATALYST-AOBIM, CATALYST-AOBP, and CATALYST-AOBOG<sup>a</sup>**

Characteristic	AOBIM (n=52)	AOBP (n=36)	AOBOG (n=108)
<b>Gender</b>			
Female	15 (29)	29 (81)	71 (66)
Male	37 (71)	7 (19)	35 (32)
Prefer not to answer	0	0	2 (2)
<b>Practice Setting</b>			
Inpatient only	30 (58)	2 (6)	4 (4)
Outpatient only	10 (19)	17 (47)	7 (6)
Mixed inpatient and outpatient	4 (8)	10 (28)	72 (67)
Academic and patient care	7 (13)	5 (14)	15 (14)
Administrative and patient care	1 (2)	1 (3)	7 (6)
Academic only	0	1 (3)	0
Administrative only	0	0	3 (3)

<sup>a</sup> CATALYST is a longitudinal formative assessment program. Data presented as No. (%).

**Abbreviations:** AOBIM, American Osteopathic Board of Internal Medicine; AOBOG, American Osteopathic Board of Obstetricians and Gynecologists; AOBP, American Osteopathic Board of Pediatrics.

**Table 2.**  
**Participants' Evaluation Survey Results From CATALYST<sup>a</sup>**

Evaluation Survey Item	AOBIM Participants (n=52)		AOBP Participants (n=36)		AOBOG Participants (n=108)	
	Strongly Agree	Agree	Strongly Agree	Agree	Strongly Agree	Agree
Registering for CATALYST was easy and quick.	36 (69)	16 (31)	25 (69)	11 (31)	84 (78)	20 (19)
Total	52 (100)		36 (100)		104 (96)	
The process for answering the questions was convenient.	29 (56)	18 (35)	22 (61)	11 (31)	78 (72)	26 (24)
Total	47 (93)		33 (92)		104 (96)	
The CATALYST format provided a good opportunity to learn/reinforce Internal Medicine/ Pediatrics/ Obstetrics & Gynecology material.	37 (71)	13 (25)	23 (64)	13 (36)	72 (67)	26 (24)
Total	50 (96)		36 (100)		98 (91)	
The rationales included with the answers to the CATALYST questions were effective learning tools.	17 (33)	27 (52)	23 (64)	11 (31)	63 (58)	30 (28)
Total	44 (85)		34 (95)		93 (86)	
I found the references included with the answers to the CATALYST questions to be useful.	28 (54)	14 (27)	15 (42)	12 (33)	66 (61)	28 (26)
Total	42 (81)		27 (75)		94 (87)	
I would rather answer a fixed number of CATALYST questions periodically to help me keep current in my Internal Medicine/ Pediatrics/ Obstetrics & Gynecology rather than take the traditional high-stakes certification/recertification examination administered in a test center.	42 (81)	8 (15)	31 (86)	5 (14)	101 (94)	6 (6)
Total	50 (96)		36 (100)		107 (99)	
Participation in the CATALYST program will help me provide better care to my patients.	30 (58)	15 (29)	19 (53)	15 (42)	73 (68)	26 (24)
Total	45 (87)		34 (95)		99 (92)	
Participation in the CATALYST program will help me stay current in Internal Medicine/ Pediatrics/ Obstetrics & Gynecology.	34 (65)	18 (35)	23 (64)	13 (36)	78 (72)	21 (19)
Total	52 (100)		36 (100)		99 (92)	
I would recommend CATALYST to others in my specialty.	34 (65)	18 (35)	25 (69)	11 (31)	79 (73)	25 (23)
Total	52 (100)		36 (100)		104 (96)	

(continued)

**Table 2 (continued).**  
**Participants' Evaluation Survey Results From CATALYST<sup>a</sup>**

Evaluation Survey Item	AOBIM Participants (n=52)		AOBP Participants (n=36)		AOBOG Participants (n=108)	
	Strongly Agree	Agree	Strongly Agree	Agree	Strongly Agree	Agree
<b>How did you access the questions?</b>						
Smartphone	24 (46)		14 (39)		44 (41)	
Tablet		5 (10)		1 (3)		2 (2)
Latop		17 (33)		17 (47)		33 (31)
Desktop		6 (12)		4 (11)		29 (27)

<sup>a</sup> CATALYST is a longitudinal formative assessment program. Data are presented as No. (%). Percentages may not total 100 because participants were not required to answer each question.

**Abbreviations:** AOBIM, American Osteopathic Board of Internal Medicine; AOBOG, American Osteopathic Board of Obstetricians and Gynecologists; AOBP, American Osteopathic Board of Pediatrics; CATALYST.

cohort—about one-third—had no improvement suggestions. Those who suggested improvements focused on streamlining the user interface, allowing more time to respond to questions, and providing more detailed rationales (Figure 2).

### Correlation

In stage 1, the correlation between performance on items administered through the CATALYST platform and scores on previously administered AOBIM and AOBP board certification examinations was analyzed. For the AOBIM, data were merged from the 53 engaged participants (those who responded to at least 87.5% of the CATALYST items) and AOBIM certification examination scores using participants' names. All of the 30 matched participants earned passing scores on the AOBIM certification examination. The Pearson product-moment correlation analysis showed a statistically significant relationship between the percentage correct on CATALYST items and the percentage correct on the AOBIM certification examination scores ( $r=0.51$ ,  $P<.001$ ).

For AOBP, again using participants' names, data were merged from the 37 engaged participants in

CATALYST with the AOBP certification examination scores administered between 2010 and 2017, inclusive, resulting in 30 matched participants. The Pearson product-moment correlation analysis between examination scores and number of correct CATALYST questions showed no statistically significant correlation ( $r=0.197$ ,  $P=.296$ ).

In stage 2, the relationship between CATALYST performance and an upcoming traditional AOBOG recertification examination was studied. Data were merged from 113 engaged CATALYST participants with the examination scores of the 24 CATALYST participants who took the March 2018 AOBOG recertification examination, and a Pearson product-moment correlation analysis was performed. The analysis shows a correlation between performance on CATALYST and the AOBOG recertification examination scores ( $r=0.370$ ,  $P=.075$ ). However, the sample size was small ( $n=24$ ), and the results were not statistically significant.

### Discussion

The outcomes of the evaluation survey provided substantial support for participants' satisfaction with the

- "Very convenient; allowed for quick review of pediatric topics. Overall a great program."
- "The material was clinically significant."
- "Convenient with high-yield topics."
- "It kept me up to date on pediatric topics, and I loved the explanations and references. This is so convenient and makes so much more sense to do this than the high-stakes examination."
- "I thought the format of CATALYST is an excellent idea. I enjoyed doing the questions each week and it was readily accessible. I think it is an excellent alternative to the traditional maintenance of certification examination."
- "I MUCH prefer the idea of regular questions and information to keep current in pediatrics as opposed to the stress and time commitment of studying and traveling for a high-stakes exam. This is a better way to learn, and would be much less interruption to my practice/income (no need to take off work for study or the test)."
- "Convenient and familiar testing style. Easy to commit to every week. I looked forward to the questions. Helped me to see which topics I need to review."
- "I like the fact that it made me look up information that I was unsure about. It also made me realize that some changes occurred in patient management that I was not aware of."
- "I liked that I could do it on my own time. It wasn't as pressured as the formal examination in the test center."
- "I really liked the CATALYST format. I definitely learned new things and reinforced things I already knew."
- "The format is definitely more relevant to the life of a practicing physician... iterative, ongoing, and incremental learning of topics that we deal with commonly and uncommonly. The traditional mode of testing really forces the learner to artificially, and therefore temporarily, learn a wide range of topics to pass a single examination, I must say that this is a very rational and practical solution for an important competency maintenance/development plan."

**Figure 1.**

Sample responses from American Osteopathic Board of Internal Medicine, American Osteopathic Board of Obstetricians and Gynecologists, and American Osteopathic Board of Pediatrics participants to the question, "What did you like about CATALYST?" CATALYST is a longitudinal formative assessment program.

relevance and convenience of CATALYST as an assessment vehicle. Importantly, 91% of the total group believed that participation in the CATALYST program would help them provide better care to their patients, which is the same percentage reported by the ABA.<sup>8</sup>

Based on the findings of learning theory research,<sup>3,5</sup> this pilot study employed principles that have been shown to accelerate learning, including using quizzes to ask learners to retrieve previous learning, providing immediate feedback, spacing learning, and interleaving different topics. The large majority of participants in all 3 cohorts found this method to be effective.

Ninety-four percent of the total group agreed or strongly agreed that the CATALYST format provided a good opportunity to learn or reinforce content in their specialties.

The data suggested great interest in pursuing a longitudinal formative assessment in place of the traditional high-stakes specialty board certification or recertification examination. The physicians in these pilots expressed an overwhelming preference (98% of the total group of 196 diplomates) for a format that permits responding to questions periodically to maintain specialty knowledge as a replacement for the traditional recertification examination.

- “Make the questions a bit more difficult.”
- “More questions per week (3 potentially).”
- “The format is a bit tedious and old-fashioned looking—not as user friendly or intuitive, which makes the process a bit more difficult. Streamlined would be better; simpler design.”
- “The explanations could be better: a little more thorough, and bullet points with space separating each point. It would be easier to read on cell phones.”
- “List the practice bulletin or article to read first then ask questions. I want to stay up to date and some questions are on topics I don’t deal with in practice, so knowing the article prior would be of great help.”
- “Too many clicks to get into questions. Maybe more questions (4) each week about a single topic.”
- “I would give a longer time frame to answer than 4 minutes. Difficult with interruptions in office, patient care, and on-call duties to complete in that time frame.”
- “Please give material BEFORE asking questions. It is disheartening and annoying to have gotten 60% of these questions incorrect because I hadn’t read the most up-to-date reference. That way you know I’ll have read the material AND tested my retention of the important ‘take-home’ points.”
- “Really liked...don’t have any suggestions. Would prefer to answer questions like this than sit for a high-pressure examination in a test center on a vacation day that is a far cry from being a vacation day.”
- “In the future, it would be prudent to have the questions more specific to ones’ specialty instead of just general OB/GYN.”

**Figure 2.**

Sample responses from American Osteopathic Board of Internal Medicine, American Osteopathic Board of Obstetricians and Gynecologists, and American Osteopathic Board of Pediatrics participants for improving CATALYST, a longitudinal formative assessment program.

The 2-stage plan of this study permitted information gained during stage 1 to be used in the planning of stage 2. It permitted variation and customization in board-specific parameters, specifically time permitted to respond and participation requirements. However, the data suggest a number of elements of the process that require further exploration. These elements include the amount of time permitted to respond to the questions (how many minutes), the response interval (quarterly, monthly), and the number of questions per response interval. While a number of participants in these pilots requested an unlimited response time per question, the response time limit encourages examinees to answer questions on their own, without use of external resources or other peers, and thus helps maintain integrity of the process. Whereas this pilot study employed a weekly response interval in an effort to gather data more quickly, a production setting might

use a quarterly basis to accommodate physicians’ busy schedules.

In addition, it will be necessary to investigate and determine parameters for required participation. Specifically, is it sufficient for diplomates to answer 80% of the questions, or should 100% be required? Regarding setting standards for recertification, what number of questions must be answered correctly? What other variables should be considered in making accurate recertification decisions? These questions will require careful deliberation from both clinical and psychometric experts.

Preliminary analysis showed a statistically significant correlation between performance on content delivered through CATALYST and performance on a traditional certification examination in internal medicine, suggesting some comparability in the assessment scores. The correlation was not significant in either the pediatrics

or the obstetrics-gynecology cohort. However, the small number of participants in these cohorts limits both the interpretation and generalizability of these initial results and indicates that further research is warranted. For example, the Spearman correlation coefficient can be used in future studies to examine the linear relationship between the ranking or relative positions of CATALYST and certification scores.

In this pilot study, 3 common sources of validity evidence were considered: examination content, statistical relationships with participants' scores, and theoretical models of the underlying construct being measured. The items used in this study were developed by content experts, and the alignment of this content with the AOBIM, AOBP, and AOBOG examination blueprints provide contributions to content validity. Regarding the statistical relationship with participants' scores, the results were not straightforward, likely a function of the small sample sizes. This type of validity evidence can be further enhanced by collecting additional data to obtain a larger sample for AOBP and AOBOG to increase the reliability of the results. However, considering physician concerns about the validity and practice relevance of traditional multiple-choice examinations, modest score correlations may be desirable as CATALYST-delivered test content may indeed be measuring something different. Concerning the theoretical models of the construct being measured, nearly all participants believed that CATALYST's formative approach will enable them to better care for their patients and stay current in their specialties.

Another limitation is the study duration. Considering that CATALYST is intended as a longitudinal assessment, with learning likely to develop over longer periods than the 16 weeks of this pilot study, conclusions are limited. Furthermore, while the present study reports data on participants' gender and primary practice setting, it may be of interest to consider other variables, such as diplomate age. One might postulate that younger physicians would be more comfortable with accessing formative test content via technology such as smartphones or

tablets, but that this cohort might also have higher expectations about the technological interface. Based on survey responses that suggested improving the user interface, the NBOME is actively developing an alternative platform for future CATALYST implementations to make the tool more user friendly and provide additional functionality such as adaptive learning, which is expected to become increasingly important for medical education in the future.<sup>11</sup> The new platform will also permit customization for different audiences and learning opportunities. For example, the NBOME has used the CATALYST platform in other formative assessment studies, such as the Comprehensive Osteopathic Medical Self-Assessment Examination (COMSAE)-CATALYST beta study, in which osteopathic medical students were provided access to COMSAE test items<sup>12,13</sup> presented in a formative assessment format. Outcomes of the COMSAE CATALYST beta study supported the results of the present study, indicating both substantial interest in learning through formative assessment and calls for a more modernized user-friendly platform.

## Conclusion

The outcomes of this study provide strong support for participants' satisfaction with the relevance and convenience of CATALYST as an assessment platform, as well as for the potential of this assessment to help them stay current in their specialty. The data suggest participants' overwhelming preference for this continuous formative assessment over the current traditional examination format. Additional research is planned to investigate and clarify various parameters for optimal learning.

## Acknowledgments

We thank the following people for their assistance in planning and implementing the CATALYST pilots: Ingrid Brown, DO, Chair, AOBIM; Paul Whitham, DO, Examination Chair, AOBOG; Mary Cameron, MA, AOA Senior Certification Director; Danny Hart, AOA Certification Director; Angi Beranek, former AOA Certification Director; the NBOME CATALYST Team, in particular, Sandra Waters, MEM, NBOME, Vice President for Collaborative Assessment and Initiatives; and Melissa Turner, MS, Associate Vice President for Strategy, Quality, and Communications.

## Author Contributions

All authors provided substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; all authors drafted the article or revised it critically for important intellectual content; all authors gave final approval of the version of the article to be published; and all authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## References

1. Scheinthal S, Gross C. Osteopathic Continuous Certification: it's here— are you prepared? *J Am Osteopath Assoc.* 2013;113(6):479-483.
2. Drolet BC, Tandon VJ. Fees for certification and finances of medical specialty boards. *JAMA.* 2017;318(5):477-479. doi:10.1001/jama.2017.7464
3. Brown PC, Roediger HL III, McDaniel MA. *Make It Stick: The Science of Successful Learning.* Cambridge MA: Harvard University Press; 2014.
4. Davis DA, Mazmanian PE, Fordis M, Harrison RV, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA.* 2006;296(9):1094-1102. doi:10.1001/jama.296.9.1094
5. Moulton CA, Dubrowski A, MacRae H, Graham B, Grober E, Reznick R. Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. *Ann Surg.* 2006;244(3):400-409.
6. Part 3: MOCA Minute. American Board of Anesthesiology website. <http://www.theaba.org/MOCA/MOCA-Minute>. Accessed January 23, 2019.
7. MOCA timeline. American Board of Anesthesiology website. <http://www.theaba.org/MOCA/MOCA-Timeline>. Accessed October 4, 2019.
8. Sun H, Zhou Y, Culley DJ, Lien CA, Harman AE, Warner DO. Association between participation in an intensive longitudinal assessment program and performance on a cognitive examination in the maintenance of certification in anesthesiology program. *Anesthesiology.* 2016;125(5):1046-1055. doi:10.1097/ALN.00000000000001301
9. MOCA-Peds. The American Board of Pediatrics website. <https://www.abp.org/mocapeds> Accessed January 23, 2019.
10. 2018 MOCA-Peds learning objectives and guidelines. American Board of Pediatrics website. <https://www.abp.org/mocapeds/2018-learning-objectives>. Accessed January 23, 2019.
11. Harden RM, Lilley P, McLaughlin J. Forty years of medical education through the eyes of Medical Teacher: from chrysalis to butterfly. *Med Teach.* 2018;40(4):328-330. doi:10.1080/0142159X.2018.1438593
12. COMSAE: a self-assessment exam for the COMLEX-USA exam series. National Board of Osteopathic Medical Examiners website. <https://www.nbome.org/exams-assessments/comsae>. Accessed January 23, 2019.
13. Horber D, Waters S. CATALYST: Transforming physicians' assessment into learning. Paper presented at the American Board of Medical Specialties Conference; Chicago, IL; September 2019. [https://static1.squarespace.com/static/5581c84de4b08614acb89b1ff/5d8a65a981eb3d15b7df52b8/1569351082961/Horber\\_Waters\\_How+Catalog+Can+Transform+Physicians+Assessment+into+Learning.pdf](https://static1.squarespace.com/static/5581c84de4b08614acb89b1ff/5d8a65a981eb3d15b7df52b8/1569351082961/Horber_Waters_How+Catalog+Can+Transform+Physicians+Assessment+into+Learning.pdf)

© 2020 American Osteopathic Association