

ORIGINAL ARTICLE

Knowledge Self-Assessment Engagement and Family Medicine Board Examination Outcomes

Peter M. Wingrove, MD^a; Andrew Bazemore, MD, MPH^{b,c}; Ting Wang, PhD^c; Keith Stelter, MD^{c,d}; David W. Price, MD^{c,e}

AUTHOR AFFILIATIONS:

^aDepartment of Radiology, University of Pittsburgh Medical Center, Pittsburgh, PA

^bCenter for Professionalism and Value in Health Care, Washington, DC

^cAmerican Board of Family Medicine, Lexington, KY

^dMayo Clinic, Rochester, MN

^eUniversity of Colorado Anschutz School of Medicine, Aurora, CO

CORRESPONDING AUTHOR:

David W. Price, American Board of Family Medicine, Lexington, KY,
dprice@theabfm.org

HOW TO CITE: Wingrove PM, Bazemore A, Wang T, Stelter K, Price DW. Knowledge Self-Assessment Engagement and Family Medicine Board Examination Outcomes. *Fam Med.* 2025;57(7):500–507. doi: [10.22454/FamMed.2025.200510](https://doi.org/10.22454/FamMed.2025.200510)

PUBLISHED: 13 June 2025

KEYWORDS: engagement, family physicians, formative assessment, summative assessment

© Society of Teachers of Family Medicine

ABSTRACT

Background and Objectives: Evidence on the relationship between formative assessment engagement and summative assessment outcomes in practicing physicians is sparse. We evaluated the relationship between engagement in the American Board of Family Medicine (ABFM) formative Continuous Knowledge Self-Assessment (CKSA) and performance on high-stakes summative assessments.

Methods: This retrospective cohort study included 24,926 ABFM diplomates who completed CKSA modules and summative assessments between 2017 and 2023. We analyzed CKSA engagement metrics—such as the number of quarters completed, time of completion, and self-reported confidence—against performance on summative assessments, measured by z scores. Multivariable regression models controlled for demographic factors and prior assessment performance.

Results: The overall cohort summative assessment pass rate during the study period was 90.3%. Greater CKSA engagement was strongly associated with higher summative assessment performance. Diplomates who completed all four CKSA quarters had significantly higher summative assessment z scores than those completing fewer quarters ($P < .001$). Early CKSA completion and spending more time on low-confidence questions were also positively correlated with both CKSA and summative assessment scores ($P < .001$). These effects were observed across different levels of prior exam performance.

Conclusions: Engagement in formative assessments like CKSA, particularly early and consistent participation and reviewing incorrect or low-confidence questions, is linked to better outcomes on high-stakes assessments. Future research should explore the mechanisms underlying these associations and consider developing an index of engagement to identify physicians at risk of poor performance. Incorporating structured, longitudinal self-assessments like CKSA into certification requirements could enhance continuous learning and improve summative exam readiness.

INTRODUCTION

Some physicians question the importance and role of high-stakes summative assessments (assessments of learning with comparisons to peers and minimal performance standards required to achieve or maintain a credential) after graduate training.^{1–3} They argue that continuing medical education (CME) or formative assessment (feedback without minimal performance standards designed to guide self-directed learning) is sufficient.⁴ While some CME formats are associated with changes in clinical practice, known limitations exist to relying on CME alone for maintenance of knowledge. These include CME selection based on interest rather than data-defined gaps⁵ and decay of knowledge and quality of care over time.^{6,7} Amer-

ican Board of Family Medicine (ABFM) certification signifies that family physicians possess knowledge across the breadth of the specialty, even though some family physicians may narrow their scope of practice over time. Thus, ABFM, as well as all American Board of Medical Specialties (ABMS) boards continue to incorporate summative knowledge assessments as part of the continuing certification process.⁸

Studies on the relationship between formative and summative assessments in trainees have had mixed results. Some studies found no correlation,^{9,10} while other studies found a positive association between serial formative assessments and summative assessment performance.^{11–13} However, evidence on the assessments' relationship in practicing health care

professionals is sparse.¹⁴

Research has shown that assessment consequences (stakes) impact performance and learning.^{15–17} Previous work has shown that physicians perform better, albeit with lower confidence, on higher-stakes examinations compared with lower-stakes self-assessment. This evidence suggests that higher-stakes examinations may be a better reflection of physicians' fund of knowledge.¹⁴

Physicians may concentrate less or put less effort into low-stakes formative self-assessments, given the lack of consequences for low performance. The degree of engagement in formative assessment may influence the association between formative and summative assessment performance.¹⁸ Engagement can be categorized into behavioral, cognitive, and affective dimensions.^{19,20} Behavioral engagement is observable and includes attendance, persistence, paying attention, asking questions, and avoiding disruptions. Cognitive engagement is deeper, including efforts to process and understand concepts and retrieve previous knowledge on which to build. Affective engagement includes emotional reactions, general interest, attitude, beliefs, values, sense of psychological safety in the learning environment, and social aspects of learning with peers and teachers.

The degree to which engagement in formative assessments affects the association between formative and summative assessments in practicing health care professionals remains underexplored. In this retrospective analysis, we examined the relationship between engagement in the ABFM Continuous Knowledge Self-Assessment (CKSA), a popular longitudinal online low-stakes formative assessment, and performance on ABFM summative assessments.

METHODS

CKSA

The CKSA¹² is a low-stakes, longitudinal assessment that presents 25 online multiple-choice questions each quarter. Participants rate their confidence in each answer on a six-point scale and receive immediate feedback, including answer accuracy, in-depth critiques, and pertinent references. The assessment has no time limit, and external resource use is allowed. More than 30,000 physicians complete a CKSA each quarter. In this study, we used data from Diplomate completions of CKSA modules from 2017 to 2023.

Dependent Variables

The primary dependent variable was each Diplomate's score on their first attempt at the high-stakes (summative) ABFM knowledge assessment from 2018 to 2023. The score was derived from either the periodic, 1-day, secure ABFM knowledge assessment (Family Medicine Certification Examination [FMCE]) or the newer quarterly Family Medicine Continuous Longitudinal Assessment (FMCLA)¹² in use since 2019. To quantify performance of each Diplomate relative to their peers and minimize confounding from Diplomates' ability to choose the type of summative assessment, we standardized each test score as a z score based on the distribution of scores for a given

year and type of assessment (FMCE vs FMCLA). We used the continuous z score distribution for all Diplomates completing a summative exam in the study period to create three terciles of current performance (low, medium, and high). A secondary dependent variable was the score on CKSA itself because we hypothesized that some CKSA behaviors may influence the CKSA score while having no effect (or even the opposite effect) when the Diplomate eventually took the summative assessment.

Independent Variables

We selected four measures of behavioral engagement and three measures of cognitive engagement as independent variables in our analysis. The first behavioral variable included was amount of participation, defined as the number of CKSA quarters (0–4) completed in the year prior to FMCE or FMCLA. Timing of engagement, the second variable, included the percentage of questions completed during late hours (10 PM–6 AM the next day) or on weekends (10 PM Friday–6 AM Monday). The third variable, expediency of completion (a proxy of prioritizing participation, as opposed to waiting until the last minute to complete the activity), was measured as the percentage of questions completed within the first month of the CKSA quarter. The fourth variable, concentrated participation (as opposed to distributing questions across the CKSA quarter), was assessed by dividing Diplomates into those more versus less likely than average to complete all quarterly questions in a single session.

Cognitive engagement was assessed by median time spent on questions (stratified by confidence) and median time spent reviewing critiques (stratified by correct vs incorrect answers). Most often, Diplomates answered questions in less than 5 minutes. We considered a calculated answer time greater than 10 minutes as a break in engagement; those completion times were recorded as “unknown” and were not used in calculating the median time. After looking at the frequency distributions, we excluded critique review times greater than 180 seconds because we assumed this reflected a Diplomate taking a break or interspersing CKSA with another activity.

We collapsed the six-point CKSA confidence scale into three categories: “extremely confident” and “very confident” as “high confidence,” “pretty confident” and “moderately confident” as a reference category, and “not at all confident” and “slightly confident” as “low confidence.” Less than 1% of Diplomates had missing values for high, medium, or low confidence median answer time. We replaced these missing values with the mean value among all Diplomates so as not to bias the regression results.

We included self-assessment accuracy as a proxy for metacognition. We calculated accuracy identifying strengths as the percentage of times that the Diplomate reported high confidence minus the percentage of times they reported low confidence in correctly answered questions. We calculated accuracy identifying weaknesses as the percentage of low confidence minus the percentage of high confidence on incorrectly answered questions.

Covariates

Participant ability, as measured by performance on previous knowledge assessments, is associated with engagement in learning²¹ and with performance on summative knowledge assessments.^{22,23} We therefore included performance on the most recent prior ABFM summative examination (either certification or recertification examination) as a covariate. We controlled for gender, degree (MD/DO), active provision of patient care, location of training (domestic vs international medical graduate), and age (as a z score relative to other Diplomates being examined during a given year and with a given format [FMCE/FMCLA]). This data was derived from Diplomate recertification surveys from 2018–2023, required as part of registration for FMCE or FMCLA.

Statistical Analysis

To standardize scores across years and between examination options, we transformed FMCLA and FMCE scores into z scores. We then created terciles so that we could compare Diplomates with the lowest exam performance and the best performance with the middle third. This approach allowed us to compare CKSA performance (percentage of correct questions) with examination performance.

We used χ^2 tests to assess the relationship between demographic variables and CKSA completion. We used t tests to examine how patterns of engagement with CKSA varied for the Diplomates in the highest versus lowest terciles of performance on the recertification exam relative to Diplomates with average performance. We performed multivariable regression to assess the effect of CKSA engagement patterns with performance on both CKSA and the subsequent summative exam. To help ensure that observed differences were related to CKSA, we controlled for performance on the previous ABFM exam and demographic characteristics. To avoid problems with circular logic, variables contingent on whether a *particular* CKSA question was correct (ie, time reviewing correct and incorrect answers as well as both variables relating correctness to confidence) were omitted from the regression on CKSA score.

As part of routine internal program evaluation, ethical approval was granted by the American Academy of Family Physicians Institutional Review Board.

RESULTS

This analysis included 24,926 Diplomates who completed a recertification exam between 2018 and 2023 (Table 1). Of these, 6,275 (25.2%) completed at least two quarters of CKSA and were therefore included in subsequent analyses of CKSA metrics.

The overall cohort summative assessment pass rate during the study period was 90.3%. Figure 1 shows the results of a primary regression examining summative assessment performance (as a z score relative to other Diplomates who took that specific exam) as a function of CKSA completion. Regression coefficients are reported for all Diplomates and three separate regressions by tercile of performance on the prior summative assessment. After controlling for demographic variables as well as prior FMCLA/FMCE performance, we found a statistically

significant association between completing three out of four ($P<.001$) or four out of four ($P<.001$) CKSA quarters relative to not engaging with CKSA. Highly statistically significant associations between exam performance and completing four out of four CKSA quarters persisted for all three prior performance tercile subgroups. Additionally, we found a clear benefit to completing three out of four quarters for those in the middle tercile ($P<.01$).

Of the Diplomates scoring in the highest tercile of exam performance, 29.1% completed at least two quarters of CKSA vs 21.3% in the lowest tercile (Table 2). Compared with the middle tercile, those scoring in the lowest tercile were more likely to have answered questions between 10 PM and 6 AM ($P<.001$), spent more time per question on questions with medium or high self-reported confidence ($P<.001$ for both), spent less time reviewing incorrect questions ($P<.001$), and self-reported confidence less accurately on questions they answered correctly ($P<.001$). Compared with those in the middle and lower terciles, those in the highest tercile completed a higher percentage of questions during the first month of the cycle ($P<.001$), spent less time on questions with medium or high self-reported confidence ($P<.001$ for both), spent more time reviewing both correct and incorrect questions ($P<.01$ and $P<.001$, respectively), and were more accurate self-reporting confidence on correct questions ($P<.001$). We found no differences in answering questions on weekends, concentrated participation (answering all questions in one session instead of across the quarter), or in accuracy identifying weaknesses.

Figure 2 shows the coefficients for two separate regressions that consider the relationship of (formative) CKSA metrics with CKSA performance and later performance on the (summative) recertification exam. Prior exam score is a strong and statistically significant predictor of performance on both formative and summative assessments ($P<.001$ for both). Even after controlling for demographic confounders and prior score, we found an association between completing all four quarters of CKSA and higher performance on the formative and summative assessments relative to those who completed only two out of four quarters. For those completing three out of four quarters, we found a statistically significant positive effect on CKSA score ($P<.001$), but no corresponding effect on the later summative assessment.

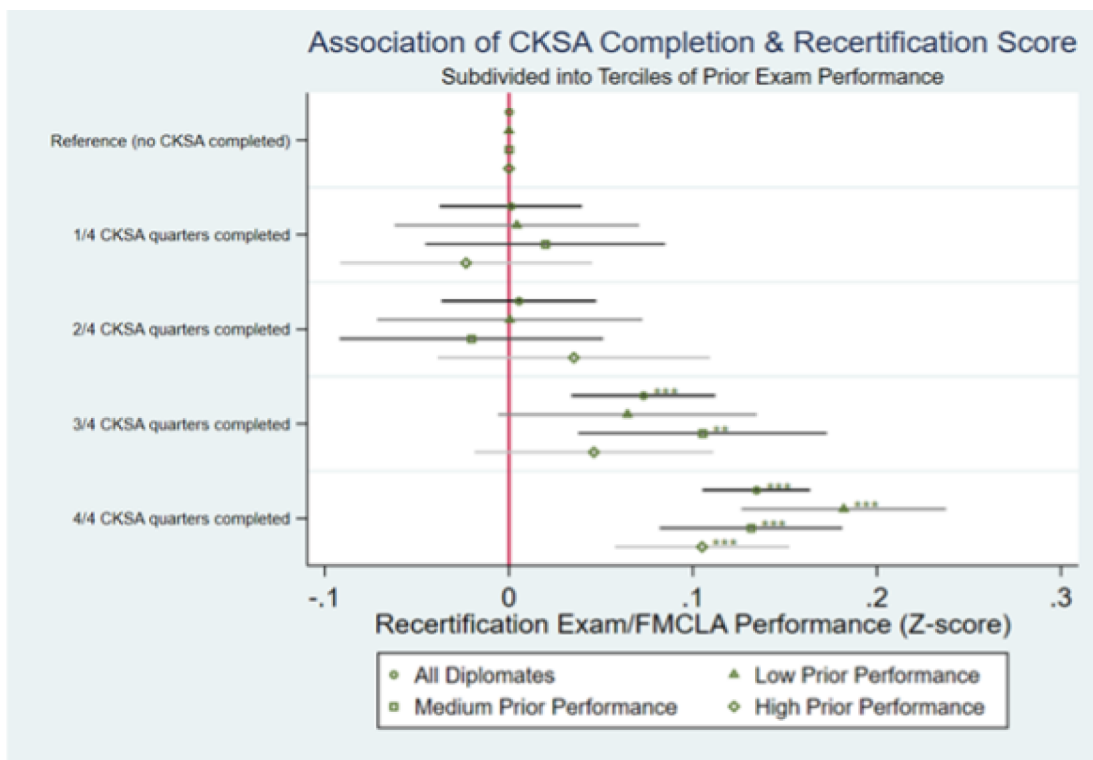
Completing CKSA questions late at night or in the early morning (between 10 PM and 6 AM) had a statistically significant negative effect on the CKSA score itself but did not reach statistical significance for a deleterious effect on recertification exam performance. We found no apparent harm/benefit to completing a higher percentage of questions on the weekend (vs a weekday) or concentrating CKSA completion into a lower number of sessions. However, we found a strong association between consistently completing CKSA questions during the first month of a quarterly cycle and summative exam performance ($P<.001$).

TABLE 1. Study Cohort Demographics, Overall and by Quarters of CKSA Completion

	Overall cohort	Quarters of CKSA Completed in Prior Year					P value *
		0	1	2	3	4	
Number of Diplomates	24,926	16,954	1,697	1,410	1,656	3,209	
Age (in years), mean (SD)	51.2 (9.2)	51.6 (9.2)	49.5 (8.7)	49.8 (8.8)	50.8 (9.4)	50.6 (9.3)	<.001
Female, %	43.38	41.39	45.73	48.87	47.04	48.33	<.001
Degree, % DO	9.83	9.71	9.49	9.79	10.02	10.53	.668
International medical graduate, %	23.94	23.49	30.29	26.95	26.03	20.57	<.001
Currently providing patient care, %	85.95	85.67	84.74	87.52	79.11	90.90	<.001

*Reported P values are for χ^2 tests except for age, which is a one-way analysis of variance.

Abbreviations: CKSA, Continuous Knowledge Self-Assessment; DO, Doctor of Osteopathy; FMCLA, Family Medicine Continuous Longitudinal Assessment; FMCE, Family Medicine Certification Examination; SD, standard deviation

FIGURE 1. Association of Continuous Knowledge Self-Assessment Completion With Recertification Exam Score

P values shown alongside markers:

* $P < .05$

** $P < .01$

*** $P < .001$

Regressions were controlled for score on prior exam, age (as a z score), gender, type of degree (MD/DO), international medical graduate status, and whether a diplomate reported providing patient care.

Abbreviations: CKSA, Continuous Knowledge Self-Assessment; FMCLA, Family Medicine Certification Longitudinal Assessment

Spending more time on questions with self-reported low confidence was positively associated with CKSA score ($P < .001$) and recertification exam score ($P < .001$). Spending more time on questions with medium confidence was associated with higher CKSA score ($P < .001$) but not better performance on the summative assessment. We observed an association between more review time after incorrect questions and higher summative assessment scores ($P < .001$) but did not see a parallel effect with questions answered correctly. After controlling for CKSA demographic variables and prior score, we found significant

independent associations between accuracy self-identifying strengths/weaknesses and subsequent performance on summative assessment ($P < .001$).

To verify the robustness of our results among more homogenous pools of test-takers, we ran three further regressions that subdivided Diplomates by terciles of prior performance (Figure 3). The association between completing four out of four CKSA quarters and higher summative performance appeared strongest for those with worse performance on the prior exam ($P < .001$ for the lowest tercile,

TABLE 2. Tercile of Most Recent Exam Performance Among Those Who Completed at Least Two CKSA Quarters in the Year Prior to the Exam

	Overall	Lowest tercile	Middle tercile	Highest tercile
Number of Diplomates	24,926	8,321	8,313	8,292
Number of diplomates with ≥ 2 quarters CKSA completed, (% of Diplomates)	6,275 (25.2)	1,775 (21.3)	2,086 (25.1)	2,414 (29.1)
z score current exam	0.17	−1.05***	0.03	1.20***
z score prior exam	0.41	−0.186***	0.262	.988***
CKSA questions correct, %	62.2	55.3***	61.2	68.2***
Answered at night (10 PM–6 AM), %	12.4	14.2***	11.9	11.7
Answered on weekend, %	33.3	32.8	32.9	34.1
Concentrated questions answered in one session, % of Diplomates	56.4	56.9	55.3	57.0
Questions completed during the first month, %	48.1	44.9	46.4	51.9***
Median time on low-confidence questions, seconds	67.9	70.1	67.1	67.0
Median time on medium-confidence questions, seconds	54.9	60.0***	54.6	51.4***
Median time on high-confidence question, seconds	46.7	53.1***	46.8	42.3***
Median time reviewing correct questions, seconds	24.6	23.0	24.2	26.2**
Median time reviewing incorrect questions, seconds	36.0	33.1***	35.5	38.5***
Accuracy identifying strengths, %	−15.8	−23.3***	−15.8	−10.4***
Accuracy identifying weaknesses, %	40.9	41.1	40.2	41.4

The overall cohort summative assessment pass rate during the study period was 90.3%.

Note: *t* tests are relative to the middle tercile except “Spread quarterly CKSA questions across multiple days,” which is an overall χ^2 test for the distribution of Diplomates in this group across the three terciles.

P Values:

* $P < .05$

** $P < .01$

*** $P < .001$

Abbreviation: CKSA, Continuous Knowledge Self-Assessment

$P < .01$ for the middle tercile, and no statistical significance for the highest tercile). We found a particularly strong association between consistently completing CKSA during the first month of a cycle and performance for those in the highest tercile of prior performance ($P < .001$), though this association held for the other two subgroups as well ($P < .05$ for both). Regardless of prior performance, all CKSA users appear to benefit from spending more time on low-confidence questions as well as more time reviewing incorrect questions. Finally, we continued to observe the same association between higher accuracy self-identifying strengths/weaknesses and summative exam performance for all terciles ($P < .001$ for both metrics across all subgroups).

DISCUSSION

Consistent completion of (formative) CKSA is associated with higher performance on subsequent summative assessments. Specific types of behavioral and cognitive engagement during formative assessment may contribute to this effect.

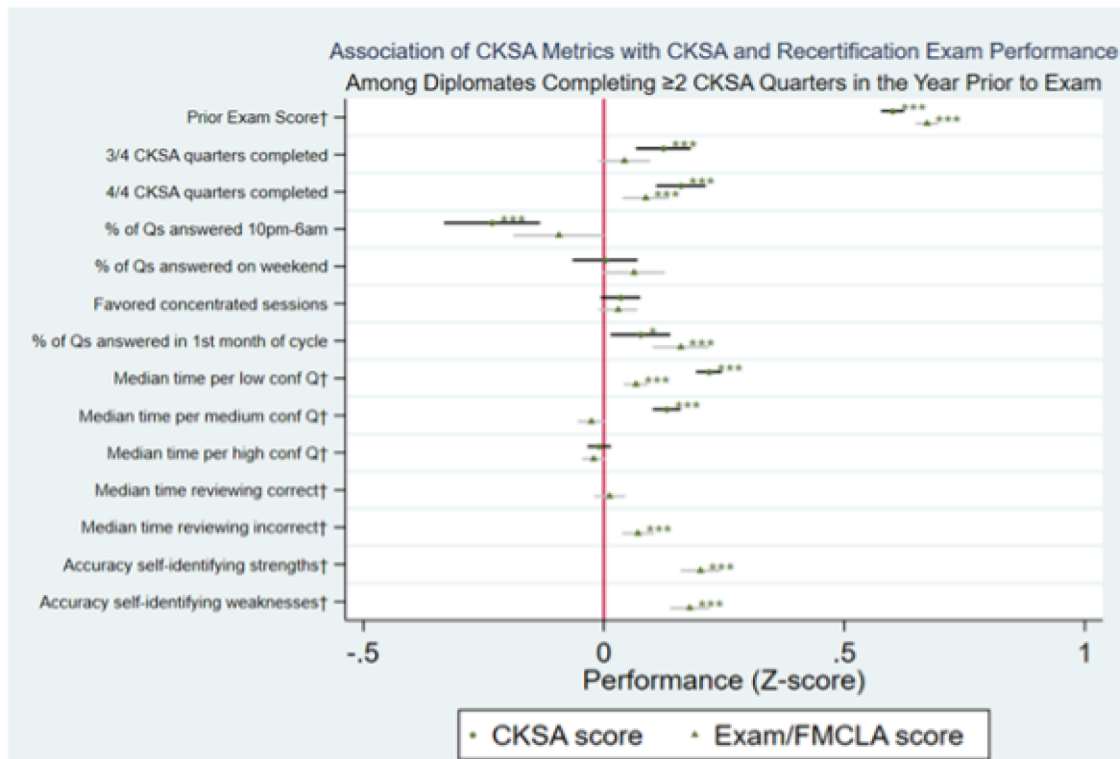
If schedules allow, physicians should consider completing self-assessments early in a quarter. Doing so may be a proxy for enthusiasm about, or at least prioritization of, the task, which may lead to more engagement compared with rushing to complete the task by the deadline. Physicians should devote sufficient time to questions where they have low confidence in their answer. They should prioritize incorrect questions, even if it means de-emphasizing review of correct answers. They

should avoid completing questions late at night or early in the morning, when fatigue (perhaps after a full day of other responsibilities) may make fully processing and reflecting on questions and remembering what is learned more difficult. This advice may be particularly important for physicians who scored in the lowest tercile on the previous summative examination.

Though further research is needed to clarify the mechanism, we observed that a higher ability self-identifying strengths and weaknesses during formative assessment correlated with better performance on subsequent summative assessment. This trend held across subgroups regardless of initial differences in knowledge/test-taking ability reflected in performance on the prior exam. This ability to identify strengths and weaknesses may represent a difficult-to-measure metacognitive skill. This skill may spur improvement by highlighting discrepancies between what participants think they know versus what they actually do know.²⁴

The choice to complete questions on weekends or in a single day (as opposed to distributing effort across multiple days) did not appear to affect subsequent summative examination performance. All things considered, our findings appeared to suggest that physicians should do parts of CKSAs when they are relatively nonfatigued and can focus on learning, rather than trying to rush to meet deadlines or while attending to other competing demands.

FIGURE 2. Regressions That Consider the Relationship of Continuous Knowledge Self-Assessment (CKSA) Metrics With CKSA Performance and Later Performance on the Recertification Exam



†Denotes variable was transformed into z score.

P values shown alongside markers:

* $P < .05$

** $P < .01$

*** $P < .001$

Regressions were controlled for age (as a z score), gender, type of degree (MD/DO), international medical graduate status, and whether a diplomate reported providing patient care.

Abbreviations: CKSA, Continuous Knowledge Self-Assessment; FMCLA, Family Medicine Certification Longitudinal Assessment

Strengths

Strengths of our study included a large, diverse sample of 24,926 Diplomates, enhancing the generalizability of the findings across a broad spectrum of practicing family physicians. The use of standardized z scores for performance comparison provided a robust method for evaluating the impact of CKSA engagement on summative assessment outcomes. The study's comprehensive analysis of engagement metrics, such as timing, consistency, and confidence levels, offers detailed insights into how specific behaviors may influence exam performance. We controlled for a range of demographic and prior performance variables, minimizing several potential confounders that could impact the relationship between formative and summative assessment performance.

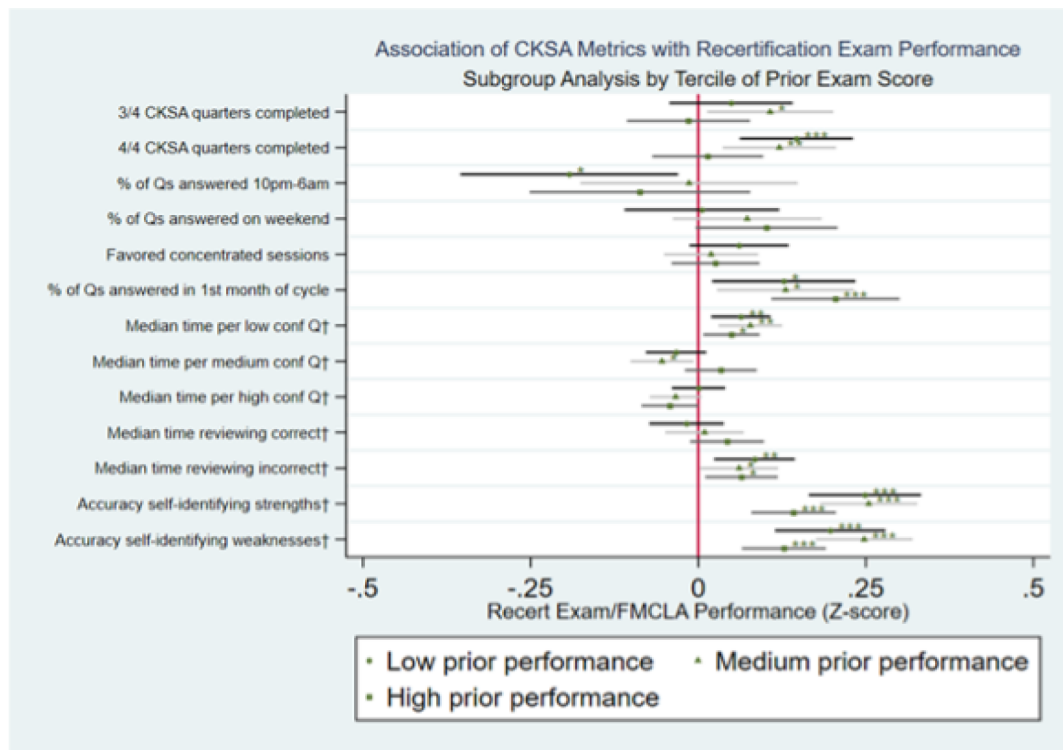
Limitations

This study had several limitations. As a retrospective analysis, one can infer association but not causality. Participants who engage more consistently with CKSA might inherently differ from those who do not; therefore, results cannot be extrapolated to individuals who do not engage in such activities. While we examined several measures of behavioral and cognitive

engagement, other factors that would influence our results may be missing. For example, we could not examine the correlation between flexibility or demanding work schedules and time of day of CKSA completion. We did not include measures of affective engagement, which would have required participant interviews or well-structured surveys. While we controlled for several demographic and performance-related variables, unmeasured confounders could still have influenced the results. Reliance on self-reported confidence levels and engagement metrics could introduce subjective bias. Participants were limited to physicians who completed training; findings may vary in undergraduate or graduate trainees. Finally, our findings are specific to ABFM and may not be generalizable to other specialties, professions, or certification processes. Further prospective studies are needed to validate these findings in other populations and to explore the mechanisms underlying the observed associations.

CONCLUSIONS

This study examined the relationship between behavioral and cognitive engagement in a low-stakes, longitudinal formative knowledge assessment and performance on high-stakes sum-

FIGURE 3. Continuous Knowledge Self-Assessment Engagement and Subsequent Examination Performance, by Tertiles

†Denotes variable was transformed into z score.

P values shown alongside markers:

* $P < .05$

** $P < .01$

*** $P < .001$

Regressions were controlled for score on prior exam, age (as a z score), gender, type of degree (MD/DO), international medical graduate status, and whether a diplomate reported providing patient care.

Abbreviations: CKSA, Continuous Knowledge Self-Assessment; FMCLA, Family Medicine Certification Longitudinal Assessment

mative assessments among practicing physicians. Physicians could use findings from this analysis to potentially enhance their performance on summative examinations. Future work should prospectively examine the relationship of these formative assessment engagement factors in different cohorts of physicians (eg, by years in practice, practice type, or arrangement) with summative examination performance. Additional studies could then examine associations of these relationships with performance in practice. Participant interviews with qualitative analyses could elucidate factors including how competing demands and after-hours administrative work (“pajama time”) affect engagement. If the relationships hold, medical specialty boards and others involved in formative knowledge assessment could use these findings to develop an index of engagement. Such an index could cross-sectionally or longitudinally help identify individuals at risk of poorer examination performance (eg, those who have failed or minimally exceeded the passing threshold on previous summative assessments). Educators could use some of these strategies to coach their learners on optimizing formative assessment. Consistent with Standards for Continuing Certification, ABMS boards could reach out to potentially at-risk Diplomates to “provide [them] with opportunities to address performance

or participation deficits prior to the loss of a certificate.”⁸ This study further supports the value of structured, lifelong learning and self-assessment activities such as CKSA as part of a program of continuous specialty board certification.

FUNDING

This work was internally funded by the American Board of Family Medicine.

ACKNOWLEDGMENTS

The authors thank Zachary J Morgan, MS, for assistance with data acquisition for this project.

REFERENCES

1. Watchter R. The ABIM controversy: a brief history of board certification and MOC. *The Health Care Blog*. 2015. <https://thehealthcareblog.com/blog/2015/07/02/the-abim-controversy-a-brief-history-of-board-certification-and-moc>.
2. ACEP Now. AMA opposes mandatory American Board of Medical Specialties' recertification exams. 2016. <https://www.acepnow.com/article/american-medical-association-opposes-mandatory-american-board-medical-specialties-recertification-exams>.

3. Rosner MH. Maintenance of certification: framing the dialogue. *Clin J Am Soc Nephrol*. 2018;13(1):161–163.
4. Teirstein PS. Boarded to death—why maintenance of certification is bad for doctors and patients. *N Engl J Med*. 2015;372(2):106–108.
5. Cook DA, Price DW, Wittich CM, West CP, Blachman MJ. Factors Influencing physicians' selection of continuous professional development activities: a cross-specialty national survey. *J Contin Educ Health Prof*;2017(37):154–160.
6. Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. *Ann Intern Med*. 2005;142(4):260–273.
7. Caddick ZA, Fraundorf SH, Rottman BM, Nokes-Malach TJ. Cognitive perspectives on maintaining physicians' medical expertise: II. acquiring, maintaining, and updating cognitive skills. *Cogn Res Princ Implic*. 2023;8(1):47.
8. Standards for continuing certification. *American Board of Medical Specialties*. 2024. <https://www.abms.org/board-certification/board-certification-standards/standards-for-continuing-certification>.
9. Anziani H, Durham J, Moore U. The relationship between formative and summative assessment of undergraduates in oral surgery. *Eur J Dent Educ*. 2008;12(4):233–238.
10. McNulty JA, Espiritu BR, Hoyt AE, Ensminger DC, Chandrasekhar AJ. Associations between formative practice quizzes and summative examination outcomes in a medical anatomy course. *Anat Sci Educ*. 2015;8(1):37–44.
11. Arja SB, Acharya Y, Alezaireg S, Ilavarasan V, Ala S, Arja SB. Implementation of formative assessment and its effectiveness in undergraduate medical education: an experience at a Caribbean medical school. *MedEdPublish*. 2018;7:131.
12. Ekelu SO. *Correlation between formative and summative assessment results in engineering studies. Presented at: The 6th African Engineering Education Association Conference*. 2016. <https://hdl.handle.net/10210/214036>.
13. Bazelaïs P, Doleck T, Lemay D. Exploring the association between formative and summative assessments in a pre-university science program. *J Form Des Learn*. 2017;1(12):1–8.
14. Price DW, Wang T, Neill O, Bazemore TR, Newton A, P W. Differences in physician performance and self-rated confidence on high and low stakes knowledge assessments in board certification. *J Contin Educ Health Prof*. 2024;44(1):2–10.
15. Fraundorf SH, Caddick Z, Rottman B. Conceptual foundations for designing continuing certification assessments for physicians. *American Board of Family Medicine*. 2024. <https://www.theabfm.org/app/uploads/2024/02/Conceptual-Foundations.pdf>.
16. Thiede KW. The relative importance of anticipated test format and anticipated test difficulty on performance. *J Exp Psychol A*. 1996;49(4):901–918.
17. Fraundorf SH, Caddick ZA, Nokes-Malach TJ, Rottman BM. Cognitive perspectives on maintaining physicians' medical expertise: III. strengths and weaknesses of self-assessment. *Cogn Res Princ Implic*. 2023;8(1):58.
18. Kemp PR, Bradshaw JM, Pandya B, Davies D, Morrell MJ, Sam AH. The validity of engagement and feedback assessments (EFAs): identifying students at risk of failing. *BMC Med Educ*. 2023;23(1):866.
19. Fredricks JA, Blumenfeld PC, Paris AH. School engagement: potential of the concept, state of the evidence. *Rev Educ Res*. 2004;74(1):59–109.
20. Ben-Eliyahu A, Moore D, Dorph R, Schunn CD. Investigating the multidimensionality of engagement: affective, behavioral, and cognitive engagement across science activities and contexts. *Contemp Educ Psychol*. 2018;53:87–105.
21. Dong A, Jong MY, King RB. How does prior knowledge influence learning engagement? the mediating roles of cognitive load and help-seeking. *Front Psychol*. 2020;11:591203–591203.
22. O'Neill TR, Li Z, Peabody MR, Lybarger M, Royal K, Puffer JC. The predictive validity of the ABFM's in-training examination. *Fam Med*. 2015;47(5):349–356.
23. Stewart R, Cooling N, Emblen G. Early predictors of summative assessment performance in general practice post-graduate training: a retrospective cohort study. *Med Teach*. 2018;40(11):166–167.
24. Price DW, Wang T, Neill O, Bazemore TR, Newton A, P W. Differences in physician performance and self-rated confidence on high and low stakes knowledge assessments in board certification. *J Contin Educ Health Prof*. 2024;44(1):2–10.