

Reports and Recommendations

Novel Formative Approach of the ESAP-ITE Provides Strong Predictive Value for ABIM Certification Outcomes

William B. Horton,¹ James T. Patrie,² Lauren M. Duhigg,³ Maggie Graham,⁴ Mark W. True,⁵ Elaine M. Pelley,⁶ and Alan C. Dalkin¹

¹Division of Endocrinology and Metabolism, Department of Medicine, University of Virginia School of Medicine, Charlottesville, VA 22908, USA; ²Division of Biostatistics, Department of Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA 22908, USA; ³American Board of Internal Medicine, Philadelphia, PA 19106, USA; ⁴Endocrine Society, Washington, DC 20036, USA; ⁵San Antonio Uniformed Services Health Education Consortium, Brooke Army Medical Center, JBSA Fort Sam Houston, San Antonio, TX 78234, USA; and ⁶Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI 53705, USA

ORCID number: 0000-0001-8081-1214 (W. B. Horton); 0000-0001-5258-4448 (A. C. Dalkin).

Abbreviations: ABIM, American Board of Internal Medicine; ECE, Endocrinology, Diabetes, and Metabolism Certification Examination; ESAP, Endocrine Self-Assessment Program; ITE, In-Training Examination; MVLr, multivariable logistic regression.

Received: 28 June 2021; Editorial Decision: 12 August 2021; First Published Online: 20 August 2021; Corrected and Typeset: 25 September 2021.

Abstract

Background: The Endocrine Self-Assessment Program In-Training Examination (ESAP-ITE) has the novel formative approach of allowing open access to all questions and answers after secure examination administration is complete, resulting in the creation of an entirely new in-training examination annually.

Objective: To determine whether scores on the novel ESAP-ITE predict pass/fail outcomes on the American Board of Internal Medicine Endocrinology, Diabetes, and Metabolism Certification Examination (ABIM-ECE).

Methods: All endocrine fellows-in-training who took the ESAP-ITE between 2016 and 2019 and then subsequently attempted the ABIM-ECE within the same calendar year were included (n = 982). Primary analyses used the ESAP-ITE score from the final year of fellowship training. Covariates included sex, age on date of ABIM-ECE, medical school country, fellowship program region, pass/fail outcomes on the ABIM Internal Medicine Certification Examination, and ESAP-ITE score. All variables were analyzed using multivariable logistic regression.

Results: ESAP-ITE score ($P < 0.001$), ABIM Internal Medicine Certification Examination outcome ($P < 0.001$), and age ($P = 0.005$) were each significant predictors of passing the ABIM-ECE on the first attempt. ESAP-ITE score was the strongest predictor of passing

the ABIM-ECE, and this relationship was such that a score of 75% correct yielded a 97% probability of passing the ABIM-ECE, whereas a score of 50% correct generated only a 70% probability of doing so. Sex, fellowship program region, and medical school country were not significant predictors of ABIM-ECE outcomes.

Conclusions: In addition to serving as an important learning instrument for endocrine fellowship programs, ESAP-ITE is a robust predictive tool for pass/fail outcomes on the ABIM-ECE.

Key Words: graduate medical education, certification, specialty boards, knowledge

The use of an In-Training Examination (ITE) has long been an important component in the annual assessment of progress in medical knowledge during residency training in internal medicine [1]. The Internal Medicine ITE is composed and administered by the American College of Physicians and provides a postassessment report consisting of topical feedback to both residents and program directors. Results from this annual evaluation are used by program directors to focus educational topics in subsequent years, as well as to monitor trainees' progress toward competency and preparedness to sit for the board certification examination at the completion of training. Indeed, it is well-recognized that performance on the Internal Medicine ITE provides some predictive value for trainee success when attempting board certification [2, 3].

In recent years, ITE development and implementation has been part of most medical subspecialty fields [4-11]. The composition of these tools generally reflects the blueprint of content coverage in the respective American Board of Internal Medicine (ABIM) subspecialty certification examinations. Each of these ITE products are given in an online, secure fashion and comprises multiple-choice questions. For nearly all subspecialties, the questions are developed, refined, and retained in a confidential manner to generate a secure question bank that can be reused over time, with feedback to the examinee only being provided by topical score with the number of correct/incorrect responses [12-16].

In response to the need for better tools by which to assess endocrine trainee progress in medical knowledge, the Endocrine Society developed the Endocrine Self-Assessment Program ITE (ESAP-ITE). A unique feature of the ESAP-ITE is its design to serve as both an assessment of current medical knowledge and a formative learning tool. As such, after secure examination delivery, all fellows-in-training and fellowship programs: (1) receive performance feedback that includes all questions, answers, and an associated discussion with learning objectives and references for each question; and (2) are exclusively eligible to attend a 90-minute session at the annual Endocrine Society Scientific Meeting dedicated to reviewing that year's examination. Given that all questions are released to participating programs,

each year's examination consists of an entirely new set of questions. When the Endocrine Society chose this unique model, it prioritized the educational value of participation in ESAP-ITE, while acknowledging that its ability to predict outcomes on the ABIM Endocrinology, Diabetes, and Metabolism Certification Exam (ABIM-ECE) might be lost. An important component of this decision included surveying key stakeholders (eg, fellowship program directors) to determine if they would be willing to sacrifice the formative uses of ESAP-ITE in favor of a secure, predictive examination. Survey respondents strongly indicated that the formative nature of the examination was more valuable than a secure, predictive examination because it allowed for more directed learning. The purpose of this investigation is to retrospectively determine if the ESAP-ITE can predict pass/fail outcomes on the ABIM-ECE and thereby serve as both a gauge of trainees' knowledge as they prepare for board certification and as a formative educational tool for fellowship program directors.

Materials and Methods

Study Sample

All data included in this study were deidentified and all protocols were approved by the University of Virginia Institutional Review Board for Social and Behavioral Sciences (Protocol #2920). The study sample consisted of 982 endocrine fellows-in-training who took the ESAP-ITE between 2016 and 2019 and then subsequently attempted the ABIM-ECE within the same calendar year. There were 1052 total endocrine fellows who met these criteria, and the ABIM was able to match 982 (93.3%) of them by first name, last name, and date of birth provided by the Endocrine Society. We note that 24/982 (2.4%) fellows completed 3 years of training (ie, academic training pathway), whereas the remaining fellows completed 2 years of training (ie, clinical training pathway).

ESAP-ITE

The ESAP-ITE is the only in-training examination available for endocrine fellows-in-training. The examination

consists of 90 case-based questions written in a single best answer, multiple-choice format, and is developed annually by a committee of volunteer medical educators from Endocrine Society membership. This committee (ESAP-ITE Steering Group) is responsible for the selection of examination content and decisions related to examination utility, with its ultimate mission being to ensure relevance and usefulness of the examination to endocrinology training programs and fellows-in-training. The ESAP-ITE Steering Group consists of a chair and 8 topical teams.

The examination-building process occurs each year from approximately May to December. Questions are drawn from banks of rigorously developed and peer-reviewed content available from other Endocrine Society question-based learning products (modified to be appropriate for fellows-in-training). Topical teams select questions to align with the ABIM-ECE blueprint [17] and provide a balance of fundamental and advanced questions (based on judgment of the faculty members and any data available from prior use of that question) in each topical area. This is followed by several rounds of peer review and question editing, after which the examination is uploaded to the online platform and beta tested. The examination is proctored by fellowship program administrators from mid-January to mid-February annually and administered via a secure website. Each examination is graded with an overall score, as well as a score within each major content area. Scores are reported as percent correct (ie, 45/90 correct responses overall would be reported as a score of 50%). In addition, percentile ranks are also provided to compare the performance of each fellow-in-training with that of other endocrine trainees. As expected, given the lack of ability to curate questions over time, each ESAP-ITE has demonstrated acceptable [18] but modestly lower examination reliability than is seen with high-stakes standardized examinations such as the ABIM-ECE. Specifically, ESAP-ITE reliability coefficients (ie, Cronbach's alpha) [19] ranged from 0.72 to 0.77 in the 2016 to 2019 examination administrations (Supplemental Table 1 [20]).

ABIM internal medicine certification examination

The ABIM Internal Medicine Certification Examination was developed to assess whether an individual has the knowledge, diagnostic reasoning, and clinical judgment skills expected of a certified internist. The examination is computer-based, consists of 205 live multiple-choice questions, and is generally taken within the first year after completion of internal medicine residency training. The examination score is computed using item response theory, and scores are equated and standardized on a scale of 200 to 800 (with a mean of 500 and SD of 100 points) so that

they are comparable across years. Reliability coefficients of the ABIM Internal Medicine Certification Examination were approximately 0.90 across recent administrations [8].

ABIM-ECE

Shortly after completion of Endocrinology, Diabetes, and Metabolism fellowship training, endocrinologists take the ABIM-ECE to gain certification in the subspecialty. Certification is not mandated for endocrinologists to practice medicine in the United States; however, it is often required by employers, hospitals, and payers [21]. The ABIM-ECE is designed to test the Accreditation Council for Graduate Medical Education core competencies of medical knowledge and patient care in the major endocrine areas, including adrenal disorders; pituitary disorders; lipids, obesity, and nutrition; female reproduction; male reproduction; diabetes mellitus and hypoglycemia; calcium and bone disorders; and thyroid disorders [17]. The ABIM-ECE is administered via a secure computer-based format and is generally taken in the first year after completion of endocrinology fellowship training. Similar to the ABIM Internal Medicine Certification Examination, the ABIM-ECE consists of 200 live multiple-choice questions and is scored using item response theory. The scaled score ranges from 200 to 800 (with a mean of 500 and SD of 100 points) and are equated and standardized to ensure comparability across years. The ABIM-ECE has demonstrated good [18] reliability, with reliability coefficients consistently ranging from 0.85 to 0.86 in the 2016 to 2019 examination administrations (Supplemental Table 1 [20]).

Statistical Analyses

Categorical scaled data were summarized by frequencies and relative frequencies (ie, percentages). Continuous scaled data were summarized by mean, SD, and range of the empirical distribution.

Multivariable logistic regression (MVLRL) analysis was used to determine if the ESAP-ITE score provided unique predictive information about a fellow-in-training's odds of passing the ABIM-ECE on the first attempt. Primary analyses used the ESAP-ITE score from the final year of fellowship training (ie, only 1 score per fellow-in-training). For model specification, the MVLRL-dependent variable (Y) identified whether the fellow-in-training passed the ABIM-ECE on first-attempt ($Y = 1$) or failed to pass on first attempt ($Y = 0$). MVLRL model predictor variables were provided to us in the dataset received from the ABIM and Endocrine Society. These variables were consistent

with methodology from prior reports evaluating the predictive ability of internal medicine subspecialty ITEs for performance on board certification examinations [4, 5, 7, 9] and included sex (self-identified by each fellow), age on date of ABIM-ECE, country in which medical school degree was obtained (United States or international), fellowship program region (Northeast, South, Midwest, West, or Puerto Rico), performance on the ABIM Internal Medicine Certification Examination (pass/fail), and ESAP-ITE score. With regard to hypothesis testing, the Wald type III χ^2 statistic was used to test the null hypothesis that the predictor variable provides no unique predictive information about whether a fellow-in-training will pass the ABIM-ECE on the first attempt above and beyond the predictive information attributed to the remaining MVLR predictor variables. An a priori P value ≤ 0.05 decision rule applied as the null hypothesis rejection criterion for each hypothesis test. Directionality and magnitude of the association between the odds of passing the ABIM-ECE on first attempt and the predictor variable were quantified by the MVLR adjusted odds ratio. Prespecified hypotheses were that pass/fail outcomes on the ABIM Internal Medicine Certification Examination and ESAP-ITE score would each significantly predict ABIM-ECE pass/fail outcomes. MVLR analysis was conducted via the statistical software package Spotfire Splus version 8.2 (TIBCO Software, Inc; Palo Alto, CA).

Results

Demographic Characteristics and Examination Scores

Table 1 details pertinent descriptive statistics for the study cohort. From 2016 to 2019, there were 982 endocrine fellows-in-training who took the ESAP-ITE and were subsequently matched with ABIM-ECE data. Each subject had ABIM Internal Medicine Certification Examination and ABIM-ECE outcome data available for study inclusion. The final study cohort identified as 71.9% female, 58.2% graduated from a medical school outside the United States, and 89.3% passed the ABIM-ECE on the first attempt.

Predictors of ABIM-ECE Outcomes

After MVLR analysis, ANOVA Wald Type III testing revealed that ESAP-ITE score was the strongest predictor of first attempt passage of the ABIM-ECE (Table 2). Passing the ABIM-Internal Medicine Certification Examination on the first attempt and younger age on date of the ABIM-ECE were significant, but less strong, predictors of passing the ABIM-ECE. Sex, fellowship program region, and country in which medical school degree was obtained were not significant predictors of ABIM-ECE outcomes.

Table 3 presents the adjusted odds ratios for passing the ABIM-ECE on the first attempt. Scoring 10 percentage

Table 1. Demographic characteristics and examination performance of 982 endocrinology Fellows who took the ESAP-ITE between 2016 and 2019 and then subsequently attempted the ABIM-ECE within the same calendar year

Sex, n (%)	
Female	706 (71.9)
Male	276 (28.1)
Medical school country, n (%)	
USA	410 (41.8)
International	571 (58.1)
Unknown	1 (0.001)
Age on date of ABIM-ECE, mean years \pm SD (range)	34.5 \pm 3.6 (28.5-56.8)
USA Fellowship program region, n (%)	
South	344 (35.0)
Northeast	271 (27.6)
Midwest	220 (22.4)
West	125 (12.7)
Puerto Rico	22 (2.2)
Passed ABIM-IMCE on first attempt, n (%)	917 (93.4)
Performance on ESAP-ITE	
Score, mean \pm SD (range)	64.9 \pm 8.7 (34.0-91.0)
Performance on ABIM-ECE	
Passed on first attempt, n (%)	877 (89.3)

Abbreviations: ABIM-ECE, American Board of Internal Medicine Endocrinology, Diabetes, and Metabolism Certification Examination; ABIM-IMCE, American Board of Internal Medicine Internal Medicine Certification Examination; ESAP-ITE, Endocrine Self-Assessment Program-In-Training Examination.

Table 2. ANOVA Wald type III test summary for testing the null hypothesis that there is no unique association between the predictor variable and log-odds that the fellow-in-training passed the ABIM-ECE on the first attempt

Predictor	Wald χ^2 statistic	DoF	(Wald χ^2 statistic) – (DoF)	P value
ESAP-ITE score	52.26	1	51.26	<0.001
Passed ABIM-IMCE on first attempt	12.74	1	11.74	<0.001
Age on date of ABIM-ECE	7.80	1	6.80	0.005
Fellowship program region	7.32	4	3.32	0.120
Sex	2.07	1	1.07	0.151
Medical school country	1.61	1	0.61	0.205
Total	91.14	9	82.14	0.000

Abbreviations: DoF, degrees of freedom; ABIM-ECE, American Board of Internal Medicine Endocrinology, Diabetes, and Metabolism Certification Examination; ABIM-IMCE, American Board of Internal Medicine Internal Medicine Certification Examination; ESAP-ITE, Endocrine Self-Assessment Program-In-Training Examination.

Table 3. Multiple logistic regression results for predicted probability of passing the ABIM-ECE on the first attempt

Predictor	Odds ratio; 95% CI	P-value
ESAP-ITE score (X + 10: X)	2.68; 2.05-3.51	<0.001
ESAP-ITE score (X + 5: X)	1.64; 1.43-1.87	<0.001
Passed ABIM-IMCE on first attempt (yes: no)	3.18; 1.68-6.00	<0.001
Age on date of ABIM-ECE (X - 5 years: X y)	1.50; 1.13-2.00	0.005
Sex (male: female)	1.48; 0.87-2.52	0.150
Medical school country (USA: international)	1.28; 0.82-2.28	0.205
Fellowship program region (global test)	-	1.20

Abbreviations: ABIM-ECE: American Board of Internal Medicine- Endocrinology, Diabetes, and Metabolism Certification Examination; ABIM-IMCE, American Board of Internal Medicine Internal Medicine Certification Examination; ESAP-ITE, Endocrine Self-Assessment Program-In-Training Examination.

points higher on the ESAP-ITE provided 2.68 times higher odds (95% CI, 2.05-3.51; $P < 0.001$) of passing the ABIM-ECE. In this scenario, a fellow-in-training who got 60% correct on the ESAP-ITE will have 2.68 times higher odds of passing the ABIM-ECE on the first attempt than someone who got 50% correct. The odds ratio remained significantly higher when we examined the difference in scoring 5 points higher on the ESAP-ITE (odds ratio: 1.64; 95% CI, 1.43-1.87; $P < 0.001$). Passing the ABIM-Internal Medicine Certification Examination on the first attempt resulted in 3.18 times higher odds (95% CI, 1.68-6.00; $P < 0.001$) for passing the ABIM-ECE on the first attempt.

The predicted probability of passing the ABIM-ECE based solely on ESAP-ITE score is presented in [Fig. 1](#). An ESAP-ITE score of 75 (ie, 75% correct) generated a 97% probability of passing the ABIM-ECE, whereas a score of 50 provided only a 70% probability of doing so. Estimated probability of passing the ABIM-ECE increased by $\geq 2\%$ for each 1-point increase in ESAP-ITE score up to 53, and increased by $\geq 1\%$ for each 1-point increase in ESAP-ITE score from 54 to 63 ([Table 4](#)).

Discussion

The current study is the first to investigate performance on the ESAP-ITE, a formative ITE constructed by a novel process leading to an original examination each year, as predictor of ABIM-ECE pass/fail outcomes. Our analysis showed a significant positive correlation between increasing ESAP-ITE scores and probability of passing the ABIM-ECE on the first attempt. Indeed, the strongest predictor of passing the ABIM-ECE was ESAP-ITE score; followed by previously passing the ABIM Internal Medicine Certification Examination on the first attempt. These findings are consistent with reports from other internal medicine subspecialties showing that ITEs and prior standardized testing performance also predict outcomes on their respective subspecialty certification examinations [4-9]. Younger age on date of ABIM-ECE was the only other significant predictor, and this result also mirrors prior studies that found younger test-takers were more likely to pass initial board certification examinations on the first attempt [3, 7, 22]. Although reasons for this are speculative, Jeffe and Andriole have suggested that such results indicate older graduates may experience greater difficulties in timely advancement toward board certification [22]. We suggest

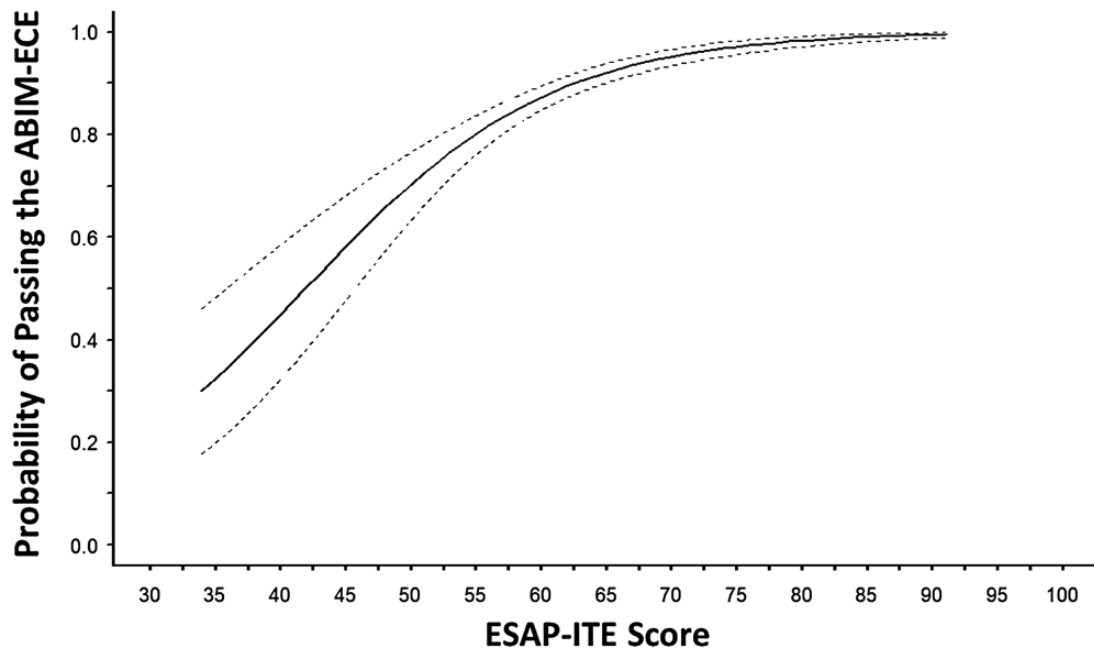


Figure 1. Probability of passing the ABIM-ECE as a function of ESAP-ITE score. The solid black line identifies the predicted probability; the dashed lines identify the 95% simultaneous confidence band for the location of the solid black line. ABIM-ECE, American Board of Internal Medicine Endocrinology, Diabetes, and Metabolism Certification Examination; ESAP-ITE, Endocrine Self-Assessment Program-In-Training Examination.

that it may be prudent for older endocrine fellows (especially those with poorer performance on the ESAP-ITE) to consider starting a dedicated board examination study program early in their training and/or consider using a board review course before taking the ABIM-ECE. Finally, sex, fellowship program region, and country in which medical school degree was obtained did not significantly affect odds of passing the ABIM-ECE, indicating that there are few biases in the examination itself. These results confirm the ESAP-ITE as a valuable assessment tool that offers support to endocrine fellows-in-training and their fellowship programs as they seek to identify and evaluate knowledge gaps for both clinical care and board certification purposes.

Before development of the ESAP-ITE, endocrine fellows-in-training did not have a standardized tool for assessment of medical knowledge during training. Since inception of its current format in 2016, the ESAP-ITE has effectively filled this assessment gap and become the primary testing tool during Endocrinology, Diabetes, and Metabolism fellowship training. Nearly all ($\geq 98\%$) Accreditation Council for Graduate Medical Education-accredited Endocrinology, Diabetes, and Metabolism fellowship programs participate in the ESAP-ITE annually, and the examination is an important resource for fellowship program curriculum design and development across the country. However, 1 key unanswered question regarding ESAP-ITE was whether this novel test development process could provide adequate examination reliability to predict ABIM-ECE outcomes. The current study firmly establishes the relationship

between ESAP-ITE scores during the final year of fellowship training with pass/fail outcomes on the ABIM-ECE, and provides clear evidence that the ESAP-ITE can serve as a valuable tool by which fellowship program directors can assess fellow-in-training readiness for board certification while not sacrificing its utility as a learning tool. Furthermore, these results combined with the consistent reliability coefficients between ESAP-ITE administrations (Supplemental Table 1 [20]) demonstrate that a de novo annual ITE (created without the benefit of being able to keep high-performing questions and modify or replace poor performing questions over serial examination administrations) can still provide trainees a useful tool for knowledge assessment. When used as intended, the ESAP-ITE represents an important resource for assessment of medical knowledge and self-identification of potential knowledge gaps. McDonald et al have shown that “a systematic, frequent (eg, daily), modest (eg, ≤ 20 minutes per day), intentional approach to medical knowledge acquisition” is associated with significant gains in knowledge over time [3, 23]. Such an approach, when guided by ESAP-ITE results, could conceivably increase odds of an “at-risk” test-taker passing the ABIM-ECE. The success of this approach could also be assessed by serial performance on the ESAP-ITE (ie, a personalized learning plan implemented after year 1 ESAP-ITE could be assessed by performance on the year 2 examination).

Although there are several strengths to our study, it is not without limitations. First, there are likely unknown

Table 4. Predicted probability of passing the ABIM-ECE as a function of ESAP-ITE score

ESAP-ITE score (% correct)	Probability of passing ABIM-ECE; 95% CI
34	0.300; 0.177-0.461
35	0.323; 0.197-0.481
36	0.347; 0.219-0.502
37	0.371; 0.242-0.522
38	0.396; 0.267-0.543
39	0.422; 0.293-0.563
40	0.448; 0.321-0.583
41	0.475; 0.349-0.603
42	0.501; 0.380-0.623
43	0.528; 0.411-0.642
44	0.554; 0.442-0.661
45	0.580; 0.474-0.680
46	0.606; 0.506-0.698
47	0.631; 0.538-0.715
48	0.656; 0.570-0.732
49	0.679; 0.601-0.749
50	0.702; 0.631-0.765
51	0.724; 0.659-0.780
52	0.744; 0.687-0.795
53	0.764; 0.712-0.809
54	0.783; 0.737-0.823
55	0.800; 0.759-0.836
56	0.817; 0.780-0.849
57	0.832; 0.799-0.861
58	0.847; 0.816-0.872
59	0.860; 0.832-0.884
60	0.872; 0.847-0.894
61	0.884; 0.859-0.904
62	0.894; 0.871-0.913
63	0.904; 0.882-0.922
64	0.913; 0.891-0.930
65	0.921; 0.900-0.938
66	0.928; 0.908-0.944
67	0.935; 0.915-0.951
68	0.941; 0.922-0.956
69	0.947; 0.928-0.961
70	0.952; 0.933-0.966
71	0.957; 0.938-0.969
72	0.961; 0.943-0.973
73	0.965; 0.947-0.976
74	0.968; 0.951-0.979
75	0.971; 0.955-0.982
76	0.974; 0.959-0.984
77	0.977; 0.962-0.986
78	0.979; 0.965-0.987
79	0.981; 0.967-0.989
80	0.983; 0.970-0.990
81	0.985; 0.972-0.991
82	0.986; 0.974-0.992
83	0.987; 0.976-0.993
84	0.989; 0.978-0.994
85	0.990; 0.980-0.995
86	0.991; 0.981-0.996
87	0.992; 0.983-0.996
88	0.993; 0.984-0.997
89	0.993; 0.985-0.997
90	0.994; 0.987-0.997
91	0.995; 0.988-0.998

Abbreviations: ABIM-ECE: American Board of Internal Medicine Endocrinology, Diabetes, and Metabolism Certification Examination; ESAP-ITE: Endocrine Self-Assessment Program-In-Training Examination.

variables, interactions, and/or social dynamics not included in this analysis that may be associated with ESAP-ITE and/or ABIM-ECE performance. Second, we only had pass/fail results from each fellow-in-training's ABIM Internal Medicine Certification examination; thus, we were unable to investigate whether exact scores on that examination predicted ABIM-ECE outcomes. Likewise, we only had access to ESAP-ITE scores from each fellow-in-training's final year of fellowship, so we were unable to assess whether first year performance or performance trends during fellowship (eg, increasing scores from year 1 to year 2) predicted ABIM-ECE outcomes. An earlier marker of board performance, such as ESAP-ITE in year 1 of fellowship training, would allow for more lead time to assist at-risk fellows. Third, the ABIM-ECE had an initial pass rate that ranged from 84% to 91% during the study period [24], and this relatively high pass rate may limit generalizability to other specialties with substantially lower pass rates for initial certification examinations. Finally, although the ESAP-ITE process outlined above has yielded stable examination metrics (ie, difficulty, reliability) over several years with cumulative results that are highly predictive of ABIM-ECE pass rate, this can only suggest (but not guarantee) the predictive ability of any given individual ESAP-ITE.

In conclusion, the ESAP-ITE is a robust predictive tool for pass/fail outcomes on the ABIM-ECE. These results provide fellows-in-training and fellowship program directors with important data and resources to help gauge trainee standing as they prepare for board certification while continuing to serve as a highly valued question-based learning tool.

Acknowledgments

The authors thank the Endocrine Society and the American Board of Internal Medicine for collaborating on this project. The authors wish to recognize the critical contribution of Kristen Ndure who, while employed by the Endocrine Society, was pivotal to preparing, publishing, and presenting the annual ESAP-ITE; and the critical contribution of Nancy Rohowyj, who facilitated data transfer between the American Board of Internal Medicine and the Endocrine Society. We also thank the faculty members of the 2016-2019 ESAP-ITE Steering Groups, whose excellent work led to the development and creation of the ESAP-ITE examinations analyzed in the current study. The 2016-2019 ESAP-ITE Steering Groups consisted of the following individuals: Derick Adams, DO; Shehzad Basaria, MD; Darrick Beckman, MD; Eve Bloomgarden, MD; Alan Chang, MD; Natsurang Chongkairatanakul, MD; David Cooper, MD; Ann Danoff, MD; Ines Donangelo, MD, PhD; Leigh Eck, MD; Gunjan Gandhi, MD; Salila Kurra, MD; Alan Malabanan, MD; Sarah Mayson, MD; Vijay Shivaswamy, MD; Elias Siraj, MD; Sirinart Sirinvaravong, MD; Lisa Tannock, MD; Nick Tritos, MD; Nicole Victor, MD; Robert Yanagisawa, MD; and Roberto Salvatori, MD.

Financial Support: This work was supported in part by National Institutes of Health research grants KL2TR003016 and UL1TR003015 (to W.B.H.).

Additional Information

Correspondence: Alan C. Dalkin, MD, University of Virginia, 450 Ray C. Hunt Drive, Box 801412, Charlottesville, VA 22908, USA. Email: ACD6V@virginia.edu.

Disclosures: W.B.H. and J.T.P. have nothing to declare or disclose. L.M.D. is employed by the American Board of Internal Medicine. M.G. is employed by the Endocrine Society. M.W.T., E.M.P., and A.C.D. have each chaired the ESAP-ITE Steering Group, written for the examination, and participated in the annual performance reviews.

Data Availability: Some or all datasets generated during and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

References

- Garibaldi RA, Subhiyah R, Moore ME, Waxman H. The In-Training Examination in Internal Medicine: an analysis of resident performance over time. *Ann Intern Med.* 2002;137(6):505-510.
- Babbott SF, Beasley BW, Hinchey KT, Blotzer JW, Holmboe ES. The predictive validity of the internal medicine in-training examination. *Am J Med.* 2007;120(8):735-740.
- McDonald FS, Jurich D, Duhigg LM, et al. Correlations between the USMLE step examinations, American College of Physicians In-Training Examination, and ABIM Internal Medicine Certification Examination. *Acad Med.* 2020;95(9):1388-1395.
- Lohr KM, Clauser A, Hess BJ, et al.; American College of Rheumatology Committee on Rheumatology Training and Workforce Issues. Performance on the adult rheumatology in-training examination and relationship to outcomes on the rheumatology certification examination. *Arthritis Rheumatol.* 2015;67(11):3082-3090.
- Grabovsky I, Hess BJ, Haist SA, et al. The relationship between performance on the Infectious Diseases In-Training and Certification Examinations. *Clin Infect Dis.* 2015;60(5):677-683.
- Collichio FA, Hess BJ, Muchmore EA, et al. Medical knowledge assessment by hematology and medical oncology in-training examinations are better than program director assessments at predicting subspecialty certification examination performance. *J Cancer Educ.* 2017;32(3):647-654.
- Indik JH, Duhigg LM, McDonald FS, et al. Performance on the cardiovascular in-training examination in relation to the ABIM Cardiovascular Disease Certification Examination. *J Am Coll Cardiol.* 2017;69(23):2862-2868.
- Jurich D, Duhigg LM, Plumb TJ, et al. Performance on the nephrology in-training examination and ABIM nephrology certification examination outcomes. *Clin J Am Soc Nephrol.* 2018;13(5):710-717.
- Kempainen RR, Hess BJ, Addrizzo-Harris DJ, et al. Pulmonary and critical care in-service training examination score as a predictor of board certification examination performance. *Ann Am Thorac Soc.* 2016;13(4):4810-4488.
- Rosner MH, Berns JS, Parker M, et al.; ASN In-Training Examination Committee. Development, implementation, and results of the ASN in-training examination for fellows. *Clin J Am Soc Nephrol.* 2010;5(2):328-334.
- Kuvin JT, Soto A, Foster L, et al. The cardiovascular in-training examination: development, implementation, results, and future directions. *J Am Coll Cardiol.* 2015;65(12):1218-1228.
- American Society of Clinical Oncology. Medical oncology in-training examination. <https://www.asco.org/training-education/education-career-resources/resources-program-directors/medical-oncology-training>. Accessed April 19, 2021.
- American Society of Nephrology. 2021 in-training examination for nephrology fellows. <https://www.asn-online.org/education/training/ite/>. Accessed April 19, 2021.
- Infectious Diseases Society of America. Fellows in-training exam. <https://www.idsociety.org/professional-development/fellows-in-training-career-education-center/fellows-in-training-exam/fite-main-page/>. Accessed April 19, 2021.
- American College of Cardiology. The American College of Cardiology In-Training Exam. <https://www.acc.org/membership/sections-and-councils/cardiology-training-and-workforce-committee/features/the-american-college-of-cardiology-in-training-exam-acc-ite>. Accessed April 19, 2021.
- American Gastroenterological Association. AGA gastroenterology training examination. <https://gastro.org/fellows-and-early-career/aga-gastroenterology-training-examination/>. Accessed April 19, 2021.
- American Board of Internal Medicine. Endocrinology, diabetes, and metabolism certification examination blueprint. <https://www.abim.org/Media/wxbjt5o3/endocrinology-diabetes-metabolism.pdf>. Accessed May 27, 2021.
- Cortina JM. What is coefficient alpha? An examination of theory and applications. *J Appl Psychol* 1993; 78(1):98-104.
- Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res Sci Educ.* 2018; 48(6):1273-1296.
- Supplemental Material. Available at: <https://osf.io/aqf6d/>.
- Freed GL, Dunham KM, Gebremariam A. Changes in hospitals' credentialing requirements for board certification from 2005 to 2010. *J Hosp Med* 2013; 8(6):298-303.
- Jeffe DB, Andriole DA. Factors associated with American Board of Medical Specialties member board certification among US medical school graduates. *JAMA.* 2011;306(9):961-970.
- McDonald FS, Zeger SL, Kolars JC. Factors associated with medical knowledge acquisition during internal medicine residency. *J Gen Intern Med.* 2007;22(7):962-968.
- American Board of Internal Medicine. Initial certification pass rates. <https://www.abim.org/Media/yeqiumdc/certification-pass-rates.pdf>. Accessed June 2, 2021.