A Systematic Review of the Relationship Between In-Training Examination Scores and Specialty Board Examination Scores

Hilary C. McCrary, MD, MPH
Jorie M. Colbert-Getz, PhD, MS
W. Bradley Poss, MD, MMM
Brigitte K. Smith, MD, MHPE

ABSTRACT

Background In-training examinations (ITEs) are intended for low-stakes, formative assessment of residents' knowledge, but are increasingly used for high-stake purposes, such as to predict board examination failures.

Objective The aim of this review was to investigate the relationship between performance on ITEs and board examination performance across medical specialties.

Methods A search of the literature for studies assessing the strength of the relationship between ITE and board examination performance from January 2000 to March 2019 was completed. Results were categorized based on the type of statistical analysis used to determine the relationship between ITE performance and board examination performance.

Results Of 1407 articles initially identified, 89 articles underwent full-text review, and 32 articles were included in this review. There was a moderate-strong relationship between ITE and board examination performance, and ITE scores significantly predict board examination scores for the majority of studies. Performing well on an ITE predicts a passing outcome for the board examination, but there is less evidence that performing poorly on an ITE will result in failing the associated specialty board examination.

Conclusions There is a moderate to strong correlation between ITE performance and subsequent performance on board examinations. That the predictive value for passing the board examination is stronger than the predictive value for failing calls into question the "common wisdom" that ITE scores can be used to identify "at risk" residents. The graduate medical education community should continue to exercise caution and restraint in using ITE scores for moderate to high-stakes decisions.

Introduction

In-training examinations (ITEs) have been used as an objective measure of residents' and fellows' medical knowledge since the 1970s. ITE scores and reports provide program directors with information on the strengths and weaknesses of their trainees' medical knowledge in various content areas, which can be used in a low-stakes, formative fashion to support development of individualized learning plans. ITE scores may also be utilized by program directors at the program level, with areas of poor performance across trainees suggesting potential gaps in program curricula and identifying areas on which to focus for continuous program improvement. Ultimately, graduate medical education (GME) programs are responsible for ensuring their trainees are equipped to succeed in passing the qualifying examination (QE) and/or certifying examination (CE), administered by their respective specialty board, at the conclusion of their training. It is unclear, however, if ITEs are predictive of trainees' success in the board certification process.

DOI: http://dx.doi.org/10.4300/JGME-D-20-00111.1

Validity evidence for the interpretation of scores from assessment tools can be organized into 5 categories, based on Messick's unified framework, including content, response process, relationship to other variables, internal structure, and consequences.[1] The category most relevant to gather evidence for ITE scores is relationship to other variables. If the ITE and respective specialty board examinations had similar test content, ITE scores would share a strong relationship with board examination scores. The predictive ability of ITEs has been an area of interest since the early 1990s, and the number of investigations of this topic has continued to increase in recent years. Furthermore, some specialties and programs have begun to expand the use of ITEs beyond the original low-stakes formative intent to more high-stakes decisions, including formal academic actions, such as formal remediation, probation, non-advancement, and non-retention within the training program, which has significant implications for the consequences of ITE scores.[2–4]

Given that ITEs could be utilized in a manner that impacts a trainee's future in terms of promotion and program completion, ensuring that there is validity

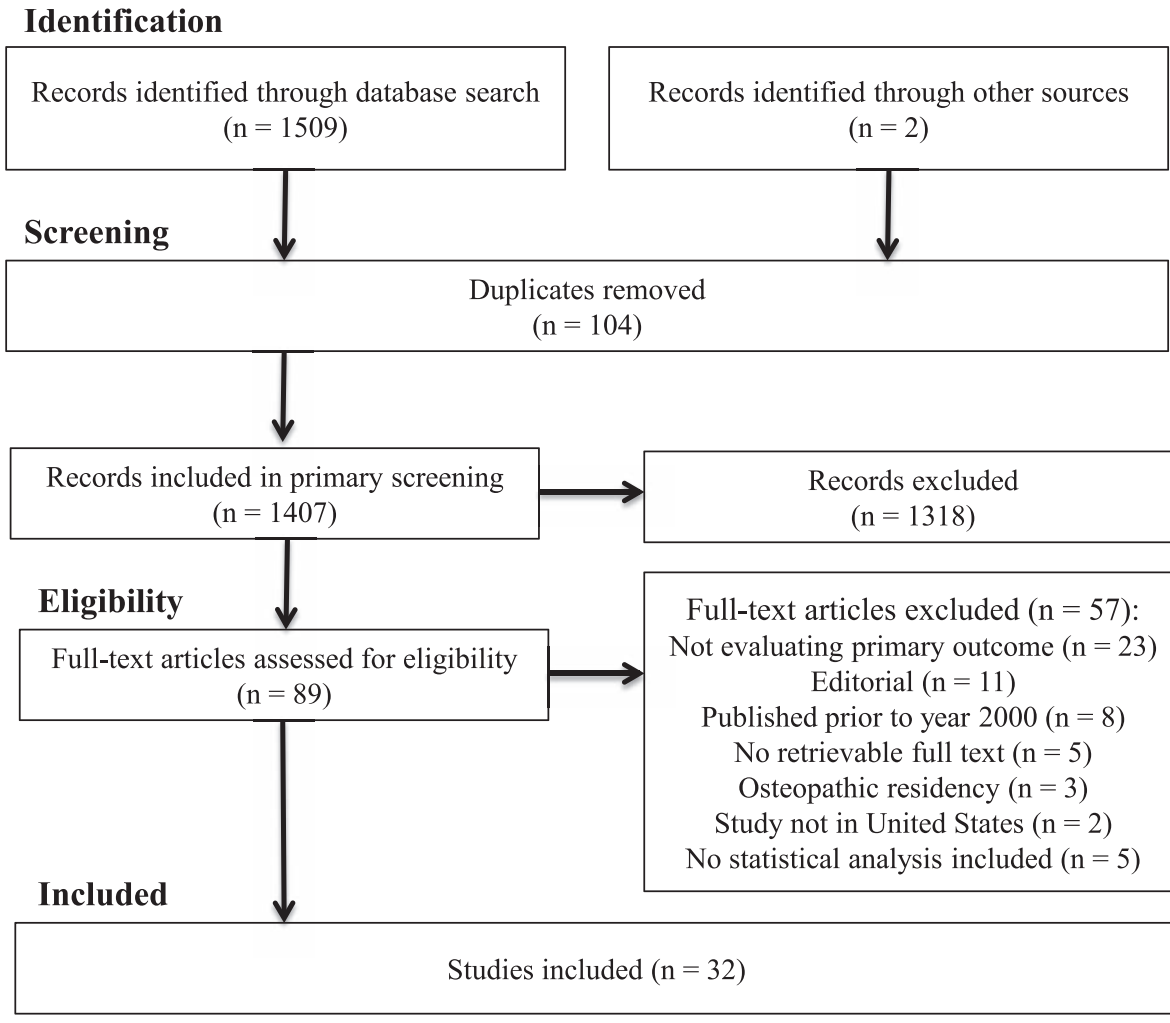

FIGURE
PRISMA Diagram Demonstrating Study Selection

evidence for the relationship between ITE scores and board examination scores is of the utmost importance. To date, there has neither been a review synthesizing the literature on the use of ITEs across medical specialties nor a synthesis of correlations/prediction results between ITE scores and board examination scores. Thus, the purpose of this study was to complete a systematic review of the literature on relationships to other variables' evidence for interpretation of GME ITE scores, with the other variable being performance on board examinations. A secondary aim of the study was to identify current use of ITEs across specialties.

Methods

Selection of Studies

We conducted a systematic review of the research on the association between ITEs and board examinations published from January 2000 to March 2019 using the following databases: PubMed, Embase, Cochrane Library, and Scopus. Major medical subject heading terms used for the systematic review included: in-training examination, in-service examination, medical education, and certification. Two authors (B.K.S. and H.C.M.) independently reviewed titles, abstracts, and full-text articles to determine if they met inclusion criteria. This process was completed with the assistance of systematic review software (Covidence, 2019). Phase 1 included screening of titles and abstracts for relevance. Phase 2 included evaluation of the full text. The search methods are reported using relevant items of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) checklist (FIGURE).

Eligibility Criteria

Studies were included if: (1) they reported quantitative analysis of an association between performance

on the ITE and performance on the respective specialty board examinations; (2) the study population included US GME trainees (residents or fellows); (3) manuscripts were available in the English language; (4) the full-text article was able to be obtained; and (5) articles were published after the year 2000. The criteria to include studies published after 2000 was established given our assessment of the availability of literature, which increased substantially after the year 2000.

Title/Abstract and Full-Text Review

Two authors (B.K.S. and H.C.M.) independently reviewed the titles and abstracts of all 1407 articles captured by the search, removing duplicates and articles obviously not meeting predetermined eligibility criteria. Discrepant opinions were discussed until consensus was reached during the abstract and full-text review stages. Two authors (B.K.S. and H.C.M.) completed the abstract review phase, while all 4 authors participated in the full-text review. A full-text review of 89 articles determined eligibility for inclusion in the final review, with a total of 32 articles ultimately included (FIGURE).

Relationship to Other Variables' Evidence

In the Messick validity evidence framework, relationship to other variables evidence refers to gathering information to show that assessment scores relate to scores from similar assessments. Such evidence generally takes 3 forms, including correlation coefficient, regression equation, and Area Under the ROC Curve (AUC). For continuous scores (eg, 0%–100%), relationships are measured with a correlation coefficient, where a strong positive correlation value is a metric for validity evidence. For educational purposes, correlation values > 0.50 are considered strong, 0.30–0.49 moderate, and < 0.30 low.[5] A significant regression equation is another potential metric for validity evidence where either continuous scores or dichotomous outcomes (eg, pass/fail) are used to predict future performance on another variable measured on a continuous scale (linear regression) or as dichotomous outcomes (logistic regression). Finally, an AUC with good accuracy/predictive value is a third potential metric for validity evidence where a particular score (eg, cut score) or outcome is used to discriminate between true positives and false positives of future performance.

Data Extraction and Analysis

Results were categorized based on the type of statistical analysis used to determine the relationship between ITE performance and board examination performance: correlation, linear regression, logistic regression, and/or AUC. Additionally, the type of ITE performance data (eg, percent score or rank) used for the analysis were extracted. Data were also collected from publicly available websites for each specialty society in terms of the format and number of ITE questions, and national pass rates for board examinations (TABLE 1).

Two authors (B.K.S. and H.C.M.) independently assessed the quality of the studies included in the final analysis using the Medical Education Research Study Quality Instrument (MERSQI). The MERSQI scoring system includes 10 items that are used to evaluate the quality of medical education research, including study design, institutions, response rate, type of data, validity, appropriateness of analysis, sophistication of analysis, and outcome.[6] Each item is scored (total possible score of 18), with Reed et al citing the mean as 9.6 in a cross-sectional study of 100 medical education research studies.[6] The validity and response rate items were not applicable to the studies included in our analysis; thus, these criteria were discarded, resulting in a total possible score of 13.5 points. Any discrepancies in scoring were resolved through group consensus. Importantly, the MERSQI scoring system is not intended to generate an absolute indicator of the validity or reliability of the research results. Furthermore, "cut-points" for "excellent" or "poor" quality have not been defined. Rather, the scores can be used to compare the quality of evidence between studies within a specific body of literature.

Given that there are differences in language across specialties in terms of what QE and CE means, the term *board examination* will henceforth refer to the written examination for each given specialty, unless a study evaluated how the ITE compared with oral board examination results. This study is consistent with the definition of non-human subjects research, therefore, no Institutional Review Board review was sought.

Results

Thirty-two articles were included in the final review, representing 21 medical specialties. National first-time pass rates for specialty board examinations are high across these specialties, ranging from 83% to 99% (TABLE 1). TABLE 2 includes a summary of the characteristics, results, and quality assessment of all studies included in our final analysis.

ITE Performance Data

The statistical analyses in the studies utilized a variety of quantification methods for ITE performance. Two studies (5%) grouped ITE performance into stanines (scaling of test scores on a 9-point scale with a mean

TABLE 1
Summary of Specialty ITEs and Board Examinations

Specialty	In-Training Examination		Board Examination	
	Creating Organization	Format, No. of Test Items, Interpretation of Score	Creating Organization	National Pass Rate
Allergy and immunology[a]	American Academy of Allergy, Asthma, and Immunology	200 MCQs	American Board of Allergy and Immunology	83%[39]
Anesthesiology	American Board of Anesthesiology	200 MCQs	American Board of Anesthesiology	95% (written) 89% (oral)[40]
Cardiovascular disease	American College of Cardiology	150 MCQs	American Board of Internal Medicine	96%[41]
Dermatology[a]	American Board of Dermatology	N/A	American Board of Dermatology	89.9%[42]
Emergency medicine[a]	American Board of Emergency Medicine	225 MCQs	American Board of Emergency Medicine	92% (written) 95% (oral)[43]
Endocrinology, diabetes, and metabolism[a]	Endocrine Society Center for Learning	90 clinical case vignettes	American Board of Internal Medicine	84%[41]
Family medicine	American Board of Family Medicine	240 MCQs	American Board of Family Medicine	98.6%[44]
Gastroenterology[a]	American Gastroenterological Association	180 MCQs	American Board of Internal Medicine	97%[41]
General surgery	American Board of Surgery	250 MCQs	American Board of Surgery	94%[45]
Geriatric medicine[a]	No ITE	No ITE	American Board of Internal Medicine	89%[41]
Hematology	American Society of Hematology	200 MCQs	American Board of Internal Medicine	91%[41]
Hematopathology	American Society for Clinical Pathology	MCQ NO	American Board of Pathology	96.4%[35]
Infectious disease	Infectious Diseases Society of America	150 MCQs	American Board of Internal Medicine	98%[41]
Internal medicine	American College of Physicians	300 MCQs	American Board of Internal Medicine	91%[41]
Medical genetics[a]	Medical Genetics Residency Program Directors	125 MCQs	American Board of Medical Genetics and Genomics	91%[46]
Medical oncology	American Society of Clinical Oncology	200 MCQs	American Board of Internal Medicine	90%[41]
Nephrology	American Society of Nephrology	150 MCQs	American Board of Internal Medicine	83%[41]
Neurology	American Academy of Neurology	400 MCQs	American Board of Psychiatry and Neurology	98%[47]
Neurological surgery[a]	No ITE	No ITE	American Board of Neurological Surgery	90.9% (written)[8] 82.5% (oral)[4]
Nuclear medicine[a]	American Board of Nuclear Medicine	N/A	American Board of Nuclear Medicine	87.7%[49]
Obstetrics and gynecology	Council on Resident Education in Obstetrics and Gynecology	397 MCQs	American Board of Obstetrics and Gynecology	82.6%[15]
Ophthalmology	American Academy of Ophthalmology	260 MCQs	American Board of Ophthalmology	87.8%[33]
Oral and maxillofacial surgery	American Board of Oral and Maxillofacial Surgery	250 MCQs	American Board of Oral and Maxillofacial Surgery	93%[50]

TABLE 1
Summary of Specialty ITEs and Board Examinations (continued)

Specialty	In-Training Examination		Board Examination	
	Creating Organization	Format, No. of Test Items, Interpretation of Score	Creating Organization	National Pass Rate
Orthopaedic surgery	American Academy of Orthopaedic Surgeons	275 MCQs	American Board of Orthopaedic Surgery	97% (written)[51] 93% (oral)[52]
Otolaryngology–head and neck surgery	American Board of Otolaryngology	300 MCQs	American Board of Otolaryngology	90%[30]
Pediatrics	American Board of Pediatrics	150 MCQs	American Board of Pediatrics	91%[53]
Physical medicine and rehabilitation[a]	American Academy of Physical Medicine and Rehabilitation	150 MCQs	American Board of Physical Medicine and Rehabilitation	94.6% (written) 96.9% (oral)[54]
Plastic surgery[a]	American Society of Plastic Surgeons	N/A	American Board of Plastic Surgery	91.3% (written) 93.6% (oral)[55]
Preventative medicine	American College of Preventative Medicine	110 MCQs	American College of Preventative Medicine	88.6%[7]
Psychiatry	American College of Psychiatrists	300 MCQs	American Board of Psychiatry and Neurology	89%[47]
Pulmonary and critical care	Association of Pulmonary and Critical Care Medicine Program Directors	150 MCQs	American Board of Internal Medicine	94% (pulmonary) 93% (critical care)[41]
Radiology diagnostic[a]	American College of Radiology	270 MCQs	American Board of Radiology	84%[56]
Radiation oncology[a]	American College of Radiology	450 MCQs	American Board of Radiology	99% (written)[57] 92% (oral)[58]
Rheumatology	American College of Rheumatology	200 MCQs	American Board of Internal Medicine	91%[41]
Sleep medicine[a]	American Academy of Sleep Medicine	N/A	American Board of Internal Medicine	95%[41]
Thoracic surgery[a]	Thoracic Surgery Directors Association	N/A	American Board of Thoracic Surgery	86% (written) 84% (oral)[59]
Urology	American Urological Association	180 MCQs	American Board of Urology	90%[60]
Vascular surgery[a]	American Board of Surgery	200 MCQs	American Board of Surgery	90% (written) 97% (oral)[61]

Abbreviations: MCQs, multiple-choice questions; N/A, not available; ITE, in-training examination.
[a] Specialty not included in review.

of 5 and standard deviation of 2), 14 studies (38%) used ITE absolute scores, 11 studies (30%) used ITE percentiles, and 10 studies (27%) used both absolute scores and percentile rank. A total of 16 studies used board examination pass/fail rates (43%), 13 studies (35%) used absolute or percentile board examination scores, and 8 (22%) used both absolute and percentile scores.

Relationship to Other Variables' Validity Evidence

About half of the studies (17, 53%) conducted a single type of statistical analysis to show evidence of relationship to other variables' evidence, 8 (25%) conducted 2 types of statistical analyses, 6 (18%) conducted 3 types of statistical analyses, and 1 (3%) conducted all 4 types of analyses. Nineteen studies used correlations, 12 used linear regressions, 18 used logistic regressions, and 6 used AUC values for the statistical analysis. Two studies reported sensitivity and specificity values, but did not provide an AUC value and thus were not include in the AUC category.

Forty-seven percent (9) of the 19 correlation studies found a strong relationship[2,7–14] between ITE performance and board examination performance for all residents and fellows in the respective study samples, and 1 found a moderate relationship (Withiam-Leitch

TABLE 2
Summary of Included Studies

Author, y	Specialty; Board	Subjects	Study Design and Methods	Types of Statistical Analysis	Strength of Correlation Results	Significant Linear Regression Results	Significant Logistic Regression Results	AUC Results for Maximized Sensitivity and Specificity	Quality Assessment Using MERSQI Score (Total Possible Score = 13.5)
Aeder, 2010	Pediatrics; American Board of Pediatrics	Pediatrics residents from 2002 to 2008 at an inner-city hospital in New York City (N = 207)	Retrospective cohort	Correlation	Strong positive relationship	N/A	N/A	N/A	8.5
Althouse, 2008	Pediatrics; American Board of Pediatrics	Evaluation of first-time board examination takers 2001–2005 (N = 14 525)	Retrospective cohort	Linear regression	N/A	Significant prediction, with prediction ability increasing with training year	N/A	N/A	8
Babbott, 2004	Internal medicine; American Board of Internal Medicine	4 internal medicine programs 2000–2002 (N = 170)	Retrospective cohort	AUC	N/A	N/A	N/A	Good accuracy (> 80%) of scoring below the 21st–23rd predicating failing	8
Bedno, 2011	Preventive medicine; American Board of Preventative Medicine	4 military programs 2002–2009 (N = 140)	Retrospective cohort	Correlation, linear regression	Strong positive relationship	Significant prediction controlling for other variables (grade point average, examination deferral)	N/A	N/A	8
Carey, 2014	Ophthalmology; American Board of Ophthalmology	Ophthalmology residents who completed their postgraduate training 1999–2011 (N = 41)	Retrospective cohort	Correlation, logistic regression, AUC	Moderate positive relationship for ITE 1 and ITE 3, strong for ITE 2	N/A	Scoring below 20th percentile on ITE 2 predicted failing; failing 1 ITE predicted failing	Good accuracy (> 80%) only for ITE 2	7
Collichio, 2016	Hematology or medical oncology; American Board of Internal Medicine	2008–2012 Hematology N = 1020 Oncology N = 1536	Retrospective cohort	Correlation, linear regression, logistic regression	Strong positive relationship	Significant prediction	Hematology and oncology ITE scores ''modestly'' predicted likelihood of passing	N/A	8
de Virgilio, 2010	General surgery; American Board of Surgery	Residents who graduated from 17 surgery programs 2000–2007 (N = 207)	Retrospective cohort	Logistic regression	N/A	N/A	Scoring below 35th percentile predicted failing	N/A	8
Dougherty, 2010	Orthopaedic surgery; American Board of Orthopaedic Surgery	Graduates of 4 residency programs in one geographic area from 1996 to 2009 (N = 202)	Retrospective cohort	Correlation	Strong positive relationship for average ITE, Strong for ITE 4, moderate for ITE 1–3	N/A	N/A	Scoring below 27th percentile predicted failing, but AUC value was not provided	8

TABLE 2
Summary of Included Studies (continued)

Author, y	Specialty; Board	Subjects	Study Design and Methods	Types of Statistical Analysis	Strength of Correlation Results	Significant Linear Regression Results	Significant Logistic Regression Results	AUC Results for Maximized Sensitivity and Specificity	Quality Assessment Using MERSQI Score (Total Possible Score = 13.5)
Ellis, 2000	Oral and maxillofacial surgery; American Board of Oral and Maxillofacial Surgery	Review of ITE scores among residents in last year of training 1992–1998 (N = 765)	Retrospective cohort	Correlation	Strong positive relationship for first to second attempt	N/A	N/A	N/A	8
Grabovsky, 2015	Infectious disease; American Board of Internal Medicine	Second-year infectious disease fellows 2008–2012 (N = 1021)	Retrospective cohort	Correlation, linear regression, logistic regression	Strong positive relationship	Significant prediction	ITE score as the only predictor had a modest R2 of 0.12	N/A	8
Indik, 2017	Cardiovascular disease; American Board of Internal Medicine	Third-year cardiovascular disease fellows 2011–2014 (N = 1918)	Retrospective cohort	Correlation, linear regression, logistic regression, AUC	Strong positive relationship	Significant prediction	Passing ITE on first attempt significantly predicted board examination scores	Scoring > 500 on the ITE predicted passing, AUC = 0.9	8
Johnson, 2010	Ophthalmology; American Board of Ophthalmology	Residents from 15 consecutive training classes at a single ophthalmologic residency 1991–2006 (N = 177)	Retrospective cohort	Logistic regression	N/A	N/A	Passing ITE in all 3 yes predicted of passing (5× more likely)	N/A	7
Jones, 2014	General surgery; American Board of Surgery	American Board of Surgery examinees 2006–2012 (N = 7372)	Retrospective cohort	Linear regression, logistic regression	N/A	Significant prediction for first- and fifth-year ITE scores	First- and fifth-year scores predicted outcome of passing	N/A	8
Jurich, 2018	Urology; American Board of Internal Medicine	Second-year nephrology fellows 2009–2014 (N = 1684)	Retrospective cohort	Linear regression, logistic regression	N/A	Significant prediction	ITE scores predicted outcome of passing	N/A	8
Juul, 2009	Psychiatry; American Board of Psychiatry and Neurology	American Board of Psychology and Neurology examinees 2002–2003 (N = 297)	Retrospective cohort	Correlation	Strong positive relationship for psychology, moderate for neurology	N/A	N/A	N/A	8
Juul, 2013	Neurology; American Board of Psychiatry and Neurology	2 cohorts of adult neurologists and 2 cohorts of child neurologists 2008–2009 (N = 982)	Retrospective cohort	Correlation	Strong positive relationship for PGY-1 and PGY-2, moderate for PGY-3	N/A	N/A	N/A	9

TABLE 2
Summary of Included Studies (continued)

Author, y	Specialty; Board	Subjects	Study Design and Methods	Types of Statistical Analysis	Strength of Correlation Results	Significant Linear Regression Results	Significant Logistic Regression Results	AUC Results for Maximized Sensitivity and Specificity	Quality Assessment Using MERSQI Score (Total Possible Score = 13.5)
Juul, 2013	Psychiatry; American Board of Psychiatry and Neurology	Psychiatry fellows (N = 342)	Retrospective cohort	Correlation	Strong positive relationship for year 1 fellows; moderate for year 2 fellows	N/A	N/A	N/A	8
Kay, 2015	Internal medicine; American Board of Internal Medicine	Single institution of internal medicine residents 2004–2012 (N = 183)	Retrospective cohort	Correlation, logistic regression	Strong positive relationship	N/A	Bottom quartile score in PGY-1, -2, or -3 predicted outcomes of failing	N/A	7
Kempainen, 2016	Pulmonary and critical care; American Board of Internal Medicine	First- and second-year fellows 2008–2012 Pulmonary N = 1484 Critical care N = 1331	Retrospective cohort	Linear regression, logistic regression	N/A	Significant prediction	ITE scores predicted outcome of passing	N/A	8
Kerfoot, 2011	Urology; American Board of Urology	US and Canadian residents who completed the ITE and participated in a sponsored online program; of the participants, 95% were from the United States 2008–2009 (N = 257)	Retrospective cohort	Correlation	Strong positive relationship	N/A	N/A	N/A	8
Kim, 2012	Anesthesiology; American Board of Anesthesiology	Anesthesiology residents at a single Midwestern residency program 1995–2007 (N = 97)	Retrospective cohort	Linear regression	N/A	Significant prediction	N/A	N/A	7
Klein, 2004	Orthopaedic surgery; American Board of Orthopaedic Surgery	Residents who graduated from a single program over a 10-year period (N = 65)	Retrospective cohort	Correlation	Strong positive relationship PGY-3, PGY-5, moderate PGY-4	N/A	N/A	N/A	7
Lingenfelter, 2016	Obstetrics and gynecology; American Board of Obstetrics and Gynecology	2 institutions of obstetrics and gynecology residents 2002–2012 (N = 80)	Retrospective cohort	Logistic regression, AUC	N/A	N/A	Score 200 in PGY-4 or 2 times predicted outcome of passing	Scoring > 195 in PGY-1 or -2, Scoring > 197 in PGY-1, > 201 in PGY-2, > 203 PGY-3, > 197 in PGY-4, AUC > 0.87	7.5

TABLE 2
Summary of Included Studies (continued)

Author, y	Specialty; Board	Subjects	Study Design and Methods	Types of Statistical Analysis	Strength of Correlation Results	Significant Linear Regression Results	Significant Logistic Regression Results	AUC Results for Maximized Sensitivity and Specificity	Quality Assessment Using MERSQI Score (Total Possible Score = 13.5)
Lohr, 2015	Rheumatology; American Board of Internal Medicine	Second-year rheumatology fellows 2008–2012 (N = 629)	Retrospective cohort	Linear regression, logistic regression	Strong positive relationship for fellow 1 and fellow 2 scores	Significant for fellow 2 scores (didn't include fellow 1 scores in model)	ITE scores fellow 2 predicted outcome of passing	NA	8
McClintock, 2010	Anesthesiology; American Board of Anesthesiology	Anesthesia trainees, split into 2 groups: (1) achieved certification on first attempt and (2) those who did not 2002–2004 (N = 2458)	Retrospective cohort	Linear regression, logistic regression	N/A	Significant prediction	ITE scores predicted outcome of passing	N/A	9
Monaghan, 2016	Hematopathology; American Board of Hematology	Hematopathology fellows 2009–2013 Fall N = 265 Spring N = 441	Retrospective cohort	Logistic regression	N/A	N/A	Spring test-takers: scores were a weak predictor of pass/fail outcome; fall test-takers: scores did not significantly predict pass/fail outcome	N/A	9
O'Neill, 2015	Family practice; American Board of Family Medicine	Family medicine residents in ACGME accredited programs. 2010–2013 (N = 6152)	Retrospective cohort	Correlation	Strong positive relationship	N/A	N/A	No AUC, but 91% correct predicting passing with 390 min score, but less predicted for people failing (47% correct)	8
Ponce, 2014	Orthopaedic surgery; American Board of Orthopaedic Surgery	Single orthopaedic surgery residency. 1999–2009 (N = 36)	Retrospective cohort	Correlation	Low for composite score and subsection scores except trauma (moderate), low, moderate, and high varied by year, but they did 52 correlation tests	N/A	N/A	N/A	7
Puscas, 2012	Otolaryngology; American Board of Otolaryngology–Head and Neck Surgery	Otolaryngology residents who took the board examination for the first time and the ITE in their final and penultimate years of training 2005–2011 (N = 1309)	Retrospective cohort	Logistic regression	N/A	N/A	ITE score in upper 3 quartiles for last PGY predicts passing outcome	N/A	8

TABLE 2
Summary of Included Studies (continued)

Author, y	Specialty; Board	Subjects	Study Design and Methods	Types of Statistical Analysis	Strength of Correlation Results	Significant Linear Regression Results	Significant Logistic Regression Results	AUC Results for Maximized Sensitivity and Specificity	Quality Assessment Using MERSQI Score (Total Possible Score = 13.5)
Puscas, 2019	Otolaryngology; American Board of Otolaryngology–Head and Neck Surgery	Otolaryngology residents who had taken the board examination for the first time, who had also taken the ITE 2007–2014 (N = 2214)	Retrospective cohort	Logistic regression, AUC	N/A	N/A	ITE score in top 6 stanines predicted pass outcome. Authors state "Due to the overall small number of failures despite the high number of examinees, the study is underpowered to analyze failure rate."	The AUC was 0.799, indicating good predictive ability of the model	8
Swanson, 2013	Orthopaedic surgery; American Board of Orthopaedic Surgery (ABOS)	Scores on at least one OITE test were located for 2852 (91%) of 3132 ABOS candidates who first took the board examination from 2002 to 2006 (N = 3132)	Retrospective cohort	Correlation, linear regression, logistic regression	Strong positive relationship for PGY-2, PGY-3, PGY-4, moderate PGY-1, low for PGY-0	ITE scores in PGY-3–4 significant prediction, PGY-1–2 didn't contribute significantly to the model	ITE score < 10th percentile predicted fail outcome relative to score > 50th percentile	N/A	8
Withiam-Leitch, 2008	Obstetrics and gynecology; American Board of Obstetrics and Gynecology	PGY-4 level ITE data at a single institution 1998–2005 (N = 69)	Retrospective cohort	Correlation, logistic regression, AUC	Moderate positive relationship	N/A	ITE scores predicted outcome of failing	Moderate accuracy (0.77) for ITE score 187.5	7

Abbreviations: N/A, not available; AUC, Area Under the Curve; ITE, in-training examination; PGY, postgraduate year; ACGME, Accreditation Council for Graduate Medical Education.
Note: Number of studies in each specialty: anesthesiology, 2; family practice, 1; general surgery, 2; hematopathology, 1; internal medicine (including subspecialties), 8; neurology, 1; obstetrics and gynecology, 2; opthalmology, 2; oral and maxillofacial surgery, 1; orthopaedic surgery, 4; otolaryngology, 2; pediatrics, 2; preventative medicine, 1; psychiatry, 2; urology, 1.

and Olawaiye, obstetrics and gynecology[15]) for all residents. The other 9 correlation studies found mixed results by postgraduate year (PGY) or specialty.[16–24] Eleven of the 12 studies using linear regression found that ITE scores significantly predicted board examination performance.[4,7,9,10,13,25–29] Only 1 study showing signicant prediction for PGY-3–PGY-4 residents, but not PGY-1–PGY-2 residents (Swanson et al, orthopaedic surgery[21]).

For logistic regression analysis, studies either used ITE scores as a predictor on a continuous scale or categorized ITE scores into 2 categories (eg, $<$ 10th percentile, $>$ 10th percentile). AUC analysis was used to determine the precision in prediction as a complement to logistic regression results or was done without logistic regression analysis. For predicting a board examination passing outcome, 6 studies showed ITE scores significantly predicted who would pass the board examination.[4,9,13,26,27,29] Three additional studies showed that a particular high score, quartile, or stanine significantly predicted who would pass the board examination (Pucas 2012, otolaryngology[30]), along with AUC good accuracy/predictive value (Lingenfelter et al, obstetrics and gynecology[31]; Pucas 2018, otolaryngology[32]). O'Neill et al (family medicine)[14] also found good AUC accuracy/predictive value for a particular high ITE score. Two additional studies showed that passing the ITE predicted passing the board examination (Johnson et al, ophthalmology[33]) with good AUC accuracy/predictive value (Indik et al, cariovascular disease fellows[10]).

For predicting a board examination failing outcome, 2 studies showed ITE scores significantly predicted who would fail the board examination (Swanson et al, orthopaedic surgery[21]), but only with a moderate AUC accuracy/predictive value (Withiam-Leitch and Olawaiye, obstetrics and gynecology[15]). Three studies showed that a particular low score or quartile significantly predicted who would fail the board examination (de Virgilio et al, surgery[3]; Kay et al, internal medicine[11]), with a good AUC accuracy/predictive value for PGY-2 and PGY-3 residents' ITE scores, but poor predictive value for PGY-1 ITE scores (Carey and Drucker, ophthalmology[16]). Babbott et al (internal medicine)[34] did not perform logistic regression and found good AUC accuracy/predictive value for a low quartile score. Only 1 study showed that failing an ITE significantly predicted failing the board examination (Carey and Drucker, ophthalmology[16]), but with a low positive predictive value and only applied to PGY-2 and PGY-3 residents' ITE scores. McClintock and Gravlee (anesthesiology)[29] applied a logistic regression to see how well the model predicted board examination fail/pass outcomes. The accuracy in prediction value was low-moderate for predicting a fail outcome and moderate-high for predicting a pass outcome. Finally, 2 studies found ITE scores had weak to no prediction for board examination pass/fail outcomes (Collichio et al, hematology and oncology[8]; Monaghan et al, hematology[35]). Additionally, Pucas (otolaryngology)[32] and O'Neill et al (family medicine)[14] were not able to predict who would fail the board examination based on their respective AUC analysis.

In terms of quality assessment of the articles included in this study, the average MERSQI score was 7.9 out of possible 13.5 points (range 7–9). This is within the range of reported MERSQI scores of medical education research more broadly.[36] All the included studies were retrospective cohorts; no studies were randomized controlled trials.

Discussion

This systematic review finds there is generally strong evidence that strong trainee performance on ITEs is predictive of subsequent passing performance on specialty board examinations. However, there is limited evidence that poor performance on the ITE predicts subsequent failure on board examinations, which calls into question the appropriateness of programs using the ITE to make high-stakes decisions. These results are important, as performance on ITEs has been widely accepted as predictive of subsequent performance on specialty board examinations, with pervasive beliefs that low-scoring residents are at risk of failing their board examination, resulting in some specialties reporting high-stakes use of ITE performance.

National first-time pass rates for specialty board examinations are high across specialties, which makes it difficult to predict trainees who will fail the examination (TABLE 1). In a cohort of otolaryngology residents, even those who scored in the bottom 3 stanines for each of the 4 years they took the ITE still had an 82% pass rate on their board examination.[32] If a nephrology program director simply predicted that all nephrology fellows would pass the nephrology board examination, they would be correct 89% of the time; using the ITE to make the same prediction, they would be correct 90% of time. This suggests that, despite correlations between ITE and board performance, prediction of board examination pass/fail using the ITE for an individual resident is of little practical benefit.[26] Even residents who perform very poorly on the ITE have a reasonable likelihood of passing their board examination.

The studies that did find a significant outcome of failing may not generalize to all trainees taking that particular ITE; thus, those results may only be useful

for the individual program since the studies that found a significant outcome of passing were more likely to use national samples of all residents and fellows. Additionally, since the number of trainees who fail an ITE is small, trying to accurately predict if all will end up failing their boards is statistically difficult since having just one of these trainees pass the board examination will greatly impact whether the outcome is significant. The number of trainees who pass the ITE is much larger so there is more wiggle room to accidently have a few fail the board examination and still find a significant outcome of predicting passing.

It is important to note the different formats of board examinations. Specialties including pediatrics, family practice, pathology, preventative medicine, neurology, internal medicine (and associated subspecialties), and psychiatry typically have 1 written examination that serves as the CE. Thus, evaluating the relationship between the ITE and CE in these fields may represent a more accurate comparison. Within surgical specialties, obstetrics and gynecology, ophthalmology, and anesthesiology there are 2 separate examinations. The QE is a written examination designed to evaluate knowledge in principles and applied science in a given specialty.[37] The CE among these specialties is an oral examination with the intent of evaluating a candidate's clinical judgement, reasoning skills, and problem-solving skills.[38] The ITE has limited ability to predict performance on oral board examinations. Additional tools that specifically assess application of knowledge and demonstration of clinical judgement in an oral format are needed to predict passage of oral CEs.

ITEs were originally developed as a formative assessment tool to assist learners and programs in identifying deficiencies in medical knowledge. Scores were meant to be used for no or low-stakes decisions and to guide development of individualized learning plans. To maintain the original intent of these examinations, further efforts at delineating "cut-scores" that predict board examination failure should not be undertaken. It remains similarly challenging to predict who will fail board examinations, with few studies designed to address this issue. Even if a significant fail outcome is found the predictive value is low. The paucity of data regarding ITE prediction of board examination failure suggests that program directors should exercise caution in the interpretation and use of low ITE scores at the individual resident level, particularly regarding high-stakes uses to inform formal academic actions (probation, repeating PGY, and requiring remediation) within a program. The majority of studies describe the use of ITE performance as low-stakes and formative for trainees or GME programs, with 2 (6%) studies in pediatrics and ophthalmology using the information for continuous program improvement.[2,33] Three studies (9%) in pediatrics and general surgery describe moderate to high-stakes use of ITE performance, including decisions regarding formal academic actions.[2–4] Finally, as expected, ITE performance increases with PGY. Therefore, when a resident is in their final year of training, when the correlations between ITE and board examination performance are strongest, it may be too late to help struggling residents "catch up" in time to pass board examinations.

This study has several limitations. First, the heterogeneity of the assessment instruments and specialties limited our ability to perform a pooled meta-analysis of the data. Furthermore, the studies included in this review vary in population size, from single institutions to a national review of how ITEs correlated with board examinations. There were also variations in study design, with some studies including data on interventions performed within a given residency versus large national data on how ITEs correlate with board examination scores. Future studies should involve national samples and investigate precision in predicting failing or passing board examinations utilizing other assessment data and contextual variables in addition to ITE scores.

Conclusions

This systematic review demonstrates that strong performance on ITEs is associated with passing subsequent board examinations, while the reverse is not necessarily true. Ultimately, this suggests that the GME community should continue to exercise caution and restraint in using ITE scores for moderate to high-stakes decisions.

References

1. Meissick S. Meaning and values in testing validation: the science and ethics of assessment. *Educ Res*. 1989;18(2):5–11.
2. Aeder L, Fogel J, Schaeffer H. Pediatric board review course for residents "at risk." *Clin Pediatr (Phila)*. 2010;49(5):450–456. doi:10.1177/0009922809352679.
3. de Virgilio C, Yaghoubian A, Kaji A, Collins JC, Deveney K, Dolich M, et al. Predicting performance on the American Board of Surgery qualifying and certifying examinations: a multi-institutional study. *Arch Surg*. 2010;145(9):852–856. doi:10.1001/archsurg.2010.177.
4. Jones AT, Biester TW, Buyske J, Lewis FR, Malangoni MA. Using the American Board of Surgery In-Training Examination to predict board certification: a cautionary

study. *J Surg Educ*. 2014;71(6):e144–e148. doi:10.1016/j.jsurg.2014.04.004.
5. Cohen J. A power primer. *Psychol Bull*. 1992;112(1):155–159. doi:10.1037//0033-2909.112.1.155.
6. Reed DA, Beckman TJ, Wright SM, Levine RB, Kern DE, Cook DA. Predictive validity evidence for medical education research study quality instrument scores: quality of submissions to JGIM's Medical Education Special Issue. *J Gen Intern Med*. 2008;23(7):903–907. doi:10.1007/s11606-008-0664-3.
7. Bedno SA, Soltis MA, Mancuso JD, Burnett DG, Mallon TM. The in-service examination score as a predictor of success on the American Board of Preventive Medicine Certification Examination. *Am J Prev Med*. 2011;41(6):641–644. doi:10.1016/j.amepre.2011.08.008.
8. Collichio FA, Hess BJ, Muchmore EA, Duhigg L, Lipner RS, Haist S, 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. doi:10.1007/s13187-016-0993-6.
9. Grabovsky I, Hess BJ, Haist SA, Lipner RS, Hawley JL, Woodward S, et al. The relationship between performance on the infectious diseases in-training and certification examinations. *Clin Infect Dis*. 2015;60(5):677–683. doi:10.1093/cid/ciu906.
10. Indik JH, Duhigg LM, McDonald FS, Lipner RS, Rubright JD, Haist SA, 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. doi:10.1016/j.jacc.2017.04.020.
11. Kay C, Jackson JL, Frank M. The relationship between internal medicine residency graduate performance on the ABIM certifying examination, yearly in-service training examinations, and the USMLE Step 1 examination. *Acad Med*. 2015;90(1):100–104. doi:10.1097/ACM.0000000000000500.
12. Kerfoot BP, Baker H, Connelly D, Joseph DB, Matson S, Ritchey ML. Do chief resident scores on the in-service examination predict their performance on the american board of urology qualifying examination? *J Urol*. 2011;186(2):634–637. doi:10.1016/j.juro.2011.03.126.
13. Lohr KM, Clauser A, Hess BJ, Gelber AC, Valeriano-Marcet J, Lipner RS, et al. Performance on the adult rheumatology in-training examination and relationship to outcomes on the rheumatology certification examination. *Arthritis Rheumatol*. 2015;67(11):3082–3090. doi:10.1002/art.39281.
14. 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.
15. Withiam-Leitch M, Olawaiye A. Resident performance on the in-training and board examinations in obstetrics and gynecology: implications for the ACGME Outcome Project. *Teach Learn Med*. 2008;20(2):136–142. doi:10.1080/10401330801991642.
16. Carey A, Drucker M. Standardized training examinations among ophthalmology residents and the American Board of Ophthalmology written qualifying examination first attempt: the Morsani College of Medicine experience. *J Acad Opthalmol*. 2014;7(1):e8–e12.
17. Dougherty PJ, Walter N, Schilling P, Najibi S, Herkowitz H. Do scores of the USMLE step 1 and OITE correlate with the ABOS part 1 certifying examination?: a multicenter study. *Clin Orthop Relat Res*. 2010;468(10):2797–2802. doi:10.1007/s11999-010-1327-3.
18. Ellis E 3rd, Haug RH. A comparison of performance on the OMSITE and ABOMS written qualifying examination. Oral and maxillofacial surgery in-training examination. American Board of Oral and Maxillofacial Surgery. *J Oral Maxillofac Surg*. 2000;58(12):1401–1406. doi:10.1053/joms.2000.18275.
19. Klein GR, Austin MS, Randolph S, Sharkey PF, Hilibrand AS. Passing the boards: can USMLE and orthopaedic in-training examination scores predict passage of the ABOS part 1 examination? *J Bone Joint Surg Am*. 2004;86(5):1092–1095.
20. Ponce B, Savage J, Momaya A, Seales J, Oliver J, McGwin G, et al. Association between orthopaedic in-training examination subsection scores and ABOS Part I examination performance. *South Med J*. 2014;107(12):746–750. doi:10.14423/SMJ.0000000000000208.
21. Swanson D, Marsh JL, Hurwitz S, DeRosa GP, Holtzman K, Bucak SD, et al. Utility of AAOS OITE scores in predicting ABOS Part I outcomes: AAOS exhibit selection. *J Bone Joint Surg Am*. 2013;95(12):e84. doi:10.2106/JBJS.L.00457.
22. Juul D, Schneidman BS, Sexson SB, Fernandez F, Beresin EV, Ebert MH, et al. Relationship between resident-in-training examination in psychiatry and subsequent certification examination performances. *Acad Psychiatry*. 2009;33(5):404–406. doi:10.1176/appi.ap.33.5.404.
23. Juul D, Flynn FG, Gutmann L, Pascuzzi RM, Webb L, Massey JM, et al. Association between performance on neurology in-training and certification examinations. *Neurology*. 2013;80(2):206–209. doi:10.1212/WNL.0b013e31827b92a3.

24. Juul D, Sexson SB, Brooks BA, Beresin EV, Bechtold DW, Lang JA, et al. Relationship between performance on child and adolescent psychiatry in-training and certification examinations. *J Grad Med Educ*. 2013;5(2):262–266. doi:10.4300/JGME-D-12-00088.1.
25. Althouse LA, McGuinness GA. The in-training examination: an analysis of its predictive value on performance on the general pediatrics certification examination. *J Pediatr*. 2008;153(3):425–428. doi:10.1016/j.jpeds.2008.03.012.
26. Jurich D, Duhigg LM, Plumb TJ, Haist SA, Hawley JL, Lipner RS, 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. doi:10.2215/CJN.05580517.
27. Kempainen RR, Hess BJ, Addrizzo-Harris DJ, Schaad DC, Scott CS, Carlin BW, 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):481–488. doi:10.1513/AnnalsATS.201601-015OC.
28. Kim PY, Wallace DA, Allbritton DW, Altose MD. Predictors of success on the written anesthesiology board certification examination. *Int J Med Educ*. 2012;3:225–235. doi:10.5116/ijme.5096.5d66.
29. McClintock JC, Gravlee GP. Predicting success on the certification examinations of the American Board of Anesthesiology. *Anesthesiology*. 2010;112(1):212–219. doi:10.1097/ALN.0b013e3181c62e2f.
30. Puscas L. Otolaryngology resident in-service examination scores predict passage of the written board examination. *Otolaryngol Head Neck Surg*. 2012;147(2):256–260. doi:10.1177/0194599812444386.
31. Lingenfelter BM, Jiang X, Schnatz PF, O'Sullivan DM, Minassian SS, Forstein DA. CREOG in-training examination results: contemporary use to predict ABOG written examination outcomes. *J Grad Med Educ*. 2016;8(3):353–357. doi:10.4300/JGME-D-15-00408.1.
32. Puscas L. Junior otolaryngology resident in-service exams predict written board exam passage. *Laryngoscope*. 2019;129(1):124–128. doi:10.1002/lary.27515.
33. Johnson GA, Bloom JN, Szczotka-Flynn L, Zauner D, Tomsak RL. A comparative study of resident performance on standardized training examinations and the American Board of Ophthalmology written examination. *Ophthalmology*. 2010;117(2):2435–2439. doi:10.1016/j.ophtha.2010.03.056.
34. 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. doi:10.1016/j.amjmed.2007.05.003.
35. Monaghan SA, Felgar RE, Kelly MA, Ali AM, Anastasi J, Bellara AP, et al. Does taking the fellowship in-service hematopathology examination and performance relate to success on the American Board of Pathology hematology examination? *Am J Clin Pathol*. 2016;146(1):107–112. doi:10.1093/ajcp/aqw085.
36. Cook DA, Reed DA. Appraising the quality of medical education research methods: the Medical Education Research Study Quality Instrument and the Newcastle–Ottawa Scale-Education. *Acad Med*. 2015;90(8):1067–1076. doi:10.1097/ACM.0000000000000786.
37. The American Board of Surgery. General Surgery Qualifying Exam. 2019. http://absurgery.org/default.jsp?certgsqe. Accessed November 28, 2020.
38. The American Board of Surgery. General Surgery Certifying Exam. 2019. http://www.absurgery.org/default.jsp?certcehome Accessed November 28, 2020.
39. The American Board of Allergy & Immunology. Statistics. 2020. https://www.abai.org/statistics_exam.asp. Accessed November 28, 2020.
40. American Board of Anesthesiology. 2017 Examination Results. http://www.theaba.org/PDFs/Presentations/2017-Exam-Results. Accessed April 13, 2019.
41. American Board of Internal Medicine. Initial Certification Pass Rates. 2018–2019. https://wfpww.abim.org/~/media/ABIM%20Public/Files/pdf/statistics-data/certification-pass-rates.pdf. Accessed November 28, 2020.
42. Hayag MV, Berman B, Weinstein A, Frankel S. American Board of Dermatology certification examination: preparation and perceptions. *Arch Dermatol*. 2002;138(4):544–546. doi:10.1001/archderm.138.4.544.
43. American Board of Emergency Medicine. Application and Exam Statistics. 2019. https://www.abem.org/public/resources/exam-certification-statistics. Accessed November 28, 2020.
44. American Board of Family Medicine. 2017 Family Medicine Certification Examination Pass Rates. https://www.theabfm.org/sites/default/files/2019-02/2017PassRates.pdf. Accessed November 28, 2020.
45. The American Board of Surgery. General Surgery Exam Pass Rates. 2018. http://www.absurgery.org/default.jsp?statgeneral. Accessed November 28, 2020.
46. American Board of Medical Genetics and Genomics. History of Pass Rates. 2011–2019. http://www.abmgg.org/pages/cert_passrates.shtml. Accessed November 28, 2020.
47. American Board of Psychiatry and Neurology. Pass Rates for First-time Takers. http://www.abpn.com/wp-content/uploads/2016/08/ABPN-Pass-Rates-5-Year-Summary.pdf. Accessed November 28, 2020.

48. The American Board of Neurological Surgery. Frequently Asked Questions 2020. https://abns.org/frequently-asked-questions/#faq-general-27b. Accessed November 28, 2020.
49. The Nuclear Medicine Technology Certification Board Inc. Annual Examination Report 2018. https://www.nmtcb.org/exam/annual-reports/Annual-Report-2018.pdf. Accessed November 28, 2020.
50. American Board of Oral and Maxillofacial Surgery. Stats/Fees/Timelines. https://www.aboms.org/become-certified/qualifying-examination/stats-fees-timelines. Accessed November 28, 2020.
51. American Board of Orthopaedic Surgery. Part I examination statistics. https://www.abos.org/certification-exams/part-i/exam-statistics.aspx. Accessed November 28, 2020.
52. American Board of Orthopaedic Surgery. Part II examination statistics. https://www.abos.org/certification-exams/part-ii/exam-statistics.aspx. Accessed November 28, 2020.
53. The American Board of Pediatrics. Initial Certifying Examination First-Time Taker Passing Rates. 2018. https://www.abp.org/sites/abp/files/pdf/exam-pass-rates-init-cert.pdf. Accessed November 28, 2020.
54. American Board of Physical Medicine and Rehabilitation. Exam statistics 2020. https://www.abpmr.org/Primary/Statistics. Accessed November 28, 2020.
55. The American Board of Plastic Surgery Inc. Statistics. https://www.abplasticsurgery.org/about-us/statistics/. Accessed November 28, 2020.
56. American Board of Radiology. Initial Certification for Diagnostic Radiology. Scoring and Results. https://www.theabr.org/diagnostic-radiology/initial-certification/core-exam/scoring-and-results. Accessed November 28, 2020.
57. American Board of Radiology. Initial Certification for Radiation Oncology. Qualifying Exam. https://www.theabr.org/radiation-oncology/initial-certification/the-qualifying-exam. Accessed November 28, 2020.
58. American Board of Radiology. Initial Certification for Radiation Oncology certifying (oral) exam. https://www.theabr.org/radiation-oncology/initial-certification/certifying-exam. Accessed November 28, 2020.
59. American Board of Thoracic Surgery. American Board of Thoracic Surgery Report to the Thoracic Surgery Directors Association. 2017. https://www.tsda.org/wp-content/uploads/2018/01/ABTS-Report-to-TSDA-10-12-17.pdf. Accessed November 28, 2020.
60. The American Board of Urology. Qualifying examination (part 1). http://www.abu.org/certification/qualifying-examination. Accessed November 28, 2020.
61. The American Board of Surgery. Vascular Surgery pass rates. 2019. http://publicdata.absurgery.org/PassRates/. Accessed November 28, 2020.

All authors are with the University of Utah School of Medicine. **Hilary C. McCrary, MD, MPH,** is an Otolaryngology Resident, Department of Surgery; **Jorie M. Colbert-Getz, PhD, MS,** is Assistant Dean, Education Quality Improvement, and Associate Professor, Department of Internal Medicine; **W. Bradley Poss, MD, MMM,** is Associate Dean, Graduate Medical Education, and Professor, Department of Pediatrics; and **Brigitte K. Smith, MD, MHPE,** is Vice Chair of Education, Program Director, Vascular Surgery Fellowship, and Assistant Professor, Department of Surgery.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

Corresponding author: Hilary C. McCrary, MD, MPH, University of Utah Hospital, hilary.mccrary@hsc.utah.edu

Received February 7, 2020; revisions received May 4, 2020, and October 13, 2020; accepted October 23, 2020.