

Exploring the Relationships Between USMLE Performance and Disciplinary Action in Practice: A Validity Study of Score Inferences From a Licensure Examination

Monica M. Cuddy, MA, Aaron Young, PhD, Andrew Gelman, PhD, David B. Swanson, PhD, David A. Johnson, MA, Gerard F. Dillon, PhD, and Brian E. Clauser, EdD

Abstract

Purpose

Physicians must pass the United States Medical Licensing Examination (USMLE) to obtain an unrestricted license to practice allopathic medicine in the United States. Little is known, however, about how well USMLE performance relates to physician behavior in practice, particularly conduct inconsistent with safe, effective patient care. The authors examined the extent to which USMLE scores relate to the odds of receiving a disciplinary action from a U.S. state medical board.

Method

Controlling for multiple factors, the authors used non-nested multilevel

logistic regression analyses to estimate the relationships between scores and receiving an action. The sample included 164,725 physicians who graduated from U.S. MD-granting medical schools between 1994 and 2006.

Results

Physicians had a mean Step 1 score of 214 (standard deviation [SD] = 21) and a mean Step 2 Clinical Knowledge (CK) score of 213 (SD = 23). Of the physicians, 2,205 (1.3%) received at least one action. Physicians with higher Step 2 CK scores had lower odds of receiving an action. A 1-SD increase

in Step 2 CK scores corresponded to a decrease in the chance of disciplinary action by roughly 25% (odds ratio = 0.75; 95% CI = 0.70–0.80). After accounting for Step 2 CK scores, Step 1 scores were unrelated to the odds of receiving an action.

Conclusions

USMLE Step 2 CK scores provide useful information about the odds a physician will receive an official sanction for problematic practice behavior. These results provide validity evidence supporting current interpretation and use of Step 2 CK scores.

In their mission statements, both the National Board of Medical Examiners (NBME) and the Federation of State Medical Boards (FSMB) frame their organizational responsibilities around notions of protecting the health of the public. As cosponsors of the United States Medical Licensing Examination (USMLE), both organizations specifically seek to ensure that patients in the United States receive high-quality health care from appropriately trained physicians. Arguably, one way to achieve this goal is to use USMLE standards to safeguard that practicing physicians possess the knowledge and skills necessary to provide safe and effective patient care. Indeed, for more than two decades, to receive an unrestricted license to practice allopathic medicine in the United States, physicians

have had to achieve passing scores on all USMLE components.

A licensing examination program such as the USMLE is resource intensive. Implicit in this investment is the assumption that USMLE standards protect the public by allowing only qualified physicians to independently practice medicine. Physicians who fail the USMLE are unable to obtain a license to practice medicine in the United States, thus precluding the possibility of establishing whether or not physicians who have met USMLE standards provide better patient care than those who have failed to meet these standards. While assessing differences in practice patterns between individuals who passed and individuals who failed the USMLE remains impossible, exploring the relationship between USMLE scores and physician performance in practice provides an alternate, though less direct, approach to examining how well achieving USMLE standards signals readiness to practice medicine.

scores can be extrapolated from the test setting to the real-world behaviors they are intended to represent, an aspect of a validity argument which relates to external criterion measures.^{1,2} To this end, a number of studies based on Canadian physicians relate Canadian medical licensing examination scores to criterion measures associated with subsequent performance in practice. For example, Tamblyn and colleagues^{3,4} demonstrated that scores from the Medical Council of Canada Qualifying Examination are associated with such practice behaviors as prescribing patterns and mammography screening rates for family physicians. Wenghofer and colleagues⁵ showed that licensing examination scores are positively related to peer assessments of the quality of care given by physicians. More recently, Norcini and colleagues⁶ uncovered a negative association, after accounting for other factors, between USMLE Step 2 Clinical Knowledge (CK) scores and patient mortality for U.S. physicians who attended international medical schools.

Experts in the field of educational assessment emphasize the importance of examining the extent to which inferences drawn from high-stakes examination

Other studies specifically focus on whether or not physicians' examination scores influence identification of problematic behavior in practice. For

Please see the end of this article for information about the authors.

Correspondence should be sent to Monica M. Cuddy, National Board of Medical Examiners, 3750 Market St., Philadelphia, PA 19104; telephone: (215) 590-9810; e-mail: mcuddy@nbme.org.

Acad Med. XXXX;XX:00–00.

First published online

doi: 10.1097/ACM.0000000000001747

Copyright © 2017 by the Association of American Medical Colleges

example, Papadakis and colleagues⁷ demonstrated a negative relationship between performance on a specialty certification examination in internal medicine and the risk of receiving a disciplinary action from a state licensing authority in the United States. In Canada, Tamblin and colleagues⁸ revealed a negative association between measures of communication and clinical decision making (as gleaned from a medical licensing performance-based assessment) and patient complaints to medical regulatory authorities for physicians licensed in Ontario or Quebec.

Overall, these studies provide valuable information about the connections between performance on medical licensing and specialty certification examinations and subsequent physician conduct; however, to our knowledge, no studies examine the associations between USMLE scores and troubling behavior in practice. To address this gap, we have conducted the present validity study to investigate the relationships between USMLE Step 1 and Step 2 CK scores and disciplinary actions taken by state medical boards for a national sample of graduates from MD-granting medical schools in the United States. Our approach for this study underscores the notion that, as a group, individuals who pass the USMLE are deemed minimally competent for unsupervised practice. Simultaneously, it presumes that scores among individuals who pass the USMLE and enter into practice reflect variation in the knowledge and skills required to effectively treat patients, with high-scoring individuals expected to be more qualified for practice than individuals obtaining moderate scores or scores just above the pass/fail standard.

Step 1 and Step 2 CK measure distinct, but related, constructs. Specifically, Step 1 is intended to assess whether an individual understands and can apply concepts related to the biomedical sciences that are fundamental to the practice of medicine. This examination ensures mastery of foundational science as well as the scientific principles required for maintenance of competence across the continuum of training and practice. Step 2 CK is intended to assess how well an individual understands and applies the medical knowledge, skills, and understanding of clinical science necessary for safe and effective patient

care under supervision. This examination focuses on both the principles of clinical sciences and the basic patient-centered skills that provide the foundation for the safe and competent practice of medicine. Both examinations assess more than medical knowledge and its applications, and each focuses additionally on practice-related principles and skills.

We posit that differences between individuals who excel on Step 1 and Step 2 CK and individuals who struggle to meet Step 1 and Step 2 CK requirements may be related to whether these individuals encounter challenges in practice that rise to the level of state medical board action. One reason physicians may be disciplined by a state medical board is that they provided substandard patient care due to a lack of clinical knowledge, which both Step 1 and Step 2 CK scores are intended to address. Substandard patient care could also be due to a misunderstanding or misuse of the scientific principles and patient-centered skills measured by Step 1 and Step 2 CK. Additionally, Step 1 and Step 2 CK scores may provide a proxy for personal characteristics associated with misconduct leading to disciplinary action. In other words, examinees with the self-discipline and motivation to achieve a high level of knowledge and skill may be less susceptible to the factors that lead to negative behaviors not directly linked to the practice of clinical medicine. Our primary research question is as follows: Are USMLE Step 1 and Step 2 CK scores related to the chance that a physician who graduated from a U.S. MD-granting medical school will receive a disciplinary action in medical practice after accounting for other factors?

Method

Data

We assembled our dataset by merging information from the databases of the NBME and FSMB. The resulting dataset included demographic variables, USMLE scores, practice-related information, and disciplinary action data. The initial sample consisted of 184,706 physicians who graduated from U.S. MD-granting medical schools between 1994 and 2006, thus representing 13 physician cohorts (we assembled the dataset in 2012, so the overall time span covered by the dataset includes disciplinary actions from 1994

to 2012). Of those physicians, 19,968 (about 11% of the initial sample) were missing information on their specialty area, and 13 (less than 0.1% of the initial sample) were missing information related to their jurisdiction of practice. These 19,981 cases were removed, and in the end we had complete data for 164,725 physicians. These physicians practiced in a range of specialty areas ($n = 16$) and licensing jurisdictions ($n = 51$, including Washington, DC). The study was reviewed by the American Institutes for Research Institutional Review Board and qualified for exempt status because it involved very minimal or no risk to study participants.

Variables

A disciplinary action taken by a state medical board represents the culmination of a medico-legal process usually originating from a complaint by a patient or a patient's family member followed by staff review, investigation by the board, and a hearing. A state medical board may receive 1,000 or more complaints annually,⁹ although the number of complaints ending in a public, punitive action is far smaller because of a multitude of factors (e.g., lack of a clear violation of state law or board regulation).¹⁰ The FSMB aggregates disciplinary data reported from state medical boards and other reporting agencies such as the Department of Health and Human Services as well as a growing number of international licensing authorities. The FSMB employs extensive quality control measures to ensure the accuracy of such disciplinary data. In our analyses, we examine only punitive disciplinary actions, such as an official reprimand or punishment, suspension of license, and revocation of license.

We used a binary variable indicating whether a physician had ever received a punitive disciplinary action from a state medical board (0 indicated no action, and 1 indicated at least one action) as our outcome measure. We used a binary variable rather than a continuous variable indicating the total number of actions received by a physician because accurately determining distinct actions proves challenging given that a single troublesome event can trigger multiple actions within and across jurisdictions.

The primary independent variables included Step 1 scores and Step 2 CK scores from physicians' first attempt on

each examination. Although currently an undifferentiated license to practice allopathic medicine in the United States cannot be obtained until physicians pass all four components of the USMLE, we focus specifically on Step 1 and Step 2 CK scores because examinees generally take Step 1 and Step 2 CK sequentially, and passing scores on these examinations are intended to signify readiness to practice medicine under supervision. Operationally, Step 1 and Step 2 CK scores are statistically adjusted to account for potential differences in the difficulty of the examination across examination forms and administration years. We analyze first-attempt Step 1 and Step 2 CK scores because the vast majority of graduates of U.S. medical schools pass Step 1 and Step 2 CK on their first attempt (more than 95%).

We also included a binary measure representing physician gender (0 indicated male, and 1 indicated female) as an independent variable since preliminary results indicated that the chance of receiving a disciplinary action varied by gender. Past state-specific studies have shown that male physicians are more likely to be disciplined in practice than female physicians.^{11,12} Moreover, a recent meta-analysis revealed that male physicians are more likely to be subject to disciplinary action across a variety of countries, the United States included.¹³ Lastly, we treated the number of years since medical school graduation as a covariate to account for the length of time that physicians had the opportunity to engage in behavior that could result in disciplinary action.

Two important dimensions categorize physicians in the United States: the specialty area in which they practice and the jurisdiction in which they are licensed. Past research focusing on a single state medical board has shown significant variation in the risk of being disciplined by medical specialty area.^{11,14} To our knowledge, no national studies specifically address potential differences in disciplinary action across licensing jurisdictions. However, given the extent of the variation among state medical boards themselves,^{15,16} as well as possible differences in physician performance by state, we believe that state-level effects may matter for understanding the disciplinary actions received by physicians practicing medicine in the

United States. In short, areas of practice differ in workload, setting, and patient population; state medical boards are diverse in terms of structure, resources, and authority; and all of these factors may cause variation in the distribution of disciplinary actions.

Analyses

Our data are not strictly hierarchically structured since all physicians licensed in a particular state are not practicing in the same specialty area, and all physicians practicing in a given specialty area are not licensed in the same state. Rather, this type of two-level data structure, sometimes referred to as cross-classified or non-nested,^{17,18} suggests that physician-level characteristics and the relationships between these physician-level characteristics and the chance of receiving a disciplinary action may vary in meaningful ways by specialty area and licensing jurisdiction. We employ non-nested multilevel modeling techniques to properly address this data structure.

We estimated physician-level effects (e.g., the impact of Step 1 scores on the [log] odds of receiving a board action), controlling for physician-level covariates and the effects of specialty and jurisdiction. For our analyses, we used physicians' first disciplinary action and its associated jurisdiction. More specifically, we fit a series of non-nested multilevel logistic regression models¹⁸ to the data to model the (log) odds of disciplinary action as a function of physician-level independent variables including Step 1 scores, Step 2 CK scores, gender, and years of exposure (i.e., years since medical school graduation). When considered separately, both Step 1 scores and Step 2 CK scores each made statistically

significant contributions to explaining the odds of receiving a disciplinary action. However, when both scores were included in the model together, the effect for Step 1 scores became nonsignificant, so we removed this variable from the model. Thus, the final model estimated the chance of receiving a disciplinary action using Step 2 CK scores, gender, and years of exposure with the intercept allowed to vary across both specialties and licensing jurisdictions.

For ease of interpretation, we converted Step 1 scores, Step 2 CK scores, and years of exposure to *z* scores, and we grand-mean centered gender. Thus, regression coefficients can be interpreted as the effect of a particular variable on the dependent measure at the average value of the other variables included in the model. Furthermore, given the use of *z* scores, a one-unit increase in an independent variable is akin to an increase of one standard deviation (SD) for that same independent variable. All statistical analyses were conducted using Stata, version 13 (StataCorp LP, College Station, Texas).

Results

Physicians in our sample had a mean Step 1 score of 214 (SD = 21) and a mean Step 2 CK score of 213 (SD = 23). Just under half of the physicians were female ($n = 74,148$; about 45%), and on average, exposure time was 12 years (SD = 4). Of the physicians included in the sample, 2,205 (1.3%) received at least one disciplinary action from a state medical board. Table 1 provides descriptive information for physicians who received a disciplinary action compared with those who did not. As shown, physicians who received a disciplinary action had

Table 1
Sample Characteristics for Physicians Receiving at Least One Disciplinary Action Compared With Physicians Receiving No Disciplinary Actions^a

Variable	Physicians receiving at least one disciplinary action, mean (SD)	Physicians receiving no disciplinary actions, mean (SD)
USMLE Step 1 score	205 (21)	214 (21)
USMLE Step 2 Clinical Knowledge score	202 (23)	213 (23)
Female	0.28 (0.45)	0.45 (0.50)
Years since medical school graduation	14 (3)	12 (4)

Abbreviation: USMLE indicates United States Medical Licensing Examination; SD, standard deviation.

^aThe authors have provided characteristics of 164,725 physicians graduating from U.S. MD-granting medical schools between 1994 and 2006. Of these, 2,205 (1.3%) received at least one disciplinary action and the remaining 162,520 (98.7%) received no disciplinary action.

Table 2
Frequency Distribution of Types of Disciplinary Action for 2,205 Physicians Sanctioned by State Medical Boards, 1994–2006^a

Disciplinary action type	No. (% of 7,929)
License restriction	1,267 (16.0)
Probation	895 (11.3)
Fine	830 (10.5)
Administrative action	829 (10.5)
Reprimand	796 (10.0)
Conditions imposed	739 (9.3)
Suspension	682 (8.6)
CME required	515 (6.5)
License revoked, surrendered, or denied	467 (5.9)
Other	909 (11.5)

Abbreviation: CME indicates continuing medical education.

^aSome physicians received more than one disciplinary action. The percentages provided are based on all of the actions the 2,205 physicians in the sample received (7,929 total actions).

lower average Step 1 and Step 2 CK scores. While just under half of the full sample of physicians were female, of those physicians who received a disciplinary action, only about 28% were women.

Table 2 provides descriptive information about the types of disciplinary actions that the physicians in our sample received. Because physicians can hold licenses in multiple jurisdictions, and sanctioned physicians often receive more than one action for a single offense, the 2,205 physicians in our sample who received at least one disciplinary action received in total 7,929 actions. Note that in our regression analyses we focus only

on initial actions. As illustrated in Table 2, actions ranged from restrictions against a license (n = 1,267; 16.0% of all actions) to a license revocation (n = 467; 5.9% of all actions). The total number of disciplinary actions an individual physician received ranged from 1 to 29, with 81% of sanctioned physicians receiving 5 or fewer actions (not shown in Table 2).

Table 3 presents the physician-level results of the final non-nested multilevel logistic regression analysis. Physicians with higher Step 2 CK scores have a lower chance of receiving a disciplinary action from a state medical board after controlling for other factors. More specifically, a 1-SD increase in Step 2 CK scores (approximately 23 score points) corresponds to a decrease in the chance of disciplinary action by about 25% (odds ratio of 0.75). With respect to physician gender, the chances of receiving an action for females are 45% less than they are for men, after accounting for other variables (odds ratio of 0.54). In terms of absolute effects (not shown in Table 3), the mean predicted probability of receiving a disciplinary action is 0.012 when an individual's Step 2 CK score is 190 (1 SD below the mean) and 0.008 when an individual's Step 2 CK score is 236 (1 SD above the mean).

Table 4 shows the results of two supplemental non-nested multilevel logistic regression models, which we estimated in an effort to demonstrate how we defined our final model, particularly in terms of our decision to remove Step 1 scores as an independent variable. Model 1 includes Step 1 scores, gender, and years of exposure as independent variables, while Model

2 includes Step 1 scores, Step 2 CK scores, gender, and years of exposure as independent variables. As shown, when Step 1 scores are included in a model without Step 2 CK scores, the Step 1 score effect is statistically significant with an expected 22% decrease (odds ratio of 0.78) in the chance of disciplinary action for every 1-SD increase in Step 1 scores. However, when both Step 1 and Step 2 CK scores are included in the model together, the Step 1 score effect becomes nonsignificant.

Discussion and Conclusions

One percent of physicians in our sample received at least one punitive disciplinary action from a state medical board. Although this percentage is clearly small, the actual number of physicians sanctioned for problematic behavior is significant—about 2,200—and the problems these physicians could create are potentially quite substantial given the number of patients physicians treat over the course of a career. Our findings indicate that USMLE Step 2 CK scores provide useful information for understanding the chances that physicians in the United States will receive a disciplinary action. Documenting a negative relationship between performance on Step 2 CK and subsequent performance in practice provides validity evidence in support of the intended interpretation and use of Step 2 CK scores. As mentioned, passing Step 2 CK is required to receive a license to practice allopathic medicine in the United States, and for most individuals with an MD degree, it is the final step in the credentialing process that allows them to practice medicine under supervision. The fact that higher scores are associated with lower odds of disciplinary action in a practice setting implies that Step 2 CK scores are a valuable tool for helping to ensure that entry into medical practice is restricted to individuals capable of providing safe and effective patient care.

Step 1 scores did not provide useful information for understanding disciplinary action above and beyond the information garnered by Step 2 CK scores. In this sense, Step 1 scores were essentially redundant in our analysis. Although Step 1 and Step 2 CK scores are related,¹⁹ the two examinations measure distinct proficiencies. As mentioned,

Table 3
Results of Non-Nested Multilevel Logistic Regression Analysis Predicting Disciplinary Action From a U.S. State Medical Board With Physicians Cross-Classified by Medical Specialty Area and Medical Licensing Jurisdiction^a

Standardized variable	B	SE β	e ^β (odds ratio)	95% CI
USMLE Step 2 Clinical Knowledge score	-0.29 ^b	0.03	0.75	0.70–0.80
Female	-0.62 ^b	0.05	0.54	0.49–0.59
Years since medical school graduation	0.44 ^b	0.02	1.56	1.48–1.59
Intercept	-4.68 ^b	0.12	0.009	0.007–0.012

Abbreviation: USMLE indicates United States Medical Licensing Examination; CI, confidence interval.

^aThe authors examined 164,725 physicians graduating from U.S. MD-granting medical schools between 1994 and 2006. The results of the likelihood ratio test comparing the null model nested in the final model are as follows: $\chi^2(3) = 475.91, P < .001$.

^bSignificant at $P < .001$.

Table 4

Results of Two Supplemental Non-Nested Multilevel Logistic Regression Models Predicting Disciplinary Action From a U.S. State Medical Board With Physicians Cross-Classified by Medical Specialty Area and Medical Licensing Jurisdiction^a

Standardized variable	Model 1		Model 2	
	Odds ratio	95% CI	Odds ratio	95% CI
USMLE Step 1 score	0.78 ^b	0.75–0.82	0.97	0.91–1.04
USMLE Step 2 Clinical Knowledge score	—	—	0.75 ^b	0.70–0.80
Female	0.49 ^b	0.45–0.55	0.53 ^b	0.48–0.59
Years since medical school graduation	1.60 ^b	1.53–1.68	1.55 ^b	1.48–1.62

Abbreviations: USMLE indicates United States Medical Licensing Examination; CI, confidence interval.

^aThe authors examined 164,725 physicians graduating from U.S. MD-granting medical schools between 1994 and 2006. The results of the likelihood ratio test comparing model 1 nested in model 2 are as follows: $\chi^2(1) = 72.71, P < .001$.

^bSignificant at $P < .001$.

Step 1 establishes command of basic science material, and Step 2 CK measures proficiency in essential clinical knowledge and skills. Step 2 CK scores reflect skills such as clinical reasoning and clinical judgment, which may in turn be more congruent with characteristics that later result in troublesome behavior. However, successful completion of both examinations requires similar motivation and training. As such, if licensure required mastery of the content of one examination, but not the other, scores might not remain related in the same way. Furthermore, the completion of Step 2 CK usually occurs close to the time at which a physician enters practice; thus, Step 2 CK scores may better reflect behavior in practice compared with Step 1 scores.

Our study has several limitations. First, we defined exposure as the length of time since graduation from medical school. Although this is a reasonable approximation of when physicians begin practice under supervision, in some cases it may not necessarily indicate when a physician began residency training. Second, the specialty information used in our analysis relates to the area in which physicians were first board certified. For some physicians, certification status was unavailable because of variation in credentialing timelines for some specialty areas (e.g., orthopedic surgery). In these cases (about 7% of the total sample), we determined specialty area by using a self-reported measure included on the USMLE Step 3 application form, which does not provide any indication of certification status.

It is possible that the relationships between USMLE scores and the chance of receiving a disciplinary action in practice vary depending on the type of offense committed and the type of action received due to that offense. For this initial study, we have investigated only associations among Step 1 scores, Step 2 CK scores, and the odds of having received at least one disciplinary action. Future research should attempt to specify how USMLE scores relate to poor patient care because of a lack of clinical knowledge or skills compared with how they relate to disciplinary actions involving major professional issues such as substance abuse, failure to recognize appropriate boundaries, or fraudulent billing practices.

Furthermore, a disciplinary sanction by a state medical board represents a high threshold as an external criterion measure. Future study of USMLE scores relative to a lower threshold (e.g., official complaints to [not sanctions by] state medical boards) may provide additional information for understanding the associations between USMLE scores and patient-related outcomes. Lastly, useful future research could focus on understanding the relationships between other USMLE components (e.g., Step 2 Clinical Skills scores; Step 3 scores) or measures (e.g., last score, best score, number of examination attempts) and disciplinary action.

Our study provides useful information supporting one validity argument relating external, real-world criterion measures to Step 2 CK scores. Specifically, in terms of the ability to better understand disciplinary action in a practice setting

for a national sample of physicians from U.S. MD-granting medical schools, our findings corroborate the intended interpretation and use of Step 2 CK scores. We note that, in this context, less evidence supports the use of Step 1 scores when Step 2 CK scores are known. Moreover, our findings add to the relatively small body of work^{7,8} examining relationships between scores on medical licensing and certification examinations and subsequent disciplinary actions in practice.

Our study is, to the best of our knowledge, the first to focus on the extent to which USMLE scores relate to performance in practice as indicated by official sanctions from a state medical board for questionable conduct for physicians who graduated from U.S. MD-granting medical schools. The central mission of the NBME and FSMB is to protect the health of the public, particularly through high-stakes, high-quality assessments. Our finding that, on average, physicians with higher Step 2 CK scores are less likely to receive a disciplinary action from a state medical board offers support for the use of USMLE Step 2 CK scores in accomplishing this mission. As such, our study provides unique information related to the aspect of validity dealing with external criterion measures for USMLE scores. Additionally, it adds to the literature related to physician assessments and performance practice more generally.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: This study was reviewed by the American Institutes for Research Institutional Review Board and qualified for exempt status.

M.M. Cuddy is measurement scientist, Center for Advanced Assessment, National Board of Medical Examiners, Philadelphia, Pennsylvania.

A. Young is assistant vice president, Research and Data Integration, Federation of State Medical Boards, Eules, Texas.

A. Gelman is professor, Departments of Statistics and Political Science, Columbia University, New York, New York.

D.B. Swanson is vice president, Academic Programs and Services, American Board of Medical Specialties, Chicago, Illinois, and professor (honorary appointment), Department of Medical Education, University of Melbourne Medical School, Melbourne, Australia.

D.A. Johnson is senior vice president, Assessment Services, Federation of State Medical Boards, Eules, Texas.

G.F. Dillon is vice president, Licensure, National Board of Medical Examiners, Philadelphia, Pennsylvania.

B.E. Clauser is vice president, Center for Advanced Assessment, National Board of Medical Examiners, Philadelphia, Pennsylvania.

References

- 1 Messick S. Validity. In: Linn RL, ed. *Educational Measurement*. 3rd ed. New York, NY: Macmillan; 1989:13–103.
- 2 Kane MT. Validation. In: Brennan RL, ed. *Educational Measurement*. 4th ed. Westport, CT: American Council on Education/Praeger; 2006:17–64.
- 3 Tamblyn R, Abrahamowicz M, Brailovsky C, et al. Association between licensing examination scores and resource use and quality of care in primary care practice. *JAMA*. 1998;280:989–996.
- 4 Tamblyn R, Abrahamowicz M, Dauphinee WD, et al. Association between licensure examination scores and practice in primary care. *JAMA*. 2002;288:3019–3026.
- 5 Wenghofer E, Klass D, Abrahamowicz M, et al. Doctor scores on national qualifying examinations predict quality of care in future practice. *Med Educ*. 2009;43:1166–1173.
- 6 Norcini JJ, Boulet JR, Opalek A, Dauphinee WD. The relationship between licensing examination performance and the outcomes of care by international medical school graduates. *Acad Med*. 2014;89:1157–1162.
- 7 Papadakis MA, Arnold GK, Blank LL, Holmboe ES, Lipner RS. Performance during internal medicine residency training and subsequent disciplinary action by state licensing boards. *Ann Intern Med*. 2008;148:869–876.
- 8 Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *JAMA*. 2007;298:993–1001.
- 9 North Carolina Medical Board. North Carolina Medical Board 2015 Annual Report. Raleigh, NC: North Carolina Medical Board; 2016. https://www.ncmedboard.org/images/uploads/disciplinary_reports/Annual_Report_2015_-final.pdf. Accessed March 8, 2017.
- 10 Medical Board of California. 2014–2015 Annual Report Medical Board of California. Sacramento, CA: Medical Board of California; 2016. http://www.mbc.ca.gov/Publications/Annual_Reports/annual_report_2014-2015.pdf. Accessed March 8, 2017.
- 11 Morrison J, Wickersham P. Physicians disciplined by a state medical board. *JAMA*. 1998;279:1889–1893.
- 12 Clay SW, Conatser RR. Characteristics of physicians disciplined by the State Medical Board of Ohio. *J Am Osteopath Assoc*. 2003;103:81–88.
- 13 Unwin E, Woolf K, Wadlow C, Potts HW, Dacre J. Sex differences in medico-legal action against doctors: A systematic review and meta-analysis. *BMC Med*. 2015;13:172.
- 14 Kohatsu ND, Gould D, Ross LK, Fox PJ. Characteristics associated with physician discipline: A case-control study. *Arch Intern Med*. 2004;164:653–658.
- 15 Federation of State Medical Boards. U.S. Medical Regulatory Trends and Actions. Eules, TX: Federation of State Medical Boards; 2014. http://www.fsmb.org/Media/Default/PDF/FSMB/Publications/us_medical_regulatory_trends_actions.pdf. Accessed March 8, 2017.
- 16 Horowitz R. In the Public Interest: Medical Licensing and the Disciplinary Process. New Brunswick, NJ: Rutgers University Press; 2013.
- 17 Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Thousand Oaks, CA: Sage Publications; 2002.
- 18 Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, NY: Cambridge University Press; 2007.
- 19 Cuddy MM, Swanson DB, Dillon GF, Holtman MC, Clauser BE. A multi-level analysis of the relationships between selected examinee characteristics and USMLE Step 2 CK performance. *Acad Med*. 2006;81(10 suppl):S103–S107.