Medical education

Correlation of scores for the Comprehensive Osteopathic Medical Licensing Examination with osteopathic medical school grades

STEVE E. HARTMAN, PhD
BRUCE P. BATES, DO
SARAH A. SPRAFKA, PhD

The authors evaluated construct validity of scores for the Comprehensive Osteopathic Medical Licensing Examination (COMLEX–USA), the examination used to evaluate osteopathic physicians for licensure. They computed correlations between students' grades in the first 2 years of osteopathic medical school and their scores on the COMLEX–USA Level 1 (N = 187) and Level 2 (N = 86), as well as correlations between third- and fourth-year clerkship grades and the COMLEX–USA Level 2. Correlations of Level 1 scores with grades for years one, two, and the first 2 years combined were .74, .80, and .81, respectively; for Level 2, correlations were .59, .70, and .71. Correlation between clerkship grades and scores for the COMLEX–USA Level 2 was .26. The strong correlation between COMLEX–USA results and grades for the didactic curriculum in the first 2 years of medical school provides evidence for the construct validity of scores for the COMLEX–USA Levels 1 and 2.

(Key words: Comprehensive Osteopathic Medical Licensing Examination, medical education)

The examination used by the National Board of Osteopathic Medical Examiners (NBOME) to evaluate newly trained osteopathic physicians for licensure is the Comprehensive Osteopathic Medical Licensing Examination (COMLEX–USA), Levels 1 through 3. In this preliminary report, we summarize data pertaining to the construct validity of scores for Levels 1 and 2 of the COMLEX–USA.

Dr Hartman is a professor in the Department of Anatomy, College of Osteopathic Medicine, University of New England, Biddeford, Maine, where Dr Bates is chairman in the Department of Family Medicine and Dr Sprafka is Director of Predoctoral Education.

Correspondence to Steve E. Hartman, Department of Anatomy, College of Osteopathic Medicine, University of New England, Biddeford, ME 04005.

E-mail: *shartman@mailbox.une.edu*

We agree with Wright and Bezruczko[1] that statistical affinities of a medical licensing examination with measures of academic aptitude taken before enrollment in medical school (for example, undergraduate GPAs and MCAT scores) are irrelevant to the validity of any medical licensing examination—though they are useful for assessing predictive validity of preadmission measures. Therefore, we focused instead on correlations between COMLEX–USA scores (for Levels 1 and 2) and performance measures for the 4 years of osteopathic medical curriculum at the University of New England College of Osteopathic Medicine (UNECOM).

Methods

We compiled scores for the COMLEX–USA examinations administered to students who began UNECOM's program in fall 1996 and fall 1997 (Level 1; N = 187; 44% female) and 1994 and 1995 (Level 2; N = 86; 40% female). All students signed consent forms for release of the relevant data. For students who took the Level 1 or Level 2 examination more than once, we used scores from their first sitting. To measure academic performance in the first 2 years of UNECOM's curriculum, each student's average percentage grade was computed in three ways. First, we calculated the average grade for courses related directly to the basic medical sciences (gross anatomy, neuroanatomy, embryology, histology, virology, parasitology, immunology, and pathology). Second, we calculated the average grade for courses with a strong clinical emphasis—our second-year "systems" courses (nervous, musculoskeletal, respiratory, cardiovascular, hematology, gastroenterology, dermatology, renal, and reproductive systems). Finally, we calculated the average grade for all of these courses combined. All courses were weighted equally.

Some students had not been required to take all courses (for example, some had completed one or more equivalent courses before matriculating). For these students, we computed an average grade for those courses completed at UNECOM. Most students we evaluated had completed all of their relevant coursework at UNECOM. Even those students who did not complete all coursework at UNECOM completed most such work at the college.

To monitor clinical proficiency through the third and fourth years, following each clerkship a supervisor assessed each student's competence using a five-point scale to rate 12 criteria. For each student who had taken the COMLEX–USA Level 2, we measured clinical performance using their average grade over all 12 criteria and all 15 clerkships.

If our students' scores on the COMLEX–USA Level 1 are a valid measure of their mastery of "the medical concepts and principles [that is, medical basic sciences] necessary for understanding the mechanisms of medical problems,"[2] the product-moment correlation between

those scores and academic performance during the first 2 years in medical school should be high. Likewise, if scores on the COMLEX–USA Level 2 are a valid measure of students' mastery of "medical concepts and principles necessary for making appropriate medical diagnoses,"[2] the correlation between those scores and grades for years 1 and 2 should be high.

Because COMLEX–USA poses questions in the context of medical diagnosis and treatment ("clinical presentations and physician tasks"[2]), we predicted that scores for both Levels 1 and 2 would be more highly correlated with grades for our second-year, more clinically oriented courses than with first-year grades. Also, to the degree that (1) third- and fourth-year clerkship grades measure the cognitive component of clinical competence, and (2) scores on the COMLEX–USA Level 2 are valid, the former should correlate with the latter.

Results

Correlations between scores on the COMLEX–USA Level 1 and grades for first-year basic-science courses, second-year clinical systems, and all courses combined were .74, .80, and .81, respectively. Correlations between Level 2 and these grade measures were .59, .70, and .71, respectively. As predicted, scores for both Levels 1 and 2 appear to be correlated more strongly with grades for the clinically focused second-year curriculum than with those for the first-year curriculum. The correlation between scores on the COMLEX–USA Level 2 and clerkship grades was .26 (N = 86).

Comments

In descriptions of the COMLEX–USA, the NBOME specifies that the purpose of the Level 1 examination is to "assess a candidate's knowledge of the basic sciences."[3] Administrators and faculty at UNECOM, as part of our college's mission, have worked to develop a curriculum offering the basic science background necessary for students' success as practicing physicians. Therefore, one indication of construct validity for scores of the COMLEX–USA Level 1 would be correspondence between our students'

scores and their grades in the first 2 years of medical school.

Although no similarly derived correlations with regard to COMLEX–USA Level 1 scores have been published, our values (.74 to .81) compare favorably with those associated with the United States Medical Licensing Examination Step 1 (.62 and .73, N = 162[4]; .72 to .83 with a median of .76, N values not provided[5]) and are consistent with construct validity for scores of the COMLEX–USA Level 1 examination. At .59 to .71, correlations of scores on the COMLEX–USA Level 2 with grades were moderate and, likewise, suggestive of construct validity.

Many students scoring at the low ends of the grade scales scored higher on the COMLEX–USA Level 1 than would have been predicted—assuming that a linear relationship exists between these measures. This may reflect lack of agreement between these measures of academic aptitude in what is arguably the most important region of the COMLEX–USA Level 1 scale—where pass/fail decisions are rendered. If this observation is corroborated following examination of other and larger samples, an explanation should be sought.

As predicted, however, scores on the clinically focused COMLEX–USA Levels 1 and 2 both were more highly correlated with second-year grades (more reflective of clinical understanding) than first-year grades. This may be partly due to the second-year curriculum being nearer in time to the administration of both COMLEX–USA examinations. Also, the COMLEX–USA scores and medical school grades of UNECOM students are not entirely independent measures because several members of our faculty have written both medical school examinations and items for the COMLEX–USA. However, COMLEX–USA item writers are asked to submit only items never used at their home college, and more than 99% of COMLEX–USA items reported on here were written by individuals unaffiliated with our medical school, as noted in a conversation with Boyd Buser, DO, chairman of the NBOME Product Committee (October 6, 1999).

The correlation between clerkship grades and COMLEX–USA Level 2 may have been low in part because several clerkship evaluation items related more to student behavior than to cognitive skills (for example, "ability to work with other healthcare professionals"). However, consideration of only those items clearly related to cognitive mastery did little to improve the correlation. This, in turn, may have occurred because of the restricted range of clerkship grades—supervisors tended to use only the upper registers of the five-point scale (unsatisfactory, marginal, average, above average, outstanding). Also, these grades may have been influenced by factors that have uncertain relationships to actual clinical competence. For example, clerkship grades were correlated (r = .24; N = 101) with students' position on the extraversion-introversion scale of the Myers-Briggs Type Indicator[6]; on average, extraverts were awarded higher grades than were introverts. This correlation is not high, but it is as high as the correlation that exists between clerkship grades and the COMLEX–USA Level 2 (r = .26). This contributes to the impression that clerkship grades are measuring a broader conceptual universe than is the COMLEX–USA Level 2, which is primarily a measure of cognitive mastery.

Conclusion

Because medical schools set different internal standards,[5] it is important that psychometrically sound licensing examinations be in place to assure professional integrity and public well-being. One expectation with regard to scores on such examinations is that they will correlate strongly with students' performance in their medical school curriculum. The fact that correlations are high between our students' scores on the COMLEX–USA Levels 1 and 2 and their grades for the first 2 years of medical school is evidence for construct validity of the COMLEX–USA.

Hartman et al • Medical education

Acknowledgment

References

1. Wright BD, Bezruczko N. *Evaluation of COMLEX–USA*. Chicago, Ill: MESA Psychometric Laboratory; 1999.

2. *Bulletin of Information*. Chicago, Ill: National Board of Osteopathic Medical Examiners, Inc; 1999.

3. *Level 1: Examination guidelines and sample exam, 1999*. Chicago, Ill: National Board of Osteopathic Medical Examiners, Inc; 1999.

4. Elam CL, Johnson MS. NBME Part I versus USMLE Step 1: predicting scores based on preadmission and medical school performances. *Acad Med* 1994;69:155.

5. Swanson DB, Ripkey DR, Case SM. Relationship between achievement in basic science coursework and performance on 1994 USMLE Step 1. 1994-95 Validity Study Group for USMLE Step 1/2 Pass/Fail Standards. *Acad Med* 1996;71:S28-S30.

6. Hammer AL. *MBTI applications: A decade of research on the Myers-Briggs Type Indicator*. Palo Alto, Calif: Consulting Psychologists Press; 1996.

Review

This Won't Hurt a Bit! And Other Fractured Truths in Healthcare

By Karyn Buxman, RN. 96 pp. Lamoine Press, 63 Pioneer, Suite 200, Hannibal, MO 63401. 2000. $11.95.

Karyn Buxman, RN, has put together a unique and entertaining series of stories related to healthcare. In the foreword of her book, Buxman states her belief that laughter is the sweetness that helps the medicine go down. Presenting stories and jokes she has collected over many years, the author suggests their use as a way of reducing stress. I believe that she accomplishes that goal.

Vice-president and contributing editor for the *Journal of Nursing Jocularity* until the publication ceased, Buxman uses constructive—rather than destructive—humor to relieve stress and perhaps improve body physiology. There are numerous positive physiologic changes associated with laughter, including improved respiratory and cardiovascular system function, relaxed muscles, elevated mood, and possibly a boost to the immune system. In the group setting, the author points out that humor can solidify ties, improve productivity, and serve as a powerfully persuasive form of communication.

Reading this book made me laugh and improved my mood after a tough day. In this context, the humor should not be offensive and is helpful in relieving stress. I recommend this book without hesitation.

The author can be contacted at *www.humorx.com.*

Gilbert E. D'Alonzo, DO
Editor in Chief