Canadian Medical Education Journal

Brief Report

Comparison of Student Performance on Internally Prepared Clerkship Examinations and NBME Subject Examinations

Pamela Veale, Wayne Woloschuk, Sylvain Coderre, Kevin McLaughlin, and Bruce Wright

University of Calgary, Canada

Published: September 30, 2011

CMEJ 2011, 2(2):e81-e85 Available at http://www.cmej.ca

Abstract

Background: This pilot study compared performance of University of Calgary students on internal clerkship examinations with corresponding National Board of Medical Examiners (NBME) subject examinations.

Methods: Between April and October 2007, students completed internal and NBME subject examinations following six mandatory rotations. Local faculty within each discipline set the minimum performance level (MPL) for internal examinations. Two methods of standard setting were considered for NBME exams and a sensitivity analysis was performed. Corresponding internal and NBME examination scores were compared using McNemar's discordant pair analysis.

Results: A significant and unexpected difference in failure rate between internal and external examinations was found in all clerkships. 1.4% of students were below the MPL for internal examinations and 27.3% (modified Angoff) or 25.9% (mean Hofstee compromise) ($p < 0.001$ for both) for the NBME. The proportion of students below MPL for internal examinations was also below the lower limit of the Hofstee compromise (14.4%).

Conclusion: Possible explanations include leniency bias in internal standard setting, discrepant content validity between local curriculum and NBME examinations, difference in student perception of examinations, and performance bias due to unfamiliar units.

Correspondence: Dr. Pamela Veale, Assistant Dean, Pre-clerkship, Faculty of Medicine, University of Calgary, 3330 Hospital Drive NW, Calgary, Alberta, Canada, T2N 4N1; Tel: (403) 220-3917; Fax: (403) 270-2681.

Introduction

In 2007, the University of Calgary Undergraduate Medical Education Program considered use of externally prepared examinations as a part of certifying student assessment in clerkship rotations. The National Board of Medical Examiners (NBME) Subject Examinations were chosen for this pilot. The primary drive to consider the external examinations was to reduce demand for faculty time to prepare high quality internal examinations. As described in the NBME Subject Examination Program Information Guide (NBME 2009),

> "The subject tests have at least four distinct advantages over locally constructed examinations in the assessment of student achievement. First, the tests provide national norms and relevant descriptive information. Second, considerable care is taken in preparing these materials, with items selected only after extensive review and pretesting. Third, these tests concentrate heavily on application and integration of knowledge rather than recall of isolated facts. Finally, they attain better accuracy of measurement."

There is a paucity of information available on the application of NBME examinations to our particular context. The NBME exams are "designed to measure students' understanding of the clinical sciences…and to be broadly appropriate as part of overall student assessment"[1]. However, the NBME also recommends that, "congruence between subject exam content and course objectives should be considered when interpreting test scores and determining grading standards."[1]

Data released to medical schools by NBME include scores of American and Canadian students, but Canadian students are in the minority of this group. There are differences in the use of NBME examinations between American and Canadian medical schools. For example, in a survey done in 2000[2] it was found that more than 90% of American schools use the Surgery Subject examination compared with 50% of Canadian schools. This same survey revealed a wide range in weighting of the NBME as a component of the final rotation evaluation (5 - 60%) and that minimum passing scores varied from 51 to 70. A similar survey in 2005[3] revealed that 69% of Canadian and American psychiatry clerkships used the NBME Subject examination and that the weighting of the score toward the final rotation grade varied.

Prior to implementing the NBME examinations as an element of our certifying clerkship assessments, we engaged in a pilot study to compare student performance on local internal examinations with corresponding NBME subject examinations. We expected that the scores on the NBME examinations would produce similar pass/fail rates as our local examinations.

Methods

Comparison between internal examination scores and NBME scores

Between April and October 2007, all students with mandatory clerkship rotations in Family Medicine, Internal Medicine, Obstetrics and Gynecology, Pediatrics, Psychiatry, and Surgery completed both the certifying internal clerkship examinations and the relevant NBME subject examinations. Scores on the NBME subject examinations were not considered part of the formal student grades during the study period. The internal examinations were completed on the last day of the relevant rotation while NBME examinations were administered during the final week.

The internal examinations were prepared and standards set by local faculty in each discipline. The NBME releases recommendations to medical schools for standard setting of the student scores. We used cut off scores from both the modified Angoff method and the Hofstee compromise. Failure rates between internal and external examinations were compared using McNemar's discordant pair analysis.

Content validity

Internal examination blueprints were obtained for each of the 6 disciplines included in the project. The NBME *Keyword Phrase Item Analysis Report* was used to identify content tested in the subject examinations. Each "keyword phrase" was compared with the internal examination blueprints and assigned to one of three categories:

1. Clear match with blueprint category.

2. Clear match with content elsewhere in the curriculum.
3. Clear mismatch of content or insufficient information to accurately match.

One rater did the initial comparison. A second rater performed a reliability check by repeating the process. Both individuals are highly informed of local curriculum content. Any items of disagreement between the raters were placed in category 3.

The project was considered to be a Quality Assurance project by the Conjoint Health Research Ethics Board (CHREB) at the University of Calgary, meaning that formal ethical approval was not required.

Results

Each clerkship had between 51 and 66 students who took both the local examination and the corresponding NBME examination. In each rotation, there was a significant correlation between scores on the internal and NBME examinations, but the mean score on the NMBE examination was always lower. All students who were unsatisfactory in the internal exams were also unsatisfactory in the corresponding NBME examinations. Use of the Angoff method or the mean Hofstee cut-off resulted in a significantly lower pass rate for all disciplines, and the minimum Hofstee cut-off resulted in a lower pass rate for four of six clerkships. (Table 1).

Of the six clerkship disciplines, four had sufficiently detailed internal examination blueprints based on clinical presentations to enable reasonably specific content matching (Family Medicine, Internal Medicine, Obstetrics and Gynecology, and Psychiatry). Blueprints for Pediatrics and Surgery were based on broad categories or subspecialty discipline groupings rather than clinical presentations.

As shown in Table 2, between 50 and 92% of the NBME keyword phrases had a clear match to an internal blueprint in the corresponding discipline. However, if content covered elsewhere in the curriculum is included, the match increases to 82-98%.

Family Medicine had the lowest clear match between NBME keyword phrases and the internal examination blueprint. However, this discipline would be expected to have broad sampling across all medical subject areas and more likely to be defined differently in the United States and Canada. This is supported by the finding that 98% of the NBME keyword phrases matched to content in either the Family Medicine rotation or elsewhere in the local curriculum.

Pediatrics had the highest clear match between NBME keyword phrases and the internal examination blueprint. This was due to the nature of the blueprint. It was based on subspecialty disciplines (e.g. Emergency Medicine, Cardiology, etc.) and thus most keywords were easily fit into such broad categories.

Discussion

The comparison of internal examinations with NBME subject examinations revealed an unexpected pattern that was consistent across all clerkship rotations. Failure rates on internal examinations were lower than on the corresponding NBME subject examinations, even when the minimum Hofstee cut-off values were used. This difference persisted in four of the six clerkships. A similar result was found at the University of Utah,[4] namely that inclusion of the NBME Subject Examination as a component of the cumulative surgery clerkship evaluation (weighted 10%) “helped 11% but hurt 64%” of the students.

We hypothesized a number of factors to explain this outcome, including local leniency bias, breach of examination security for local examinations, poor student effort for NBME examinations, discrepant content validity between the examinations, and performance bias due to use of different units for laboratory values. The last element will be addressed in a future study.

The performance difference observed was pervasive across all of the clerkships despite examination preparation and standard setting by different local faculty groups, suggesting a systematic issue rather than a problem with individual discipline examinations. Additionally, while the overall failure rate on internal examinations (1.4%) is low, it is consistent with other performance measures, such as MCC results and surveys of residency program

Table 1. Performance of University of Calgary Clerks on Local Versus NBME Examinations

	Mean Score (standard deviation)		Correlation (Local vs. NBME)	Pass Rate (%)			
Clerkship	**Local Exam**	**NBME**		**Local Exam**	**NBME (Angoff)**	**NBME (minimum Hofstee)**	**NBME (mean Hofstee)**
Family Medicine n=66	75.8 (7.0)	65.3 (7.7)	0.46	100	86.4[b]	87.9[b]	78.8[c]
Internal Medicine n=51	77.0 (8.9)	69.1 (6.8)	0.68	94.1	78.4[a]	78.4[a]	66.7[c]
Obstetrics & Gynecology n=63	74.7 (6.3)	64.9 (7.2)	0.52	100	88.9[b]	96.8	84.1[b]
Pediatrics n=59	69.7 (6.9)	65.9 (7.8)	0.54	93.2	62.7[c]	83.0	76.3[b]
Psychiatry n=59	83.8 (4.8)	66.8 (7.7)	0.56	100	76.3[c]	91.5[a]	78.0[c]
Surgery n=59	75.6 (5.7)	64.1 (8.3)	0.51	100	71.8[c]	85.9[b]	80.3[c]

[a]$p<0.05$ for comparison of pass rate with local examination

[b]$p<0.01$ for comparison of pass rate with local examination

[c]$p<0.001$ for comparison of pass rate with local examination

Table 2. Results of Content Matching

Clerkship Discipline	NBME keyword phrases with clear match to internal exam blueprint (%)	NBME keyword phrases with clear match to content elsewhere in the curriculum (%)	Clear mismatch of content or insufficient information (%)
Family Medicine	50	48	2
Internal Medicine	63	19	18
Obstetrics & Gynecology	86	11	3
Pediatrics	92	5	3
Psychiatry	72	17	11
Surgery	83	12	5

directors,[5] indicating that our students are performing at or above their peers from other Canadian institutions.

Consequently, the high failure rate of our students on the NBME examinations may be due to factors more directly related to the NBME examinations themselves. The NBME examinations are not blueprinted specifically to our clerkship objectives or learning experiences: this raises concerns about the content validity of the NMBE examinations when applied to our curriculum. Our 3-year program is organized by clinical presentations,[6] thus different from most other medical schools that contribute data on student performance and input to standard setting for NBME examinations. Given that we found a significant amount of content in each subject examination matched to content elsewhere in the curriculum rather than in the corresponding clerkship rotation, it may be more appropriate for our institution to use NBME examinations at the end of training rather than as individual rotation evaluations.

Future Direction/Implications for Medical Education

These findings have elicited internal concerns regarding our standard setting and examination security, thus a systematic review of internal examination preparation and quality assurance steps was performed. The possibility of a performance bias and reduced content validity of the NBME examinations when applied to our curriculum, have halted a plan to use the NBME examinations as part of the certifying clerkship evaluations. Until we clarify factors responsible for the observed difference in performance, NBME examinations for certifying purposes in clerkships should remain "on the shelf".

Acknowledgements

The authors wish to acknowledge the unanimous support for this project by our clerkship chairs, evaluation representatives, and support staff. The authors would like to thank Lynley Matthews for assistance in preparation of this manuscript.

References

1. National Board of Medical Examiners Subject Examination Program Information Guide –copyright 2007. [Accessed October 2009, at www.nbme.org].
2. Lind DS, Deladisma AM, Cue JI, Thomas AM, MacFadyen BV, Nesbit RR. Survey of student education in surgery. *J Am Coll Surgeons* 2000;204(5):969-974.
3. Levine RE, Carlson DL, Rosenthal RH, Clegg KA, Crosby RD. Usage of the National Board of Medical Examiners Subject Test in Psychiatry by U.S. and Canadian Clerkships. *Acad Psych.* 2005;29(1):52-57.
4. Hermanson B, Firpo M, Cochran A, Neumayer L. Does the National Board of Medical Examiners' Surgery Subtest level the playing field? *Am J Surg.* 2004; 188:520-521.
5. Woloschuk W, McLaughlin K, Wright B. Is undergraduate performance predictive of postgraduate performance? *Teach Learn Med.* 2010;22(3):202-204.
6. Mandin H, Harasym P, Eagle C, Watanabe M. Developing a "clinical presentation" curriculum at the University of Calgary. *Acad Med.* 1995;70(3):186-193.