Journal of Contemporary Medical Education

available at www.scopemed.org, www.jcmed.edu

Original Research

Predictive validity of an in-house pediatric examination towards the NBME Pediatrics Subject examination

Hassib Narchi

United Arab Emirates University, United Arab Emirates

Received: July 02, 2013

Accepted: October 12, 2013

Published Online: November 30, 2013

DOI: 10.5455/jcme.20131012102910

Corresponding Author:
Hassib Narchi
Department of Pediatrics, College of Medicine & Health Sciences, United Arab Emirates University, Al Ain, P.O. Box 17666, United Arab Emirates,
hassib.narchi@uaeu.ac.ae

Keywords: *Assessment, Clinical clerkship, Educational measurement*

ABSTRACT

Analyze the predictive validity of the scores obtained by sixth year medical students in our institution in the in-house summative pediatric examination (IHE) and in each of its individual components at the end the pediatric clerkship in relation to scores obtained in the National Board for Medical Examiners Pediatrics Subject Examination (NBME-P). In a cohort of 152 students over five academic years (August 2004 - July 2009) we analyzed the correlation and the predictive validity of the former (and of each of its individual assessment tools) for the latter. In the univariate model, the overall IHE scores ($p<0.001$), the oral examination ($p<0.001$), the non-clinical skills ($p<0.001$), the academic year ($p=0.01$) and the clinical examination ($p=0.04$) were significantly correlated with and had a positive linear relationship with the NBME scores. In the multivariable analysis, only the oral examination remained significantly correlated with the NBME-P scores ($p<0.001$). The predicted NBME scores can be calculated from the students' oral assessment score as follows: Predicted NBME-P scores = 25 marks + (1.4 x oral score). Having the best predictive validity, the oral examination in the IHE can identify which students are at risk of obtaining low NBME-P marks.

INTRODUCTION

Assessment of students' learning is vital as it determines their achievement of intended learning outcomes and also provides feedback on their progress and on the teaching methods. The chosen assessment must be aligned with the educational objectives of the teaching and learning activities in that specific part of the curriculum. At the end of each clerkship, most institutions require the students to take an in-house summative examination in the respective specialty. The available wide range of assessment tools reflects that none adequately and comprehensively evaluates, on its own, all the learning outcomes expected from medical students [1,2].

At the College of Medicine and Health Sciences (CMHS), United Arab Emirates (UAE) University, at the end of their pediatric clerkship, sixth year medical students sit an in-house summative examination (IHE) as well as the American National Board for Medical Examiners (NBME) Pediatric Subject Examination (NBME-P).

The IHE consists of: 1- a faculty-observed clinical examination testing student's professionalism, history-taking, diagnosis and treatment of common pediatric problems. 2- A structured oral examination using ten clinical vignettes testing student's skills in the interpretation of clinical photographs, radiographs, electrocardiogram's, laboratory data, ordering investigations and planning treatment. 3- an evaluation of non-clinical skills based on the documentation in the logbook of attendance at teaching sessions, attitude and professionalism as evaluated by faculty, participation in ward rounds, case discussions, in addition to the evaluation of three written case presentations.

The NBME is an American, independent, organization which develops nationwide examination which permits medical licensing authorities to evaluate candidates for licensure [3]. It assesses what students have learned about normal development and the entire range of organ systems diseases as well as important tasks carried out by physicians, such as health promotion and maintenance, understanding mechanisms of disease, establishing a diagnosis and applying principles of management. Emphasis is on application of knowledge rather than recall of isolated facts and the questions are framed in the context of clinical vignettes. The NBME-P consists of one-best-answer (A-type) and extended-matching (R-Type) multiple choice questions (MCQ).

A few studies have assessedthe correlation and predictive value of the IHE to the NBME-P scores with conflicting results. Some have reported low quality of the IHE compared to the NBME [4-7]. Others have shown that IHE results are predictive of NBME Subject Examination scores [8,9]. In addition, very few of these studies have specifically addressed or analyzed the role of each separate assessment tool used in the IHE, nor the possible confounders.

Our research objective was therefore to analyze the predictive validity of the scores obtained by the students in the pediatric IHE, as well in each of its individual components, in relation to the scores obtained at the NBME-P. Identifying which parts of the IHE better predict the marks obtained at the NBME-P and the factors behind such correlation would help to identify those students at risk of obtaining low NBME-P scores in order to provide them with advice and support in preparation for that examination. The results would also help avoiding unnecessary duplication of assessment methods.

METHODS

This retrospective study, over five consecutive academic years (from August 2004 to July 2009), involved a cohort of 152 sixth-year medical students enrolled in the pediatric clerkship at the CMHS, UAE University. Approval for the study and waiving of consent was granted by the Human Research Ethics Committee of our institution (protocol 09/34) and accepted by the University of Dundee. Participants' anonymity was preserved. The work was carried out in accordance with the Declaration of Helsinki, with no potential harm to participants whose anonymity was also guaranteed.

Students' name, university number, dates of pediatric clerkship and examination scores in both the IHE and the NBME-P were extracted from the examination records. In order to find a 10% difference between the 2 examination results with a 5% precision and 95% confidence intervals, the sample size required for the study was 140 students (Epi Info 6.04, C.D.C. and W.H.O).

The response variable was the NBME-P score. The explanatory variables included the IHE marks, including the separate components of this examination, as well as the co-variables student's gender and clerkship academic year. As students' age and the interval between the clerkship and the examinations were identical for all the students as these examinations were always held at the end of each clerkship, they were not analyzed.

As the NBME-P scores had a maximum of 100 marks, we converted the overall IHE maximum score (90 marks) to become also 100 marks to enable comparison of both examinations. The relationship and correlation between the individual components of the IHE and the NBME-P marks were examined using univariate correlation studies. As only approximately a third of students had documentation of the individual scores for the components of their clinical examination (history, examination, discussion), these could not, therefore, be separately analyzed as independent or explanatory variables. Student's gender and year of examination were also analyzed in a univariate model. Since the examination scores did not follow a Normal distribution the Kruskal Wallis test was used to compare continuous variables between independent groups. The Spearman rank order correlation test was used to evaluate the strength of the relationship between the in-house assessment scores and the NBME-P scores.

As we were interested not merely by the correlation, but more importantly by the IHE to estimate and predict the NBME-P score, we used a multivariable linear regression model. The saturated model included the continuous explanatory variables (examination results) and the categorical variables (gender and academic year) if they were, in the univariate analysis, significantly associated with the NBME examination with a p value <0.1. In a backward stepwise selection regression model, all the variables correlating with a p value <0.1 in the univariate model with the NBME-P results were initially included. The variables with the non-significant and largest p value in this saturated model were removed from the model, one by one, repeating the analysis until only variables with a p value <0.05 remained and were also significant predictors in that model. The regression coefficients were calculated, as well as the p value for the association with NBME-P scores and the coefficient of determination R2 of the model.

All statistical analyses were carried out with the software package Stata version 9.0 (Stata Corp, College Station, Texas, USA). Statistical significance was defined by a 2-tailed p value <0.05.

RESULTS

There were 118 females (78%) and the difference in the gender distribution was statistically significant (p<0.001). The results of the in-house and NBME-P are detailed in Table 1. The mean scores were significantly higher in the IHE (83.9) than the NBME-P (64.2) (paired t-test p<0.01). Although there were no significant differences between genders in the total score of the IHE and the NBME-P scores, girls had a better score in the oral component of the IHE (p=0.006). Although they slightly outperformed boys in all other components of the IHE (except for the non-clinical skills component) and the NBME-P, the difference was not statistically significant. There was a significant fluctuation in the scores of all examinations throughout the five academic years (p<0.05) but no significant trend were observed, except for the progressive decrease in the clinical examination scores throughout that period (p<0.01).

The overall IHE marks were significantly correlated with the NBME-P results in both the single regression model (p<0.001, R2 0.16) and the Spearman rank order test (rho=0.43, p<0.001) as shown in Table 2. As the overall IHE score is the sum of several assessments, in order to avoid co-linearity (non-independence), we only included, in a single linear regression model, the individual components of the IHE as explanatory variables deliberately omitting the overall IHE. The oral examination (rho 0.52, p<0.001), the non-clinical skills (attendance and presentation) scores (rho 0.33, p<0.001), the academic year (rho 0.20, p=0.01) and the clinical examination (rho 0.16, p=0.04) were, in descending order of correlation strength, significantly correlated with the NBME-P results, with a positive linear relationship between each variable on one side and the NBME scores on the other. In the single linear regression model only the oral examination (p<0.001) and the academic year (p<0.001) remained significantly correlated with the NBME-P scores.

In the saturated multiple linear regression model all variables found to be significant with a p value <0.1 in the univariate analysis were entered as explanatory variables (Table 3-A). The variables statistically significantly associated with the NBME-P results were, by descending order of significance: oral examination results (p<0.001), academic year (p<0.001) and clinical examination scores (p=0.037). Only the results of the non-clinical skills marks (attendance and presentation) were not statistically significant (p=0.1). With multiple stepwise backward regression analysis the only remaining significant variables were, by descending order of significance: the oral examination (p<0.001) with a linear regression coefficient of 1.24 (95% CI 1.03, 1.8), the academic year (p=0.001) and the clinical score (p=0.03, coefficient 0.4) with these variables explaining together 37% of the observations (Table 3-B). After adjusting for the oral and clinical examination scores in the multivariable analysis, the correlation between each academic year and the NBME-P varied between years, but as the role of the academic year varied significantly from year to year and cannot be predicted for future years, we removed it from the final model (Table 3-C). In that model, only the oral examination score remained significantly associated with the NBME-P score (regression coefficient 1.4, SE 7.9, 95% CI 1.0 and 1.79, p<0.001) with a constant (slope) of 25.1. However, the model explained well only 26% of the results (R2 = 0.26). Based on the linear regression model, with a constant coefficient (intercept) value of 25.1, the predicted NBME scores can, therefore, be simply calculated for each student from his/her oral assessment score by using the formula: Predicted NBME-P scores = 25 marks + (1.4 x oral score).

Table 1. Comparison of the median value of examinations scores between genders and academic years.

	Number of students (%)	Clinical (50)	Oral (30)	Non-clinical skills (10)	Total in-house examination (calculated for a total of 100)	NBME (100)
All students	152 (100)	40.7	25.3	9.5	83.9	64.2
Gender						
Males	34 (22)	40.1	24.2	9.6	82.3	62
Females	118 (78)	41	26.5	9.4	84.8	64
*** p value**		0.5	0.06	0.8	0.1	0.4
Academic year						
2004-2005	8 (5)	44	27.1	8.2	85.5	70.0
2005-2006	59 (39)	42.5	26.5	9.7	85.9	61.0
2006-2007	35 (23)	41.4	23.5	9.5	83.3	64.0
2007-2008	30 (20)	38.3	25.2	10	81.1	71.0
2008-2009	30 (13)	39.2	27.6	10	85.6	72.0
***p value**		<0.01	<0.01	<0.01	0.006	<0.001

*Kruskal-Wallis test

Table 2. Single linear regression model correlation between the score of the in-house examination, its separate components, the academic year and the NBME score

	Regression coefficient	Standard error	p value	Coefficient of determination R^2	Spearman rank order correlation test RhoP value	
Total in-house examination	0.62	0.11	<0.001	0.16	0.43	<0.001
Clinical score	0.35	0.2	0.08	0.02	0.16	0.04
Oral score	1.42	0.19	<0.001	0.26	0.52	<0.001
Non-clinical skills	1.72	1.03	0.09	0.02	0.33	<0.001
Academic year						
2004-05	Reference value		<0.001	0.13	0.20	0.01
2005-06	-8.4	3.5				
2006-07	-6.2	3.6				
2007-08	-4.4	3.7				
2008-09	1.3	3.8				

Table 3. Multiple linear regression model of the relationship between the score of the components of the in-house examination, the academic year and the NBME score.

	Regression coefficient	Standard error	P value	Coefficient of determination R^2
A. Saturated model containing all the explanatory variables				
Clinical score	0.4	0.2	0.037	R^2=0.38
Oral score	1.18	0.2	<0.001	
Non-clinical skills	1.7	1.06	0.1	
Academic year			<0.001	
2004-05	Reference value			
2005-06	-10.1	3.6		
2006-07	-5.2	3.7		
2007-08	-3.6	3.9		
2008-09	-1.8	4.1		
Constant	-4339	1270	0.001	
B. Model including only the statistically significant explanatory variables				
Clinical score	0.4	0.2	0.03	R^2=0. 37
Oral score	1.2	0.2	<0.001	
Academic year			0.001	
2004-05	Reference value			
2005-06	-6.8	3.0		
2006-07	-2.1	3.1		
2007-08	-0.1	3.3		
2008-09	1.9	3.4		
Constant	25.1	7.9	0.002	
C. Model including only the clinical and the oral examination				
Clinical score	0.09	0.17	0.6	R^2=0. 26
Oral score	1.4	0.2	<0.001	
Constant	25.1	7.9	0.002	

DISCUSSION

Students' score at the NBME-P can be predicted from their results in the IHE. When used as a formative assessment, the latter can identify the students at risk of obtaining low NBME-P marks and offer them remedial solutions.

Although many components of the IHE were significantly correlated with the NBME, only the oral examination and the academic year remained significant in the linear regression model, probably because that relationship was non-linear. The lack of correlation of some of the IHE individual components with the NBME-P is probably related to the differences in the learning domains and educational objectives tested by each.

Many studies confirm our findings that the gender of medical students has no relationship to performance in medicine clinical clerkship, NBME examinations or USMLE, nor influences the confidence or competence during an obstetrics and gynecology clerkship [10-13]. Other studies, however, have shown that female students perform better in obstetrics and gynecology clerkship and NBME examination, communication skills and undergraduate psychiatry examinations [14-16]. We are unaware of any similar studies carried out in pediatric examinations and are, therefore, unable to comment on our findings in a pediatric clerkship.

The finding that academic year was independently predictive of the examination scores also confirm other reports [9]. As during the study period no changes occurred in the pediatric curriculum, teaching and assessment methods, this finding reflects the differences in learning aptitudes and abilities of the students of each academic year, probably a reflection of the selection criteria for admission to medical school.

Students' attendance at teaching sessions had no correlation with either of the two examinations. This contrasts with a study where students' attendance at the basic medical sciences classes had a direct impact on their performance in examinations including the NBME Part I [17,18]. It is possible that in basic sciences classes and in part I of the NBME examination, where knowledge is the main educational domain, regular attendance to classes enhances students' knowledge. In contrast, the educational domains covered in the non-clinical in-house assessments involve attitude and professionalism, where improving performance may not necessarily correlate with attendance.

As our students are all Emirati nationals, we could not evaluate, in this study, the role of nationality. Language is also an important factor as academic success in a medical course where the curriculum is in English, such as ours, is influenced by competence in English [19,20]. However in our medical school, as all students speak English as a second language; this may affect their performance in learning and in the examinations. However, we believe this to be unlikely, as admission criteria include proficiency in the English language. In addition, as both the examinations are in English, language was not a variable we have studied. Performance in secondary school and pre-clinical years might be considered important factors in medical student's performance in examinations. However, we decided to exclude them as they had already been shown to have no correlation with the marks obtained in the examinations throughout the clinical years in our medical school [21].

We believe that the varying correlation between the individual components of the in-house and the NBME examinations may have several reasons. While relatively low quality of IHE compared to the NBME was found in some reports [4-7] other studies have shown that in-house assessments are predictive of NBME Subject Examination scores [8,9] and residents' competence correlates with their NBME results [22-24]. We believe that these varying results are due to the different educational objectives assessed in each examination. NBME examinations are designed as licensure examinations and not as academic achievement tests. Their highest precision is reached at the decision of passing or failing, but scores deviating from this pass/fail mark are only useful when combined with other methods of assessment [25].

A limitation of the study is that we could not separately analyze the scores of students' attendance at teaching sessions or the case presentations, nor each component of the in-house clinical examination (history taking, physical examination, etc.) and this should be addressed in future studies. As this project involved a single institution with single nationality students all having Arabic as their mother tongue, similar studies should be replicated in other medical schools with multinational students to test the generalizability of the findings. As, except for the clinical examination, the affective and the psychomotor learning domains were not adequately assessed, more research is needed to develop a wider range of more comprehensive tools to assess these two domains.

As prediction is one of the major roles of assessment, the predictive value of these measurements in relation to the outcomes of medical education, such as postgraduate professional training and on-the-job performance still requires further study as equivocal conclusions have been drawn [26]. Systematic reviews in medical education should include not only the outcomes used to evaluate medical education, but also if the measures obtained during medical school can predict them [27].

CONCLUSION

We still need to use a wide range of assessment tools to evaluate our students' learning, as unlike the IHE, the NBME-P explores the whole range of organ systems. The marks obtained in-house oral examination when used in a formative assessment would help develop solutions to improve the performance of those students predicted to obtain low NBME-P marks.

ACKNOWLEDGMENTS

The author wishes to thank Dr. John McAleer, Senior Lecturer, for his guidance and support for this project which was part of a research project for a Master's degree in Medical Education, Centre for Medical Education, University of Dundee, United Kingdom. The author is also indebted to Professor Michelle Mclean, Dr. Sami Shaban and Dr. Sulaiman Al Hammadi, at the Faculty of Medicine and Health Sciences, United Arab Emirates University, for allowing access to the medical students' marks.

DECLARATION OF INTEREST

The authors report no declarations of interest.

FUNDING/SUPPORT:

None

REFERENCES

1. Epstein RM. Assessment in medical education. New Engl J Med 2007;356(4):387-96.
2. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet 2001;357(9260):945-49.
3. Hubbard JP, Levit EJ. Early history of NBME. The National Board of Medical Examiners: The First Seventy Years - A Continuing Commitment to Excellence. Philadelphia, 1985.
4. Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew RH. The quality of in-house medical school examinations. Acad Med 2002;77(2):156-61.
5. Magarian GJ, Mazur DJ. Evaluation of students in medicine clerkships. Acad Med 1990;65(5):341-45.
6. Minion DJ, Donnelly MB, Quick RC, Pulito A, Schwartz R. Are multiple objective measures of student performance necessary? Am J Surg 2002;183(6):663-65.
7. Murphy JP, Rosevear GC, Gary NE. Concordance of medical school course grades and NBME Parts I and II examination subject test scores. J Med Educ 1984;59(10):828-31.
8. Nahum GG. Evaluating medical student obstetrics and gynecology clerkship performance: which assessment tools are most reliable? Am J Obstet Gynecol 2004;191(5):1762-71.
9. Ogunyemi D, De Taylor-Harris S. NBME Obstetrics and Gynecology clerkship final examination scores: predictive value of standardized tests and demographic factors. J Reprod Med 2004;49(12):978-82.
10. Erdmann JB, Hojat M, Veloski JJ. Comparing the accuracies of entire-group and subgroup models to predict NBME-I scores for medical school applicants. Acad Med 1992;67(12):860-62.
11. Holmes FF, Holmes GE, Hassanein R. Performance of male and female medical students in a medicine clerkship. JAMA 1978;239(21):2259-62.
12. Elzubeir MA, Rizk DE. Assessing confidence and competence of senior medical students in an obstetrics and gynaecology clerkship using an OSCE. Educ Health (Abingdon) 2001;14(3):373-82.
13. Ogunyemi D, Taylor-Harris D. Factors that correlate with the U.S. Medical Licensure Examination Step-2 scores in a diverse medical student population. J National Med Assoc 2005;97(9):1258-62.
14. Bienstock JL, Martin S, Tzou W, Fox HE. Medical students' gender is a predictor of success in the obstetrics and gynecology basic clerkship. Teach Learn Med 2002;14(4):240-43.
15. Omigbodun OO, Omigbodun AO. Influence of gender on undergraduate performance in psychiatry at Ibadan, Nigeria. Med Educ 2003;37(12):1091-93.
16. Wiskin CM, Allan TF, Skelton JR. Gender as a variable in the assessment of final year degree-level communication skills. Med Educ 2004;38(2):129-37.
17. Khan HU, Khattak AM, Mahsud IU, Munir A, Ali S, Khan MH, Saleem M, Shah SH. Impact of class attendance upon examination results of students in basic medical sciences. J Ayub Med Coll Abbottabad 2003;15(2):56-58.
18. Fogleman BS, Cleghorn GD. Relationship between class attendance and NBME Part I examination. J Med Educ 1983;58(11):904.
19. Parker GB. On our selection: predictors of medical school success. Med J Aust 1993;158(11):747-51.
20. Hays RB, Pearse P, Cooper CW, Sanderson L. Language background and communication skills of medical students. Ethn Health 1996;1(4):383-88.
21. Shaban S, McLean M, editors. Predictors of Medical School Performance. 9th Annual United Arab Emirates University Research Conference; Al Ain, United Arab Emirates; 2008.
22. Yindra KJ, Rosenfeld PS, Donnelly MB. Medical school achievements as predictors of residency performance. J Med Educ 1988;63(5):356-63.

23. Turner BJ, Hojat M, Gonnella JS. Using ratings of resident competence to evaluate NBME examination passing standards. J Med Educ 1987;62(7):572-81.

24. Gunzburger LK, Frazier RG, Yang LM, Rainey ML, Wronski T. Premedical and medical school performance in predicting first-year residency performance. J Med Educ 1987;62(5):379-84.

25. Bowles LT. Use of NBME and USMLE scores. Acad Med 1993;68(10):778.

26. Pearson SA, Rolfe IE, Henry RL. The relationship between assessment measures at Newcastle Medical School (Australia) and performance ratings during internship. Med Educ 1998;32(1):40-45.

27. Wolf FM, Shea JA, Albanese MA. Toward setting a research agenda for systematic reviews of evidence of the effects of medical education. Teach Learn Med 2001;13(1):54-60.