The Validity of Medical Students' Scores in their Internship Courses, a Historical Cohort Study

Haghdoost AA, MD, PhD[1]; Esmaeili A, MD[2]

[1] Assistant Professor, Kerman University of Medical
[2] Researcher, Kerman University of Medical Sciences

Abstract

Background and purpose*: Internship, undergraduate clerkships, is the last stage of training for medical students in Iran during which they work as medical doctors under supervision of academic staff for 1.5 years. We assessed the validity of students' scores in internship in one of the paramount medical universities in Iran.*

Methods*: In an historical cohort study, we collected the medical students' scores in four consecutive cohorts based on their entrance year to the medical school. We checked the validity of these scores and their internal consistencies by computing the Cronbach's alpha, and also using the discrimination index.*

Results*: The Cronbach's alpha of scores in the internship courses was lower than that in the other courses (0.72). In addition, the internship courses had the lowest discrimination index. Overall, female and younger students were more successful in their studies.*

Conclusion*: Although internship is one of the most important parts of medical education, it seems that the validity of students' scores during this phase was lower than that of other courses. These findings necessitate more work to document the predictive validity of internship evaluations by correlating them with future clinical performance.*

Key words*: Medical Education, Iran, Validity, Discrimination Index, Internal Consistency*

Journal of Medical Education Winter and Spring 2008; 12 (1,2): 29-36

Introduction

In Iran, medical students study basic sciences in five semesters and participate in a national comprehensive exam (Basic Science Comprehensive Exam: BSCE), before they are eligible to attend preclinical courses which consist of the physiopathology of the major body systems, pharmacology and pathology in three semesters. In the next step, they start theoretical and practical clinical courses for two years. A successful completion of these stages makes students eligible to participate in another national exam known as the Comprehensive Clinical Exam (CCE). Finally, they enter a 1.5 year internship programme which is comparable with the undergraduate clerkships in the US medical education system; during this phase they rotate in main clinical wards and work as a practitioner under the supervision of clinical academic staff. Internship is one of the most important phases of medical education (1, 2); however, we think that the workload of providing clinical care to patients may have a negative effect on the

Corresponding author:*Dr Aryan Esmaeili is a researcher in education development center, Deputy of Education, Kerman University of Medical Sciences, Jomhoori Islami Blvd, Postal code: 7618747653, Kerman, Iran*
Phone: +98-341-2443327
Fax: +98 341 2443327
E-mail: aryan.ei@gmail.com

training process. In our review literature, we could not find any explicit evidence to show the validity of students' scores in their internship courses in Iran, although this issue was explored deeply in other countries (3-11).

In this historical cohort study, we analyzed the students' scores of Kerman Medical University, one of the paramount medical schools in Iran, to check the validity of students' scores from different points of view. Generally, strong associations between students' scores are essential. Based on this fact, we checked the internal consistency of students' scores. Particularly, this approach is more applicable in course-based educational curriculums (3) such as the dominant model of medical curriculum in Iran.

From the other point of view, we checked how strong the students' scores could discriminate successful and unsuccessful students. It is plausible to suppose that a valid exam should differentiate these two groups markedly.

According to the above explanation, we quantified the internal consistency of medical students' scores in the internship clinical rotations to assess the validity of their exams, and also assessed the powers in discriminating successful and unsuccessful students. In addition, we explored the determinants of students' achievements using linear regression model. In these models, we explored the effects of students' scores in pre-internship courses and also their age and sex on their academic achievements.

Methods

Medical students in Kerman University of Medical Sciences (KUMS) were classified into separate cohorts based on their entry years, between 1995 and 1998. Then, their scores were obtained from the registry of KUMS; these were students' scores in theoretical courses. These data contained the students' scores in all courses and also their scores in the BSCE and CCE. In addition, the data contained the students' sex and date of birth. However, due to legal restrictions, the forms were anonymous and we could not link their data to other personal records.

In KUMS, the cutoff point for pass were 12 and 10 in exams with multiple choice questions and in assays, respectively. Only 3.4% of scores were less than these cutoff points. We entered both scores of a student if s/he failed in an exam and passed the course in the second round.

Nine academic achievement indicators (AAIs) were computed for each student. The numbers of credits of courses were used as their weights in this computation. The definitions of AAIs are as follows:

1.The student's GPA in basic science courses.

2.The student's GPA in preclinical courses, consisting of 1) the basic concepts in pharmacology, 2) the pathology of diseases, 3) the physiopathology of internal medicine, 4) semiology (part one) and 5) the epidemiology of common diseases in Iran.

3.The student's GPA in theoretical courses including 1) surgery, 2) internal medicine, 3) pediatrics, 4) gynecology & obstetric, 5) psychiatry, 6) neurology, 7) infectious diseases, 8) cardiology), 9) forensic medicine, 10) medical ethics & history (deontology), 11) public health, and 12) research dissertation.

4.The student's GPA in practical clinical courses including 1) internal medicine, 2) surgery, 3) pediatrics, 4) gynecology & obstetric, 5) neurology, 6) psychiatric, and 7) semiology (part two).

5.The student's GPA in preclinical and clinical (theoretical and practical) courses; i.e., the weighted average of indicators 2, 3 and 4.

6.The student's GPA in all pre-internship courses; i.e., the weighted average of the first four indicators.

7.The student's GPA in internship courses including 1) internal medicine, 2) surgery, 3) pediatrics, 4) gynecology & obstetric, 4) neurology, 5) psychiatric, 6) community health, 7) cardiology, 8) orthopedics, and 9&10) two optional courses out of these four choices: dermatology, ophthalmology, anesthesiology, and ENT.

8. The student's BSCE score.

9.The student's CCE score.

The students' scores in comprehensive exams, AAIs 8 and 9, were reported on a scale of 0 to 200; while other AAIs were on a scale of 0 to 20. To simplify the computations and presentation of results, we divided the students' scores in comprehensive exams by 10 to uniform the ranges of AAIs.

The associations among the AAIs and also between AAIs and the students' scores in their courses were assessed by computing the Pearson correlation coefficients. Moreover, we checked the consistency of student's scores using Cronbach's alpha(12).

In addition, 27% of students with the maximum and minimum CCE scores were labeled as successful and unsuccessful groups, respectively. The definitions for successful and unsuccessful students, 73 and 27 percentiles, were chosen just because of their similarities to the corresponding definitions in the item analysis of exams. Then the discrimination indices of all indicators were computed using the Whitney and Sabers formula for essay tests(13). The discrimination index quantifies the precision of the students' scores in a course in discriminating the successful and unsuccessful groups.

The analysis was done using the SPSS software version 13; in all analyses the significant level was set as 0.05.

Results

We analyzed the data of 243 medical students (133 males and 110 females). The mean age (SD) of students at entrance to the medical school was 19.1 (1.7) years; the difference between males' and females' ages was not statistically significant (p=0.24).

Females obtained greater scores in all indicators, except in the comprehensive exams; the differences between males' and females' scores in the comprehensive exams were not statistically significant. The average of females' and males' scores in their internship courses (AAI-7) were 17.09 (0.92) and 16.68 (0.86) respectively (p=0.001) (Table 1).

Also, we found a significant negative association between AAI-7 and their age at entrance to the medical school (r=-0.15, p=0.02). Although this association was statistically significant, it was weaker than the associations between age with other AAIs (Table 1).

All of the correlation coefficients between AAI-7 and other AAIs were greater than 0.25 and highly significant (p<0.0001). The maximum and minimum coefficients were observed between AAI-7 and the average score in clinical practice courses (AAI-4) (r=0.737) and BSCE (AAI-8) (r=0.275) respectively (Figure).

Exploring the students' scores showed that overall they were less successful in the national comprehensive exams than in their courses. Nonetheless, the correlation coefficient between BSCE and CCE was 0.62 (p<0.001) which implies that student scores in their comprehensive exams were highly correlated.

The Cronbach's alpha, as an indicator of the internal consistency of student's scores in internship courses, was 0.74 which was smaller than the alphas in other courses (Table 2). This number implies that students' scores during their internship courses were less correlated. Among the internship courses, the minimum correlation coefficient was observed between student scores in psychology and hygiene courses (r=0.05) which was very low.

The maximum and minimum discrimination indices were observed in the preclinical and internship courses, respectively (Table 3). This means that the students' scores in their internship courses had less power to discriminate successful and unsuccessful students compared to the other AAIs.

Using linear regression between student's internship scores and other potential determinants showed that among AAIs, student scores in their practical clinical courses was the most important predictors (regression coefficient=0.8, p<0.001). This means that by increasing one unit in the average score in the practical clinical courses, the student score in the internship courses increased by 0.8 units. (Table 4)

Table 1 a: The associations between the first 4AAIs and students' gender and age of entrance to the university

Group	***Average of student's score in***				
	Basic science courses (AAI-1)	**Pre-clinical courses (AAI-2)**	**Theoretical clinical courses (AAI-3)**	**Practical clinical courses (AAI-4)**	**All clinical courses (AAI-5)**
Gender					
Female (n=110)	14.9 (0.11)	15.24(0.13)	15.7 (0.10)	16.7 (0.07)	16.3 (0.07)
Male (n=133)	14.5 (0.11)	14.76(0.14)	15.5 (0.10)	16.2 (0.09)	15.9 (0.09)
p-value	0.01	0.012	0.168	<0.001	0.001
Age group					
<19 (n=102)	15.1(0.12)	15.4(0.15)	15.9(0.10)	16.7(0.08)	16.4(0.09)
19-20 (n=116)	14.5(0.10)	14.8(0.13)	15.4(0.10)	16.4(0.08)	16.0(0.08)
>20 (n=26)	14.1(0.21)	14.4(0.23)	15.1(0.22)	16.0(0.17)	15.7(0.17)
Correlation Coefficient	-0.29	-0.24	-0.28	-0.24	-0.28
p-value	<0.0001	0.001	<0.0001	<0.0001	<0.0001

Table 1 b: The associations between the second 4AAIs and students' gender and age of entrance to the university

Group	**Scores in comprehensive exam**			
	All pre-internship courses (AAI-6)	**Internship courses (AAI-7)**	**Basic science (AAI-8)**	**Pre-internship (AAI-9)**
Gender				
Female (n=110)	15.54(0.10)	17.1 (0.07)	12.8 (0.13)	11.3 (0.14)
Male (n=133)	15.14(0.10)	16.7 (0.08)	13.1 (0.15)	11.6 (0.16)
p-value	0.003	0.001	0.209	0.223
Age group				
<19 (n=102)	15.7(0.10)	17.1(0.09)	13.4(0.16)	11.9(0.17)
19-20 (n=116)	15.2(0.09)	16.8(0.08)	12.6(0.12)	11.2(0.13)
>20 (n=26)	14.8(0.18)	16.6(0.19)	12.1(0.18)	10.8(0.10)
Correlation Coefficient	-0.29	-0.15	-0.30	-0.24
p-value	<0.0001	0.05	<0.0001	<0.0001

Figure: The correlation coefficients between the average of student's score in internship courses, AAI-7, and the other AAIs; all correlations were statistically significant ($p<0.001$)

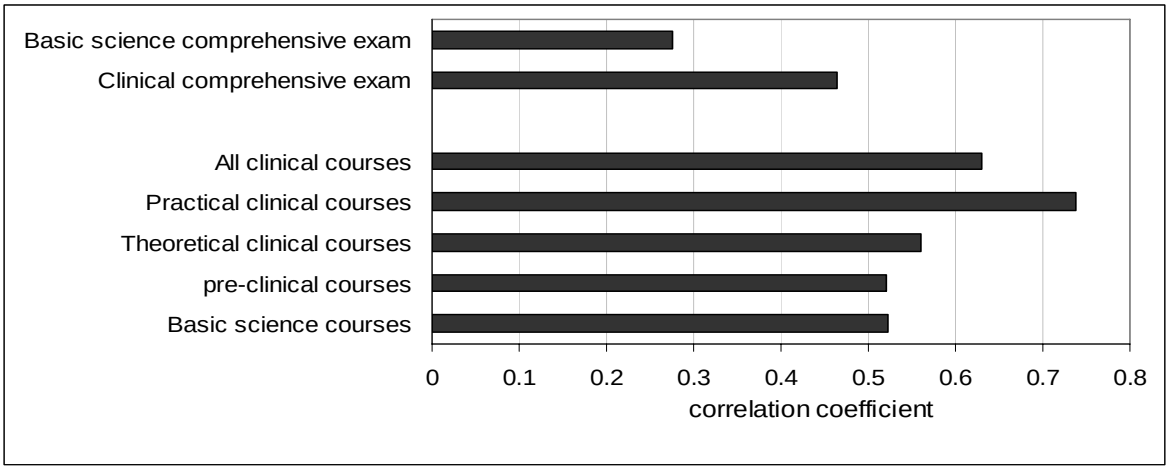

Table 2: The intra-cluster correlation coefficients, Cronbach's alpha, between courses classified by phase

Phase	Intra-cluster correlation coefficient (95% confidence Interval)
Basic science	0.93(0.90-0.95)
Preclinical course	0.94(0.92-0.95)
Clinical courses Theoretical Practical Total	0.86(0.84-0.88) 0.83(0.80-0.86) 0.91(0.89-0.92)
Internship	0.74 (0.69-0.79)

Discussion

Our findings showed that the internal consistency of student's scores in internship courses was less than that in other courses; also, their discrimination index was relatively low. The students' scores in internship courses had weaker correlations with the basic sciences scores than those of clinical courses. Overall, female and also younger students were more successful.

There were negative associations between age and AAIs. However, this association with AAI-7 was the slightest one. However, Cumlido-Hernandez et al, showed that older students had higher level of participation in their own learning process (10). We could not explore this issue explicitly; however, it may be explained by more social, financial and family engagement of older students, and also by sharper and fresher minds of younger students which also helped them to pass the entrance exam much sooner. The associations with the scores of theoretical courses were stronger, which is plausible, because learning theoretical issues usually needs more free time and fresher minds.

In Iran, female students, particularly single ones,

Table 3: The discrimination indices of AAIs; the successful and unsuccessful students were defined based on their scores in clinical comprehensive exams

AAIS	The mean score in unsuccessful group	The mean score in successful group	Discrimination index* (95% CI)
Basic science courses (AAI-1)	13.97	15.61	1.63(1.28-1.98)
Preclinical course (AAI-2)	14.00	19.20	2.20(1.80-2.59)
Theoretical clinical courses (AAI-3)	14.81	16.58	1.76(1.45-2.08)
Practical clinical courses (AAI-4)	16.01	17.11	1.10(0.83-1.37)
All clinical courses (AAI-5)	15.53	16.9	1.36(1.10-1.63)
All pre-internship courses (AAI-6)	14.30	16.24	1.61(1.32-1.91)
Internship courses (AAI-7)	16.50	17.45	0.95(0.65-1.25)
Basic Science Comprehensive Exam (AAI-8)	12.26	14.17	2.14(1.71-2.57)

**The difference between successful and unsuccessful group*

Table 4: Predictors of the student's score in Internship courses, AAI-7, using linear regression model

Predictor	Regression Coefficient	P-value
Age at entrance	-0.07	0.04
Gender	0.4	0.001
Basic Science comprehensive exam	0.03	<0.001
Clinical Comprehensive exam	0.02	<0.001
Practical clinical courses	0.8	<0.001
Theoretical clinical courses	0.5	<0.001
Basic science	0.4	<0.001

have fewer responsibilities in the family and they are mostly dependent on financial support from their families. In addition, they socialize less, and therefore have much more time to dedicate to their studies. Although these factors are culture dependent, there were evidences that showed females have been more successful in some other countries as well (11, 14, 15). This could be due to the differences in learning styles of males and females (16), and there are some studies which showed reverse results particularly in comprehensive and national exams (17, 18).

The students' scores in practical clinical courses had the strongest association with their scores in internship courses which could be explained easily. Both of these courses are mostly focusing on skills and practical capacities. These courses are hospital based; in addition, the same academic staff teams are responsible for these two types of courses. Compatible with our findings, similar associations were reported in other studies. (19, 20)

We applied the concept of the discrimination index (DI), which is commonly used in the analyses of question appropriateness, to assess the appropriateness of AAIs. The DI shows how perfectly a question can discriminate successful and unsuccessful respondents. The DI for each

question is the difference between proportions of correct responses in successful and unsuccessful respondents. With an exactly similar logic, we defined successful and unsuccessful students based on their CCE scores, and compared the other AAIs in these two groups.

Based on these findings, the DI of the students' scores during Internship courses (AAI-7) was low. In other words, the difference between successful and unsuccessful students' scores in these courses was not as large as that in other courses.

From another point of view, the internal consistency of students' scores during internship was not as strong as that in other stages. On the other hand, students' internship scores were greater than those in other stages.

Although the sample size was enough in most of the analyses, all of the data were extracted from one university which could be a point of concern in generalizing our findings to the other universities in Iran. In addition, we did not have access to the students' information such as their scores in high school and in the entrance exam to the university. Nonetheless, these two limitations could not distort our main conclusion which is the lower validity of students' scores during internship in KUMS.

All these findings imply that the validity of students' scores during internship was not as strong as the validities in other stages while the goal of developing reliable measures of faculty attitudes toward clinical evaluation of medical students has been important in medical schools (21). Interns are working as medical doctors in hospitals under the supervision of academic staff; they spend most of their time providing care for patients, and they are focusing mostly on their skills. At the same time, most academic staff do not evaluate interns deeply based on their capacities since they look at interns as colleagues not students. It should be mentioned that the evaluation of practical capacities are much more difficult than theoretical capacities and there are less valid tools for such an evaluation (5). For example, generating a multiple choice questioner is less time consuming comparing to running an ASCI exam. The cumulative impacts of these factors may deteriorate the validity of exams in the internship phase.

We found that the validity of students' scores in internship phase was low. Unfortunately, a weaker supervision on the training process during internship, which is the last phase for a medical student, has direct effect on the capacities of a medical doctor. Therefore, in conclusion, we suggest that the system should pay more attention to the evaluation of intern's programs.

Acknowledgement

The authors are grateful to registry staff of Kerman University of Medical Sciences who helped in data collection. In addition, the authors highly appreciate the contribution of Dr. Shiva Mehravaran for her valuable comments on the final draft of the paper particularly in its English.

References

1.Lavin B, Pangaro L. Internship ratings as a validity outcome measure for an evaluation system to identify inadequate clerkship performance. Acad Med 1998;73:998-1002.

2.Ray AM. The value of internship in graduate medical education: survey of emergency medicine residents and program directors. J Am Osteopath Assoc 2007;107:26-9.

3.Stone SL, Qualters DM. Course-based assessment: implementing outcome assessment in medical education. Acad Med 1998;73:397-401.

4.Guidry TD, Cohen PA. A practical examination for student assessment in an externship program. Am J Pharm Educ 1987;51:280-4.

5.Durning SJ, Pangaro LN, Lawrence LL, Waechter D, McManigle J, Jackson JL. The feasibility, reliability, and validity of a program director's (supervisor's) evaluation form for medical school graduates. Acad Med 2005;80:964-8.

6.Durning SJ, Pangaro LN, Denton GD, et al. Intersite consistency as a measurement of programmatic evaluation in a medicine clerkship with multiple, geographically separated sites.

Acad Med 2003;78:S36-8.
7.Paolo AM, Bonaminio GA. Measuring outcomes of undergraduate medical education: residency directors' ratings of first-year residents. Acad Med 2003;78:90-5.
8.Campos-Outcalt D, Witzke DB, Fulginiti JV. Correlations of family medicine clerkship evaluations with scores on standard measures of academic achievement. Fam Med 1994;26:85-8.
9.Part HM, Markert RJ. Predicting the first-year performances of international medical graduates in an internal medicine residency. Acad Med 1993;68:856-8.
10.Cumplido-Hernandez G, Campos-Arciniega MF, Chavez-Lopez A, Perez-Garcia V. [Learning approaches used by undergraduate interns in the development of a medical specialty]. Rev Med Inst Mex Seguro Soc 2006;44:321-8.
11.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;14:240-3.
12.Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297-334.
13.Whitney DR, Sabers DL. Improving essay examinations: Use of item analysis University of Iowa, 1970.
14.Smits PB, Verbeek JH, Nauta MC, Ten Cate TJ, Metz JC, van Dijk FJ. Factors predictive of successful learning in postgraduate medical education. Med Educ 2004;38:758-66.
15.Buddeberg-Fischer B, Klaghofer R, Abel T, Buddeberg C. The influence of gender and personality traits on the career planning of Swiss medical students. Swiss Med Wkly 2003;133:535-40.
16.de Saintonge DM, Dunn DM. Gender and achievement in clinical medical students: a path analysis. Med Educ 2001;35:1024-33.
17.Dawson B, Iwamoto CK, Ross LP, Nungester RJ, Swanson DB, Volle RL. Performance on the National Board of Medical Examiners. Part I Examination by men and women of different race and ethnicity. Jama 1994;272:674-9.
18.Frischenschlager O, Haidinger G, Mitterauer L. Factors associated with academic success at Vienna Medical School: prospective survey. Croat Med J 2005;46:58-65.
19.Andriole DA, Jeffe DB, Whelan AJ. What predicts surgical internship performance? Am J Surg 2004;188:161-4.
20.Pearson SA, Rolfe IE, Henry RL. The relationship between assessment measures at Newcastle Medical School (Australia) and performance ratings during internship. Med Educ 1998;32:40-5.
21.McGaghie WC, Richards BF, Petrusa ER, et al. Development of a measure of medical faculty attitudes toward clinical evaluation of students. Acad Med 1995;70:47-51.