# THE INTERNAL CONSISTENCY OF MEDICAL STUDENTS' SCORES IN THEIR PHYSIOPATHOLOGY AND CLINICAL COURSES

ARYAN ESMAEILI, ALI-AKBAR HAGHDOOST

### ABSTRACT

**BACKGROUND:** To quantify the internal consistency of medical students' scores. AIM: We assessed the associations between medical students' scores in physiopathology and clinical courses and compared these scores with their scores in their comprehensive exams. SETTINGS AND DESIGN: We collected medical students' scores in their courses and also in their comprehensive exam in six consecutive years. MATERIALS AND METHODS: We assessed the associations between students' scores and their personal characteristics, and the consistency between theoretical and practical courses. STATISTICAL ANALYSIS: We used Pearson correlation coefficient and linear regression. In addition, we computed difficulty and discrimination indices of students' scores in their courses by comparing these scores with comprehensive clinical exam (CCE). RESULTS: Generally, females and younger students were more successful. CCE were predicted by students' scores and their characteristics relatively accurate (the adjusted R2 of the model was 0.59). Students' scores in the pathology and in thesis had the maximum and minimum discrimination indices, while the difficulties of these two courses were in reverse order. The strongest association was observed between theoretical and practical scores in internal medicine while the associations between theoretical and practical scores in the other courses were not strong although all of them were statistically significant. CONCLUSIONS: Using this approach to explore the students' score, might highlight the weak points of the current educational system. For example we found that the students' score in thesis had the minimum accuracy; although students obtained very high score in this course. Hence, for better comparison of the accuracy students' scores in colleges around the world, we recommend similar quantitative approach

Key words: Difficulty index, discrimination index, Iran, medical education, validity

Dept. Educational Development Center, Kerman University of Medical Sciences, Jomhoori Islami Blvd, 7618747653, Kerman, Iran

#### Correspondence:

Dr. Aryan Esmaeili Education Development Center, Kerman University of Medical Sciences, Jomhoori Islami Blvd, Postal code: 7618747653, Kerman, Iran. E-mail: aryan\_ei@yahoo.com

The Paper was Partly Published in an Iranian Journal

## INTRODUCTION

In Iran, medical students study basic science in five semesters and participate in a comprehensive exam (Basic Science Comprehensive Exam: BSCE). After that, they study preclinical courses on physiopathology of main body systems and also pharmacology, pathology for three semesters. In the next

267

step, they start theoretical and practical clinical courses for two years. A successful completion of these periods makes students eligible to participate in another national comprehensive clinical exam (CCE). In this exam, students are evaluated for all of pre-clinical and clinical courses.

An acceptable internal consistency between students' scores in preclinical and clinical courses and also significant associations between scores in these courses and in comprehensive exams may indicate the validity of exams indirectly. Particularly, this method of validity assessment is more appropriate in course-based educational curriculums<sup>[1]</sup> such as the model of medical curriculum in Iran.

There are a great deal of studies which assessed the relationship of some of variables such as students' scores in high school,<sup>[2,3]</sup> premedical summer programs<sup>[4]</sup> and admission tests<sup>[5,6]</sup> and even their personal characteristics with students' scores in their courses.<sup>[2,3,7]</sup> Most of these studies used those scores and characteristics as predictors of students' achievement. Nonetheless, this paper mainly explored the validity of students' scores in their courses using backward approach by comparing students' scores in CCE with their scores in their courses, using the concepts of difficulty and discrimination indices. This approach is not a common approach and we believe that the concepts of difficulty and discrimination indices could be applied to check the validity of whole exams.

According to the above explanation, we quantified the internal consistency of medical students' scores in the physiopathology and

clinical courses to assess the internal validity of their scores. In addition, we computed difficulty and discrimination indices of students' scores in their courses by comparing these scores with CCE score. Additionally, we assessed the age and gender effects on the academic achievement.

### MATERIALS AND METHODS

Medical students in Kerman University of Medical Sciences (KUMS) were classified into separate cohorts based on the entry year between 1995 and 2000. Then, their physiopathology and clinical course scores were obtained from the registry of KUMS in paper forms. These forms also contained the students' BSCE and CCE scores, 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.

The data were double entered and the validity of the data entry process was assessed.

# Six academic achievement indicators (AAIs) were computed as follows:

- The average of scores in physiopathology courses consisting of basic concepts of pharmacology, pathology of diseases, physiopathology of internal medicine, general physical examination and the epidemiology of common diseases in Iran.
- The average of scores in practical clinical courses including internal disease, surgery, pediatrics, gynecology and obstetric disease, neurology, psychiatric, advance physical examination, forensic medicine, medical ethics and history (deontology), public health, and thesis.

- The average of scores in theoretical courses including surgery, internal medicine, pediatrics, gynecology and obstetric disease, psychiatry, neurology, infection diseases, and cardiology
- The total average in physiopathology, theoretical and practical courses; i.e., the weighted average of the above three indicators
- The score in the BSCE
- The score in the CCE

The scoring system in KUMS is on a scale of 0 to 20; however, the comprehensive exams are scored on a scale of 200 points. For easier comparison, BSCE and CCE scores were converted to one on a scale of 20 points.

The associations between the AAIs and also between AAIs and the students' scores in their courses were assessed by computing Pearson correlation coefficients. In addition, 27% of students with the top and lowest scores in the CCE were labeled successful and unsuccessful groups; then the discrimination and difficulty indices of all courses were computed using the Whitney and Sabers formula for essay tests.<sup>[8]</sup> The computed difficulty index implies how difficult the course was for students, while the computed discrimination index quantifies the accuracy of the students' scores in a course in discriminating the top and lowest groups.

The analysis was done using the SPSS software version 11.5; the significant level was 0.05.

## RESULTS

From 1995 to 2000, 481 medical students started their studies at KUMS (39.7% male). The minimum and maximum annual number of enrolled students was 45 (in 2000) and 99 (in 1997 and 1998), respectively.

Females were more successful in their studies based in all AAIs (*P*<0.001), except in CCE and BSCE scores. In CCE and BSCE, males' scores were slightly greater than females' scores, but the differences were not statistically significant [Table 1].

Negative associations were observed between the entrance age and the academic

Table 1: Description (mean and standard error) of the academic achievement indicators, classified by age at entrance and gender

Group	Average				Score in comprehensive exam	
	Physio pathology	Clinical (Practice)	Clinical (theory)	Total	Basic science	Clinical courses
Gender						
Female (n=290)	15.35(0.09)	16.76(0.06)	15.53(0.08)	16.27(0.06)	12.84(0.09)	11.64(0.12)
Male (n=191)	14.6(0.11)	16.15(0.08)	15.12(0.1)	15.74(0.08)	12.93(0.11)	11.85(0.15)
P-value	< 0.001	0.001	< 0.001	<0.001	0.519	0.286
Age group						
<19 (n=188)	15.58(0.12)	15.83(0.09)	16.75(0,07)	16.38(0.08)	13.45(0.12)	12.16(0.16)
19-20 (n=247)	14.83(0.09)	15.12(0.08)	16.46(0.06)	15.92(0.06)	12.59(0.09)	11.54(0.12)
>20 (n=42)	14.39(0.24)	14.97(0.21)	16.09(0.20)	15.63(0.18)	12.15(0.17)	11.02(0.29)
Correlation coefficient	-0.19	-0.16	-0.16	-0.18	-0.20	-0.18
P-value*	< 0.001	< 0.001	<0.001	< 0.001	<0.001	<0.001
All students (n=481)	15.08(0.07)	15.38(0.06)	16.54(0.05)	16.08(0.05)	12.89(0.07)	11.74(0.09)

\**P*-values were computed based on one way ANOVA. The results of post-hoc (Tukey) showed significant differences between all age groups in all variables, except the difference between age group 19-20 and >20 in the score of CCE (*P*-value=0.105)

achievement [Table 1]. Students were classified based on their entrance age into three groups: 1) under 19 years of age; most of who successfully started their academic studies right after high school, 2) 19 and 20 years of age; who started their academic studies with a one or two year gap, and 3) over 21 years of age. The trend of all achievement indicators showed that the success rate decreased with age (P<0.001). The greatest correlation coefficient observed between age and BSCE (r=-0.2).

There were strong correlations between students' scores in all of the AAIs in both genders. The strongest association was observed between the students' scores in the physio-pathology and clinical courses (in males: r=0.811, in female: r=0.802) [Table 2].

Enrolling all possible predictors of students' scores in CCE, a linear regression model was

Table 2: The correlation coefficients between the academic achievement indicators; light and medium shades show the coefficients in males and females respectively

	BSCE	Physio- pathology	Clinical courses	CCE
BSCE		0.663	0.555	0.526
Physio-pathology	0.621		0.811	0.669
Clinical courses	0.572	0.802		0.630
CCE	0.685	0.624	0.626	

All correlation coefficients were significant at the 0.01 level (2-tailed), BSCE - Basic Science Comprehensive Exam, CCE - Comprehensive clinical exam

#### Table 3: Prediction of the student score in comprehensive clinical exam using linear regression model

Predictor	Regression coefficient	P-value
Age	-0.03	0.471
Female	-0.522	<0.001
BSCE	0.361	<0.001
Physio-pathology	0.251	0.003
Clinical theoretical course	0.096	0.365
Clinical practical course	0.457	<0.001

The adjusted R2: 0.59, *P*<0.0001, BSCE - Basic Science Comprehensive Exam

Indian J Med Sci, Vol. 62, No. 7, July 2008

generated [Table 3]. The adjusted R<sup>2</sup> of the model was 0.59 which implied an acceptable accuracy. Adjusting for the effects of the other variables, age and students' scores in their theoretical courses were not significant. While, the adjusted mean difference between males' and females' score was -0.522 (*P*-value<0.001). In addition, the results showed that by one unit increase in student scores in BSCE, physiopathology and practical courses, their score in CCE was increased 0.36, 0.25, and 0.46 units respectively.

In the next step, we compared the students' scores in theoretical and practical exams, classified by courses [Figure 1]. These results showed a wide variation between coefficients, the strongest correlation coefficient was observed in internal medicine (r=0.65) and the weakest coefficient in neurology (r=0.24). Nonetheless, all of these coefficients were highly significant (*P*-values <0.0001).

Based on the results of Table 4, students got the greatest and the lowest scores in thesis (19.01) and in pathology of diseases (14.28) respectively. In fact, the score of students in thesis were considerably greater than the other courses with the minimum standard deviation



**Figure 1:** The correlation coefficients between theoretical and practical students' scores in their courses. All of these coefficients were highly statistically significant (*P*-value <0.0001).

270

	Mean of students' score as an indicator of difficulty Index		The difference of scores in top and lowest groups of students as an indicator of discrimination index			
	Mean	SD	Mean Difference	95% Cl	P-value	
Specific courses						
Internal Disease	14.94	1.52	2.38	2.05-2.7	<0.001	
Surgery	15.05	1.52	1.94	1.57-2.32	<0.001	
Pediatrics	15.52	1.66	2.17	1.73-2.6	<0.001	
Gynecology and Obstetric Disease	16.13	1.54	2.02	1.62-2.41	<0.001	
Neurology	16.04	1.59	1.39	1.03-1.76	<0.001	
Psychiatric	15.81	1.54	1.37	0.94-1.8	<0.001	
Minor courses*	17.37	1.07	1.01	0.76-1.26	<0.001	
Pathology of diseases	14.28	2.24	2.95	2.41-3.48	<0.001	
Pharmacology	15.47	2.26	2.33	1.74-2.92	<0.001	
Physical examination	16.15	1.39	1.41	1.02-1.8	<0.001	
Forensic Medicine	15.64	2.29	1.86	1.24-2.48	<0.001	
Medical Ethics and History (Deontology)	16.79	1.77	0.87	0.35-1.39	0.01	
Public Health**	16.05	1.65	1.32	0.93-1.72	<0.001	
Thesis	19.01	0.86	0.17	-0.17-0.51	0.32	
Mean scores in						
Physiopathology courses	15.06	1.62	2.46	2.1-2.82	<0.001	
Theoretical courses	15.36	1.36	1.83	1.5-2.16	<0.001	
Practice clinical courses	16.52	1.05	1.21	0.98-1.45	<0.001	
All of the above courses	16.06	1.08	1.46	1.21-1.71	<0.001	
Basic Science Comprehensive Exam	12.88	1.53	2.36	1.99-2.73	<0.001	

\*Minor courses consists of Urology, Radiology, Ophthalmology, Orthopedics Disease, Dermatology, E.N.T, \*\*Public Health consists of Biostatistics and Research Methodology, Epidemiology and Iran's Common Disease, and family medicine

(SD=0.86) which means that students' scores in thesis were much closer than their scores in the other courses. In contrast, the mean scores of students in BSCE and CCE were much lower that their scores in the other courses (12.88 and 11.72 respectively).

On the other hand, based on the computed discrimination indices. thesis score had the minimum discrimination index (0.17, P-value=0.32); in the other words, the mean difference of top and weak students' scores in thesis was only 0.17. While, the maximum discrimination index was observed for the pathology of diseases (2.95, *P*-value = <0.001). After that, the mean score in all physiopathology courses, in internal medicine and in pharmacology had the maximum discrimination indices [Table 4].

### DISCUSSION

The results showed that the associations between AAIs were relatively strong. Nonetheless, the consistency of students' scores in theoretical and practical courses in some subjects such as neurology, cardiology and infection diseases were much less than that in internal medicine. Generally, students' scores in thesis were much greater than the other courses, but it had the minimum discrimination index. In contrast, although students got the lowest scores in the pathology of diseases, it had the maximum discrimination index which implies that the score of this course could discriminate top and lowest group of students much better than other scores. Nonetheless, in all of the courses, the discrimination indices were not considerable. Moreover, the results of

271

multivariate analysis showed that the students' scores in their theoretical courses did not predict their scores in CCE.

Generally, younger students and females were more successful. There was a strong negative association between entrance age and AAIs, which has been reported in many studies.<sup>[7]</sup> In Iran, female students, particularly single ones, 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 is evidence that shows females were more successful in some other countries as well.<sup>[7]</sup> It should be added that male students were slightly more successful in the comprehensive exams, which may imply that their long term achievement is at least in the same level as females.

We applied the concept of the discrimination index commonly used in the analyses of question appropriateness to assess the appropriateness of exams. The discrimination index is an indicator that shows how perfectly a question can discriminate successful and unsuccessful respondents. For this purpose, you define successful and unsuccessful respondents based on their scores in an exam; then, you check the proportion of successful and unsuccessful respondents who provide correct responses to every question. The discrimination index 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 scores in the CCE, and

compared their scores in courses.

Based on the above logic, we can imply than the students' scores in their thesis had the minimum power to discriminate successful and unsuccessful students. Surprisingly students got the best scores in their thesis. Therefore, we can imply that the scores of thesis had the minimum accuracy.

On the other hand, the students' scores in comprehensive exams were considerably lower than their scores in their courses which might partly be explained because CCE is a national exam and its standards are different. Although it is reported in other studies as well,<sup>[9,10]</sup> we may think more about the validity of local exams. It is one of the basic concepts in exams that the questions should focus on the topics that students must learn based on the teaching curriculum.[11,12] Nonetheless, it is not hard to believe that examiner will focus on those topics that were taught if they had played the role of teacher as well. Therefore, in the best scenario we can suggest independent professionals evaluate students based on their course plan.

The correlations between students' scores in practical and theoretical courses were not as strong as we might expect. On an average, the correlation coefficients were around 0.4. These low associations also imply that there were some problems in either the teaching methods or in exams. The strongest association was found between students' scores in theoretical and practical exams of internal medicine. In addition, the discrimination index of internal medicine was among the best ones. These findings may show an acceptable validity of the exams in internal medicine courses.

274

273

Nonetheless, we may remember that internal medicine is the most important course for medical students<sup>[13]</sup> and students pay more attention to its contents and study internal medicine much deeper than the other courses.

This study only reviewed the internal consistency of medical students' scores only in Kerman University of Medical Sciences. The pedagogic shift from traditional approach to a need-based approach requires a fundamental change of the roles and commitments of educators, planners and policymakers.<sup>[14]</sup> We could not find similar analysis on the scores of students in other colleges to compare our findings. Therefore, we encourage researchers around the world to explore students' scores with similar methodology. For sure, comparison between the internal consistencies of students' scores in different colleges can expand our knowledge about the effects of different teaching curriculum on the learning of students.

## CONCLUSION

We found that the internal consistencies between students' scores in their courses were not generally strong and in a few courses such as thesis, the validity of their scores was not acceptable. The weak associations between students' scores in practical and theoretical courses was only acceptable in internal medicine course and it seems that other departments in Kerman University of Medical Sciences should review their teaching curriculums and their exams to find the sources of these weak associations. However, the strong associations between the averages of scores in comprehensive exams and also in the physio-pathology, theoretical and practical courses may imply that the averages of scores are much more accurate than the scores of individual courses.

## REFERENCES

- Stone SL, Qualters DM. Course-based assessment: Implementing outcome assessment in medical education. Acad Med 1998;73: 397-401.
- Lipton A, Huxham GJ, Hamilton D. Predictors of success in a cohort of medical students. Med Educ 1984;18:203-10.
- Höschl C, Kozený J. Predicting academic performance of medical students: The first three years. Am J Psychiatry 1997;154:87-92.
- 4. Strayhorn G. Participation in a premedical summer program for underrepresented-minority students as a predictor of academic performance in the first three years of medical school: two studies. Acad Med 1999;74:435-47.
- Dixon D. Relation between variables of preadmission, medical school performance and COMLEX-USA levels 1 and 2 performance. J Am Osteopath Assoc 2004;104:332-6.
- Carline JD, Cullen TJ, Scott CS, Shannon NF, Schaad D. Predicting performance during clinical years from the new medical college admission test. J Med Educ 1983;58:18-25.
- 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.
- Whitney DR, Sabers DL. Improving essay examinations: Use of item analysis University of lowa; 1970.
- Evans P, Goodson LB, Schoffman SI, Baker HH. Relations between academic performance by medical students and COMLEX-USA Level 2: A multisite analysis. J Am Osteopath Assoc 2003;103:551-6.

- Agostini DE, Stano AS, Parente DH. Student performance on the comprehensive osteopathic medical licensing examination-USA level 2 following a clinical evaluation, feedback, and intervention program. J Am Osteopath Assoc 2002;102:477-80.
- 11. Wiles K. The Changing Curriculum of the American High School. Prentice-Hall; 1963.
- Damjanov I, Fenderson BA, Hojat M, Rubin E. Curricular reform may improve students' performance on externally administered comprehensive examinations. Croat Med J

2005;46:443-8.

- Hemmer PA, Elnicki DM, Albritton TA, Kovach R, Udden MM, Wong RY, *et al.* The responsibilities and activities of internal medicine clerkship directors. Acad Med 2001;76:715-21.
- Majumder AA, D'Souza U, Rahman S. Trends in medical education: Challenges and directions for need-based reforms of medical training in South-East Asia. Indian J Med Sci 2004;58:369-80.

Source of Support: Indian Council of Medical Research, New Delhi, India, Conflict of Interest: None declared.

## Author Help: Reference checking facility

The manuscript system (www.journalonweb.com) allows the authors to check and verify the accuracy and style of references. The tool checks the references with PubMed as per a predefined style. Authors are encouraged to use this facility before submitting articles to the journal.

- The style as well as bibliographic elements should be 100% accurate to get the references verified from the system. A single spelling error or addition of issue number / month of publication will lead to error to verifying the reference.
- Example of a correct style Sheahan P, O'leary G, Lee G, Fitzgibbon J. Cystic cervical metastases: Incidence and diagnosis using fine needle aspiration biopsy. Otolaryngol Head Neck Surg 2002;127:294-8.
- Only the references from journals indexed in PubMed would be checked.
- Enter each reference in new line, without a serial number.
- Add up to a maximum 15 reference at time.
- If the reference is correct for its bibliographic elements and punctuations, it will be shown as CORRECT and a link to the correct article in PubMed will be given.
- If any of the bibliographic elements are missing, incorrect or extra (such as issue number), it will be shown as INCORRECT and link to possible articles in PubMed will be given.