Internal Consistency of Medical Students' Scores in General and Baisc Science Exams, Kerman University, Iran

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

[1] Assistant Professor, Kerman University of Medical Sciences and Health services and honorary lecturer in London School of Hygiene and Tropical Medicine
[2] Researcher, Kerman University of Medical Sciences

Received:March 2006 ***Accepted: May 2006***

Abstract

Background and purpose: *Course-based assessment is a method to gather, analyze, disseminate, and use course data to improve student learning. We assessed the associations between medical students' scores in basic sciences and general courses in a university in Iran and compared these scores with their scores in comprehensive exam.*

Methods: *We collected the scores of medical students in their courses and also in their comprehensive exam in Kerman University in eight consecutive years (1995-2002). Using a hierarchical cluster analysis and discrimination index, the internal consistency of students' scores was assessed.*

Results: *Generally, females were more successful. In addition, age had a strong negative correlation with academic achievement. The temporal variations in students' achievements were more or less constant. Students' scores in anatomy, biochemistry, histology, immunology, medical English, microbiology and physiology had the greatest discrimination indices and also stronger intra-cluster correlations.*

Conclusion: *It seems that a cluster analysis and the discrimination index are powerful approaches to be used in a course-based assessment and to check the validity of students' scores.*

Key words: *MEDICAL EDUCATION, IRAN, VALIDITY, CLUSTER ANALYSIS, DISCRIMINATION INDEX*

Journal of Medical Education Spring2006; 9(1); 3-10

Introduction

In Iran, medical students should pass eight general courses (GCs) for 20-23 credits, and 15 basic science courses (BSCs) for 63-65 credits during the first five semesters of their studies; compatible with the national basic science curriculum. A successful completion of this period makes students eligible to participate in the national comprehensive exam (CE). This exam tests students' knowledge in BSCs, and is a prerequisite for the second stage of their medical education. Assessing the students' scores in their courses and comparing them with their scores in CE can show the validity of students' scores indirectly. It is clear that the low variation in students' scores in their courses in consecutive cohorts, and also their high internal consistencies can show the internal validity of exams. This method of validity assessment is more appropriate in course-based educational curriculums.(1) Course-based assessment is a systematic way to gather, analyze, disseminate, and use course data to improve student learning, and it is well suited to current trends in health professions education .(1) It should be added that the current medical education method in Iran is

Corresponding author: *Dr Aryan Esmaeili, is a researcher in Kerman University of Medical Sciences.*
Address: Education development center, Deputy of Education, Kerman University of Medical Sciences, Jomhoori Islami Blvd, Postal code: 7618747653, Kerman, Iran
Tel: 98 341 2443327
Fax: 98 341 2443327
E-mail: Aryanei@yahoo.com

course-based which means that departments teach different courses to medical students and assess their uptakes with separate exams.

Assessment of the correlations between students' scores in their courses with other predictors of academic achievement is a well-known method. Among the most important predictors were the results of students in high school (2,3) participating in premedical summer program ,(4) the results of admission test (5,6) and even personal characteristics such as the results of personality tests (2,3,7) and gender (7).

This study explored the associations between medical students' scores in the first two and half years of their studies to assess the internal validity of their scores and compare them with their scores in the national comprehensive exam, classified by sex and age. In this analysis, we extended the concept of the discrimination index (DI) and used a hierarchical cluster analysis in this setting.

Methods and materials

Medical students in Kerman University of Medical Sciences (KUMS) were classified into separate cohorts based on their year of entry. Then every GC and BSC test scores of cohorts 1995 to 2002 were obtained from the registry of KUMS in paper forms. These forms also contained the students' CE 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. The average of a student's scores in a course was computed if s/he failed in that course.

To assess successfulness of a student, six academic achievement indicators (AAIs) were computed as follows:

1. The grade point average in GCs consisting of Persian literature (3 credits), general English (3 credits until 1999, and 6 credits later on), religious studies (10 credits), ethics (2 credits) and physical education (2 credits).
2. The grade point average in BSCs including anatomy (10 credits until 1999, and 12 credits later on), biochemistry (6 credits), physiology (9 credits), hygiene/epidemiology (6 credits), microbiology and virology (5credits), parasitology and mycology (4credits), immunology (3 credits), genetics (2 credits), histology (4 credits), embryology (2 credits), pathology (5 credits), bio-physic (2 credits), psychology (2 credits), nutrition (2 credits) and medical English and terminology (6 credits).
3. The grade point average in all GCs and BSCs
4. The number of failed courses
5. The number of failed credits
6.The score in the national CE

The scoring system in KUMS is on a scale of 0 to 20; however, the CE is scored on a scale of 200 points. For easier comparison, CE scores were converted to one on a scale of 20 points. Also based on the content of general courses they were grouped into 5 categories.

The associations between the AAIs and also between AAIs and the students' scores in their courses were assessed by computing the Pearson correlation coefficients. In addition, 27% of students with the top and lowest scores in the CE were labeled successful and unsuccessful groups; then the DIs of all courses were computed using the Whitney and Sabers formula for essay tests [8].

To check the intra-cluster correlation between students' scores in BSCs, the hierarchical cluster method was used; the average linkages between students' scores in different courses were computed and the results were illustrated in a dendrogram. In this graph, a longer line between two courses implies a weaker association; i.e., lower consistency between students' scores. Students were classified based on their entrance age into three groups: 1) under 19 years of age; most of which successfully passed the entrance exam right after high school, 2) 19 and 20 years of age; who mostly passed the entrance exam with a one or two year gap, 3) over 21 years of age. The analysis was done using the SPSS software version 11.5; the significant level was 0.05.

Results

From 1995 to 2002, 571 medical students started their studies at KUMS (40.7% male). The minimum and maximum annual number of enrolled students was 38 (in 2001) and 98 (in 1997), respectively.

Students' scores in the CE had stronger correlations with the students' scores in the BSCs in both genders (male: r=0.75, female: r=0.7) and age groups (all correlation coefficients were greater than 0.72).

Temporal variations in AAIs were not considerable, although the variations in the means of students' scores in BSCs and in all courses collectively were statistically significant (p=0.03 and p=0.02 respectively) (Figure 1). The variation in students' scores in GCs was bimodal, with two peaks in the cohorts of 96-97 and 2001-2. The scores in BSCs showed a drop in 1998. Interestingly, in 2000, when the students' scores in BSCs peaked, and the number of failed courses and credits was minimum, the scores in GCs and CE were not as high as might be expected; i.e., this cohort of students was more successful based on all indicators except their scores in GCs.

In general, the mean of students' scores in GCs were greater than those in BSCs (Figure 2). The maximum and minimum of mean scores in GCs belonged to physical education (mean±SD: 18±1.6) and general English (14.32±4.1), respectively; the corresponding mean scores in BSCs belonged to psychology (16.1±2.2) and histology (13.2±2.1). Similar patterns were observed in both genders and age groups.

The standard deviations (SD) of scores, as an indicator of disparity in students' scores, had considerable variations. The students' scores in biophysics (SD=4.5), general English (SD=4.1), anatomy (SD= 3.1) and medical English (SD=2.9) had maximum variations. In contrast, students' scores in Persian literature (SD=1.6), physical education (SD=1.6) and microbiology (SD=1.8) had minimum variations (Figure 2).

Figure 1: The temporal variations in academic achievement indicators. Only the variations in the scores in basic science courses and in all courses were significant (p=0.03 and p=0.02, respectively)

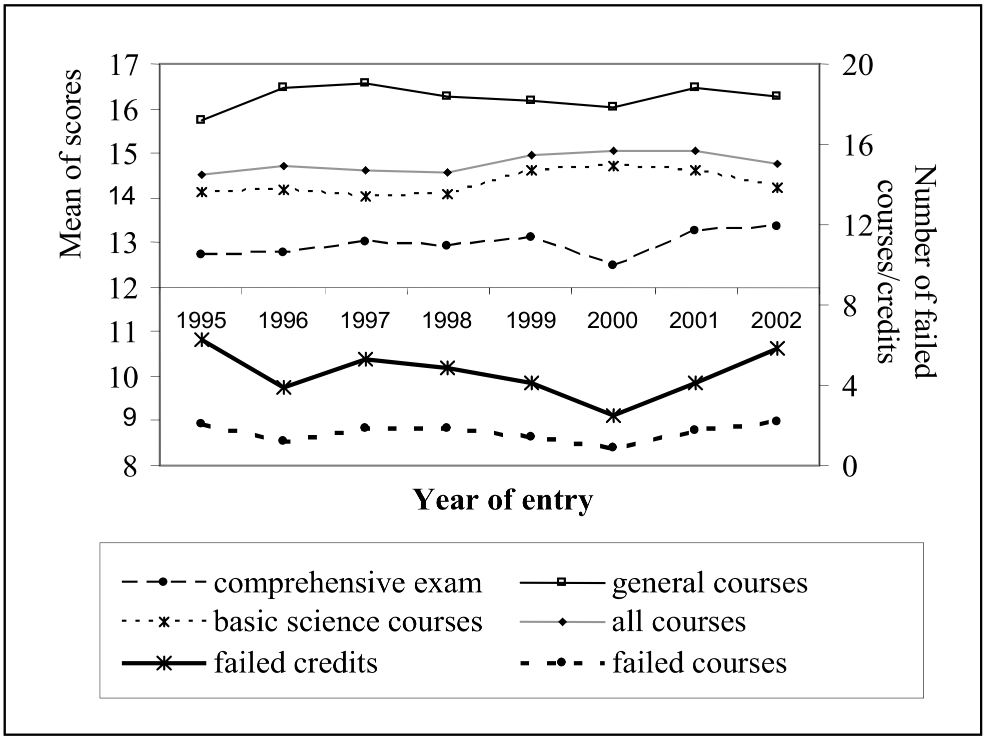
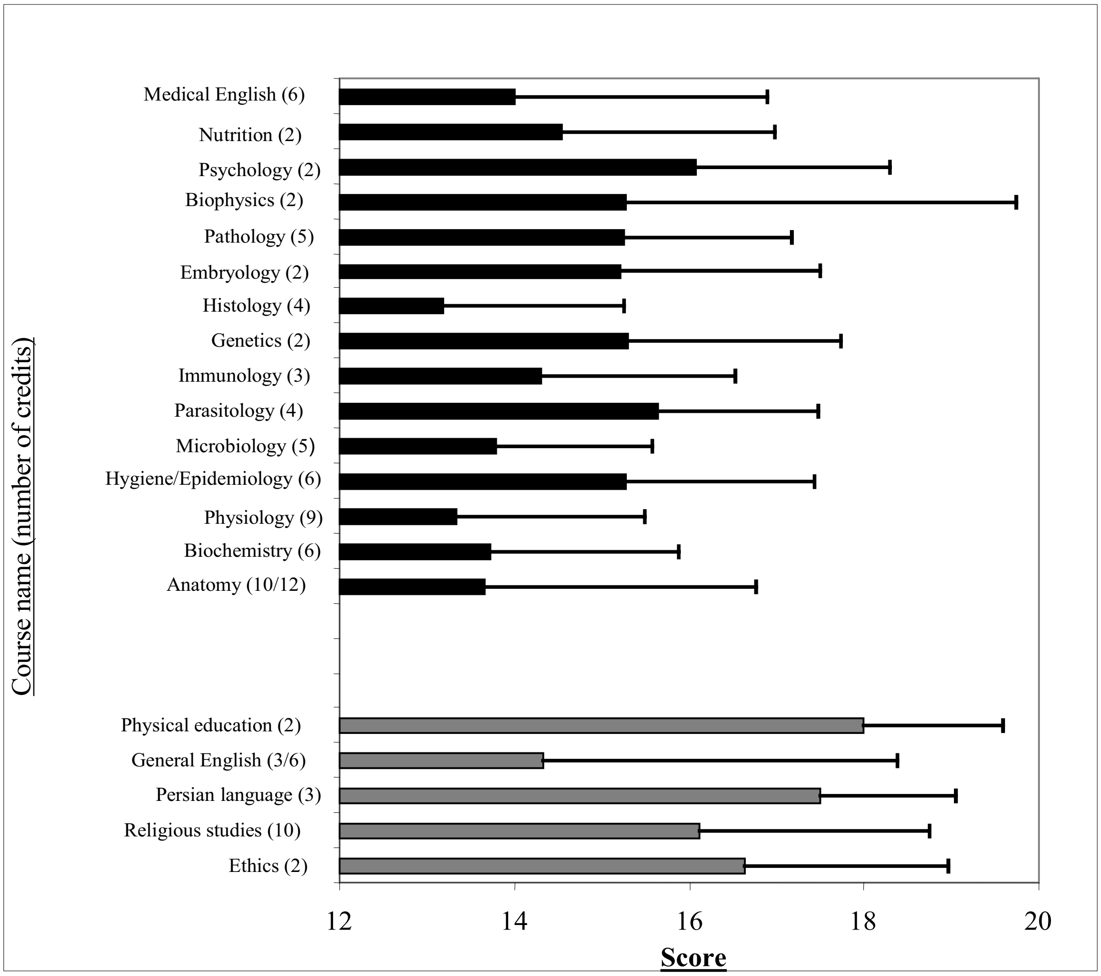


Figure 2: Means and standard deviations of students' scores in basic science and general courses. Numbers in parentheses show the number of credits for each course.



The correlation coefficients between AAIs and the students' scores were more or less compatible with the DIs of courses, particularly in BSCs (Table 1). Generally, the students' scores in GCs had weaker associations with AAIs. This means that the students' scores in GCs had lower internal consistency with their overall scores. In addition, DIs of GCs were lower than those in BSCs; this also implies that the students' scores in GCs had lower predictive validity (Table 1). However, the DIs of general and medical English were relatively acceptable (13.2, and 15.9, respectively). This means that the students' scores in these two courses could discriminate successful and unsuccessful students in the comprehensive exam much better than their scores in other courses could. Although DIs of English courses in males and females, and also age groups did not have statistical significant different, the DIs in males and younger age group were slightly greater.

In contrast, the DIs and the correlation coefficients for psychology, biophysics and genetics were low; i.e., these courses could not discriminate successful and unsuccessful students appropriately (Table 1).

The above finding on low DIs of a few courses was also supported by results of the cluster analysis. The intra-cluster correlations between students' scores in their courses are illustrated in figure 3; strong intra-cluster correlations were observed between anatomy, biochemistry, histology, microbiology, immunology, medical English and physiology (cluster one); and between parasitology, hygiene/epidemiology, pathology and embryology (cluster two). There were considerable distances between scores in nutrition, genetics, psychology and biophysics and scores in other courses. ,These four courses had also low DIs as well.

Females were more successful in their studies based on all AAIs, except the CE score, but the internal consistency of males' and female's scores were comparable. Females' scores in GCs and BSCs were statistically greater than males'(p<0.001). Moreover, females failed in their courses less frequently than males (p<0.001). Nonetheless, in the CE, males' scores were slightly greater than females' scores, but the difference was not statistically significant (12.92 versus 12.9, p=0.84) (Table 2). However, the patterns of DIs in both genders and also their dendrograms were very similar.

Negative associations were observed between the entrance age and the academic achievement (Table 2), and also their internal consistencies. The trend of all achievement indicators showed that the success rate decreased with age (p<0.001). However, generally the DIs in the last age group were weaker, particularly in BSCs.

Table 1. The correction coefficients between the academic achievement indicators and the students' scores in their courses, light, medium and dark shades show strong, intermediate and low values, respectively. Courses are sorted in ascending order of discrimination indices.

General courses	Mean score in				Number of Failed		DI[1]
	CE[1]	GCs[1]	BSCs[1]	ACs[1]	credits	courses	
Physical education	0.013	0.148	0.089	0.109	-0.098	-0.095	0.1
Persian literature	0.189	0.57	0.356	0.423	-0.258	-0.251	3.7
Ethics	0.243	0.507	0.327	0.382	-0.235	-0.253	6.8
Religious studies	0.377	0.888	0.549	0.65	-0.38	-0.39	7.3
General English	0.422	0.597	0.538	0.582	-0.424	-0.425	13.2
Basic science courses							
Psychology	0.2	0.326	0.329	0.347	-0.209	-0.206	5
Biophysics	0.143	0.056	0.257	0.231	-0.104	-0.088	6.6
Genetics	0.303	0.285	0.486	0.475	-0.329	-0.326	7.9
Hygiene & Epidemiology	0.486	0.554	0.736	0.739	-0.54	-0.552	8.7
Nutrition	0.337	0.289	0.571	0.544	-0.358	-0.369	8.8
Parasitology & Mycology	0.542	0.532	0.75	0.744	-0.532	-0.564	11.3
Biochemistry	0.563	0.577	0.815	0.81	-0.625	-0.631	11.3
Microbiology	0.567	0.507	0.78	0.766	-0.533	-0.537	11.3
Immunology	0.481	0.428	0.655	0.642	-0.468	-0.483	12
Anatomy	0.534	0.569	0.847	0.835	-0.588	-0.59	12
Histology	0.544	0.596	0.809	0.809	-0.546	-0.54	12.3
Embryology	0.537	0.474	0.743	0.727	-0.512	-0.524	13.4
General pathology	0.602	0.487	0.774	0.759	-0.461	-0.45	13.7
Physiology	0.644	0.569	0.847	0.837	-0.534	-0.528	14
Medical English	0.534	0.509	0.687	0.688	-0.468	-0.478	15.9

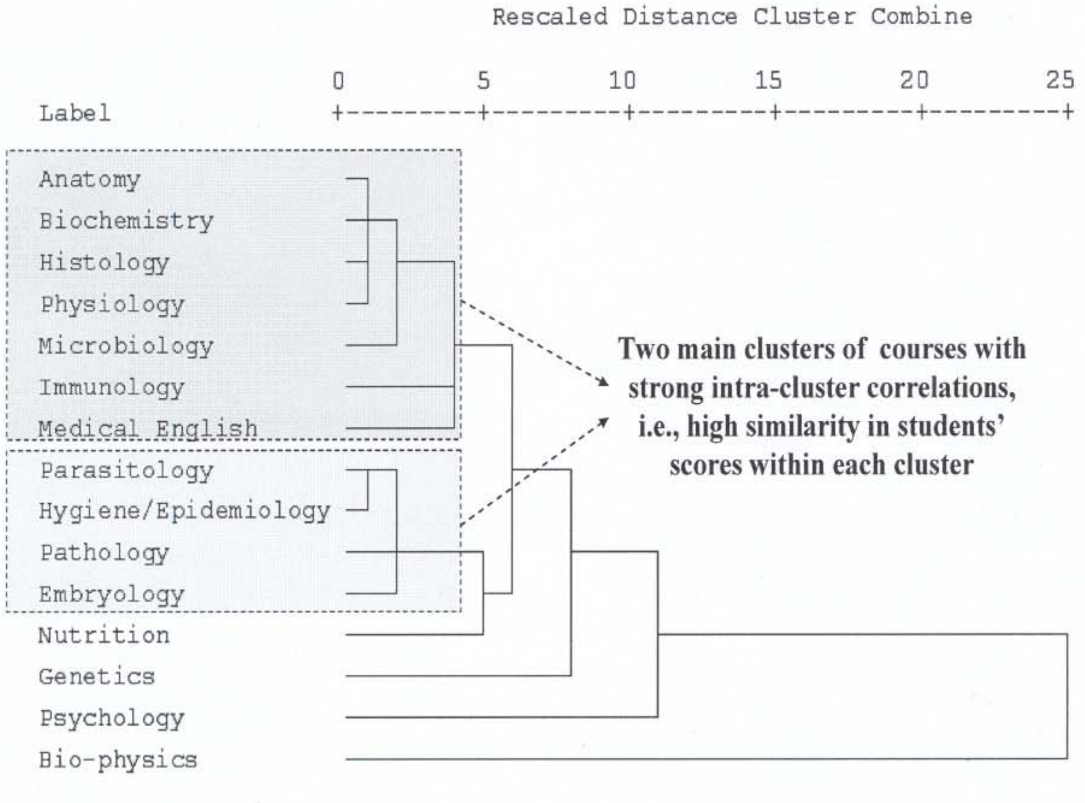
[1] *Discrimination index*
[2] *Comprehensive exam*
[3] *General courses*
[4] *Basic science courses*
[5] *All courses*

Table 2. Description (mean and standard error) of the academic achievement indicators, classified by entrance age and gender

group	Average score in			Number of failed		Score in the comprehensive exam
	General courses	Basic science courses	total	courses	credits	
Gender						
Female (n=335)	16.51(0.07)	14.54(0.08)	15.00(0.07)	1.32(0.12)	3.89(0.34)	12.90(0.09)
Male (n=230)	15.83(0.08)	13.86(0.09)	14.31(0.08)	2.08(0.19)	5.93(0.53)	12.92(0.1)
p-value	<0.001	<0.001	<0.001	<0.001	0.001	0.843
Age group						
<19 (n=222)	16.56(0.08)	14.79(0.1)	15.20(0.09)	0.92(0.10)	2.75(0.33)	13.45(0.11)
19-20 (n=293)	16.10(0.07)	14.03(0.08)	14.51(0.07)	1.91(0.15)	5.41(0.44)	12.63(0.09)
>20 (n=49)	15.62(0.16)	13.46 (0.2)	13.96(0.17)	2.90(0.48)	8.45(1.39)	12.13(0.17)
p-value*	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

**p-values were computed based on the one way ANOVA. The results of post-hoc (Tukey) showed significant differences between all age groups in all variables, except the comprehensive exam scores of the 19-20 and >20 year old age groups (p-value=0.105)*

Figure 3: Dendrogram. Average linkage illustrates correlations between students' scores in their courses. The X-axis shows the distances between students' scores in their courses; shorter lines between two courses implies greater similarity in the students' scores; i.e., less deviance



Discussion

The results show that generally, the temporal variations of AAIs were not considerable between 1995 and 2002, although the variations in BSCs and all courses collectively were statistically significant. The internal consistencies and DIs of scores in GCs were less than those in BSCs. Among the BSCs, there was also a wide range of consistencies and DIs. Students' scores in psychology, biophysics and genetics were less compatible with their scores in other courses. In addition, these three courses had the lowest DIs and internal consistency with other courses. It seems that the patterns of internal consistencies were more or less similar in both genders; however, slightly weaker consistencies were observed in the last age group.

We applied the concept of the DI commonly used in the analyses of question appropriateness to assess the appropriateness of exams. The DI 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 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 scores in the CE, and compared their scores in every course.

Results of the cluster analysis were compatible with the conclusion based on DIs. A dendrogram is a useful graph which illustrates associations between different items and is used commonly in molecular and genetic epidemiology. In the context of this paper, this graph illustrates how strong the associations between students' scores in different courses are. The students' scores in anatomy, biochemistry, histology, physiology, microbiology, immunology and medical English had the strongest intra associations, while students' scores in biophysics, psychology and nutrition had the least intra associations and also considerable differences with their scores in other courses. Interestingly, the DIs in the former courses were high, while the DIs in the latter courses were low. Therefore, we could conclude that the validity of students' scores in the former courses is greater than that is the latter courses. Generally, younger students and females were more successful. There was a strong negative association between entrance age and AAIs. 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 helped them pass the entrance exam much sooner or even right after high school. 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 [7]. It should be added that male students were slightly more successful in the CE, which may imply that their long term achievement is at least in the same level as females.

Most published studies have used regression models to predict the academic achievement of students[5,9-10-11], while we used different statistical methods to assess the association between students' scores. Our methods were more complex and findings were more difficult to present; this was our most important limitation. However, we recommend these methods because they may represent the associations between students' scores in their courses and also address to our research question more deeply.

We suggest that determining the DI and performing a cluster analysis are two useful tools to assess the internal consistency of student scores. Interestingly our conclusions based on both of these two methods were comparable and shows that the students' scores in anatomy, biochemistry, histology, immunology, medical English, microbiology and physiology had stronger consistencies. These patterns were observed in both genders and all age groups. From the medical education point of view, we can state that these seven courses construct a main core among basic science courses.

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. Stone SL, Qualters DM. Course-based assessment: implementing outcome assessment in medical education. Acad Med. 1998;73(4):397-401.
2. Lipton A, Huxham GJ, Hamilton D. Predictors of success in a cohort of medical students. Med Educ. 1984;18(4):203-10.
3. Hoschl C, Kozeny J. Predicting academic performance of medical students: the first three years. Am J Psychiatry.1997; 154 (6 Suppl):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(4):435-47.
5. Dixon D. Relation between variables of preadmission, medical school performance, and COMLEX-USA levels 1 and 2 performance. J Am Osteopath Assoc. 2004;104(8):332-6.
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(1):18-25.
7. 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(39-40):535-40.
8.Whitney DR, Sabers DL. Improving essay examinations: Use of item analysis University of Iowa, 1970.
9. Donnelly M, Yindra K, Long SY, Rosenfeld P, Fleisher D, Chao YC. A model for predicting performance on the NBME Part I examination. J Med Educ. 1986;61(2):123-31.
10. Glew RH, Ripkey DR, Swanson DB. Relationship between students' performances on the NBME Comprehensive Basic Science Examination and the USMLE Step 1: a longitudinal investigation at one school. Acad Med. 1997;72(12):1097-102.
11. Gonnella JS, Erdmann JB, Hojat M. An empirical study of the predictive validity of number grades in medical school using 3 decades of longitudinal data: implications for a grading system. Med Educ.2004;38(4):425-34.