



Investigation of Consistency Between the Students' Scores Using Bayesian Intraclass Correlation Coefficient in Postgraduate Students of Kerman University of Medical Sciences in 2013 - 2015

Fatemeh Mohtasham¹, Yunes Jahani ^{1,*}, Abbas Behrampour¹, Fariba Shariffar² and Abbas Aghaei Afshar³

¹Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

²Department of Pharmacy, Faculty of Pharmacy, Kerman University of Medical Sciences, Kerman, Iran

³Leishmaniasis Research Center, Faculty of Medicine, Kerman University of Medical Sciences, Iran

*Corresponding author: Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran. Tel: +98-3431325421, Email: yonesjahani@yahoo.com

Received 2018 September 15; Revised 2019 February 04; Accepted 2019 February 12.

Abstract

Background: Evaluation of students' scores helps us indirectly examine the status of education system in university departments.

Objectives: In this study, in order to assess the education system, consistency between the students' scores was evaluated by measuring the Bayesian intraclass correlation coefficient (ICC) in postgraduate students of School of Public Health, Kerman University of Medical Sciences during 2013 - 2015.

Methods: This cross-sectional study was conducted on all postgraduate students of the School of Public Health of Kerman University of Medical Sciences during 2013 - 2015. The students' scores were collected from the Office of Postgraduate Studies. First, the Bayesian ICC of students' scores was calculated for all fields. Next, cluster analysis was performed on Master's fields of study, and the Bayesian ICC was recalculated for each cluster. Data were analyzed using R 3.3.2 and OpenBUGS 3.2.3.

Results: Out of 117 postgraduate students, 102 (87.2%) were MSc students, and 15 (12.8%) were PhD students. The highest ICC was attributed to health education (ICC = 0.345) and the lowest to environmental health engineering (ICC = 0.023). Clustering was effective in most fields, and ICC of the clusters increased.

Conclusions: According to the results, consistency between the students' scores was low in the majority of fields; therefore, it is necessary to modify and improve teaching and evaluation methods.

Keywords: Intraclass Correlation, Academic Achievement, Students' Scores, Bayesian, Cluster Analysis

1. Background

In every education system, in addition to addressing the quantitative development of educational services, quality development should be also evaluated (1). Several approaches have been proposed to improve the quality of education, the most important of which is modification of teaching and evaluation methods in the assessment of students' academic achievement. Many experts believe that reforms in evaluation methods, as the most important aspect of education, can significantly improve education.

Evaluation is one of the important components of educational planning. Assessment of academic achievement is also one of the important steps in the teaching process. If this assessment is not carried out accurately, students with good and poor performance cannot be differentiated, and revision and decision-making cannot be applied in sub-

sequent teaching activities (2). Therefore, establishment of an effective system for examining the quality of education provides a means for universities to reform their activities, identify their strengths and weaknesses, and select an appropriate strategy for reform (3). Evaluation of academic achievement is a systematic process, which involves collecting, analyzing, and interpreting information to understand to what extent the students have achieved their learning goals. This type of assessment involves an appraisal of students' learning abilities, used in decision-making about teaching activities of university faculties (4).

In Kerman University of Medical Sciences, the most important university in Southeast of Iran, student assessment is performed by educational groups. Different groups evaluate the students' academic achievement by conducting final exams. It seems that by examining the

scores of students at different educational levels in different courses, it is possible to study the status of the educational system of schools, and in part, the educational status of the university department (5).

One of the important measures for the assessment of learning process is measurement of intraclass correlation coefficient (ICC) of the students' scores. ICC, as well as correlation coefficient, measures the magnitude of the relationship between two variables. ICC is a modification of correlation coefficient; therefore, in addition to consistency, it can take measurement differences into account and assess the agreement of two measurements. Overall, if there are "n" individuals, each with "k" measurements, there are two sources of variation: between members of one group (intra-group) and between two different groups (inter-group).

ICC is an appropriate index for measuring the consistency and similarity of intra-group measurements. It is the ratio of inter-group variance to total variance. The low difference in the scores of each student, besides high ICC, can indicate the validity of exams and scores. If there is no consistency between the students' scores in some courses, its cause should be investigated. Such inconsistencies can be found only by measuring the ICC of scores (6). Generally, in standard exams, strong students are expected to achieve higher scores in almost all courses, compared to weak students. In other words, strong students are assumed to achieve higher scores, and weak students are expected to obtain lower scores; therefore, ICC is high and close to one. In fact, a high ICC indicates that the exam can differentiate strong and weak students.

2. Objectives

In this regard, several studies have been conducted in Iran. However, these studies have been performed using classical methods (frequency-oriented) to analyze the students' scores, and Bayesian statistic has not been measured yet. Overall, evaluation of students' scores helps us indirectly examine the education system of university departments. In the present study, for the assessment of education system, consistency between the students' scores was evaluated using Bayesian ICC in postgraduate students of School of Public Health, Kerman University of Medical Sciences during 2013 - 2015.

3. Methods

This descriptive, analytical study was conducted on postgraduate students (MSc and PhD) at the School of Public Health of Kerman University of Medical Sciences from

2013 to 2015, using census sampling. All students' scores were collected from the Office of Postgraduate Studies. Then, the students' scores in each course were determined as a constant value. If a student repeated a course more than once and had more than one score for the course, the mean score was calculated and considered in the statistical analysis.

To determine the ICC of students' scores, the Bayesian model was fitted to the data. The Bayesian method was used because of the small sample size. Generally, in a small sample size, the Bayesian method is more suitable than the classical method, and the confidence interval is narrower; therefore, the precision of ICC is higher. Also, Bayesian ICC always ranges from 0 to 1, while classic ICC is sometimes negative and not meaningful. With the assumption that the structure of some courses may not be consistent with other courses and that some students show better performance in some courses, clustering method was applied to make sure that the student's score don't need to be similar in all courses. First, the courses included in each field were clustered, and then, for each cluster, ICC was calculated.

The Bayesian statistics was determined based on Bayes' theorem by multiplying the likelihood function by "prior probability" for determining the "posterior probability" (7). The Bayesian ICC was also defined using random effect models. In this study, consistency between the students' scores was measured. Consistency means that if a student obtains a high score in one of the courses, he/she will also obtain high scores in other courses; therefore, the two-way mixed-effect model was used (8):

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \quad (1)$$

Where μ is the total average, y_{ij} is the j th score of the i th student, α_i is the random effect of students, β_j is the fix effect of the course (because the course was specific for students and all courses were included in the analysis), and ϵ_{ij} is the residual of the model.

In the Bayesian method, σ_a^2 is the variance of scores between students, and σ_e^2 is variance in the scores of each student. According to the described model, Bayesian ICC was calculated in the consistency mode using the following equation (9):

$$ICC = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} \quad (2)$$

Bayesian method estimates ICC based on the posterior distribution, using Markov Chain Monte Carlo (MCMC) algorithm (10). After calculating the ICC of students' scores, a cluster analysis was conducted for the courses in each field, and ICC was calculated for each cluster. Generally, clustering is the classification of data in logical groups

so that data in similar clusters have the greatest similarity, and different clusters have the least similarity. In this study, the courses were clustered using Ward's clustering method. This method is considered the most accurate hierarchical method, where the sum of squared difference between each data from a cluster and the mean vector of the cluster is used as a criterion for evaluating the cluster (11).

In this study, the MSc fields included epidemiology, biostatistics, health education, occupational health engineering, and environmental health engineering, while the PhD fields included epidemiology and biostatistics in the School of Public Health, Kerman University of Medical Sciences. For analyzing the data, R 3.3.2 and OpenBUGS 3.2.3 were used.

4. Results

In this study, 117 postgraduate students from the School of Public Health were included during 2013 - 2015. Among these students, 102 (87.2%) were MSc students, and 15 (12.8%) were PhD students. Also, 55 (47%) students were enrolled in 2013, 28 (23.9%) were enrolled in 2014, and 34 (29.1%) were enrolled in 2015. The MSc students included 30 (29.4%) students of epidemiology with 13 common courses, 19 (18.6%) biostatistics students with 9 common courses, 18 (17.6%) students of health education with 12 common courses, 13 (12.7%) students of occupational health engineering with 13 common courses, and 22 (21.6%) students of environmental health engineering with 11 common courses.

On the other hand, PhD students included 11 (73.3%) students of epidemiology with 10 common courses and 4 (26.7%) students of biostatistics with six common courses. In terms of gender, of 102 MSc students, 58 (56.8%) were female, and 44 (44.2%) were male. Also, of 15 PhD students, 9 (60%) were female, and 6 (40%) were male. The ICC of MSc students' scores, with and without differentiation of the enrolment year, is shown in Table 1.

According to Table 1, health education students, who were enrolled in 2013, had the highest ICC. Also, these students had a higher ICC than students of other fields in all entries. On the other hand, environmental health engineering showed the lowest ICC. The ICC of this field decreased from 2013 to 2014, whereas it slightly increased since 2014; however, all three entries had a very low ICC.

Among MSc students, the scores of biostatistics and epidemiology students showed an increasing trend from 2013 to 2014, whereas a decreasing trend was observed in 2015. Also, there was an insignificant difference between the total ICC (all entries) and ICC of each entry in biostatistics; therefore, it seems that the status of students in this

field is stable. It should be noted that Kerman University of Medical Sciences did not accept any students in the field of occupational health engineering in 2014, and there were only two entries; there was a significant difference between the ICCs of these two entries, and a decreasing trend was observed. Moreover, health education showed a decreasing trend from 2013 to 2014, while it had an increasing trend from 2014 to 2015.

After measuring the ICC of MSc students' scores, ICC of PhD students' scores was also calculated. However, due to the low number of PhD students and also the low number of common courses that students had completed during these three years, only the total ICC (without entries differentiation) was calculated for PhD courses. The ICC of PhD students' scores at the School of Public Health, without entries differentiation, is presented in Table 2.

As shown in Table 2, the ICC of PhD students' scores in epidemiology and biostatistics was very low (< 0.1), indicating a very poor consistency between the scores in each field. After examining the ICC, cluster analysis was performed for each MSc field to determine which courses were homogenous with similar scores. Table 3 presents the courses grouped in different MSc clusters.

Clustering of MSc courses indicated that the courses grouped in each cluster had the greatest similarity with each other, whereas they had the lowest similarity with courses in other clusters. Next, ICC for each cluster was calculated, the results of which are presented in Table 4.

As shown in Table 4, the first cluster of biostatistics and the second cluster of health education had the highest ICCs, respectively, indicating a significant correlation between the courses in these two clusters; in other words, the scores of students in these courses were similar. Clustering of courses in occupational health engineering, health education, and biostatistics was effective in finding courses with similar scores.

5. Discussion

The results of the present study showed that by comparing the ICC of students' scores, validity of the scores can be easily determined in different educational departments. The high ICC may be attributed to the students' motivation for better learning, higher faculty experience, or high agreement between the university faculties; in other words, the students' points of view are similar. Low ICC may be related to the substitution of old faculty members with young ones, who use different teaching and evaluation methods. However, confounding variables, such as Internet addiction and social networking addiction, cannot be ignored in recent years.

Table 1. ICC of MSc Students' Scores in the School of Public Health

Field	ICC (95% Confidence Interval)			
	Total Students	Entries of 2013	Entries of 2014	Entries of 2015
Health education (n = 18)	0.34 (0.18, 0.55)	0.42 (0.16, 0.74)	0.29 (0.05, 0.67)	0.35 (0.005, 0.94)
Biostatistics (n = 19)	0.27 (0.11, 0.48)	0.25 (0.008, 0.61)	0.28 (0.001, 0.76)	0.20 (0.002, 0.63)
Occupational health engineering (n = 13)	0.19 (0.05, 0.41)	0.33 (0.11, 0.67)	-	0.11 (0.0008, 0.51)
Epidemiology (n = 30)	0.15 (0.06, 0.26)	0.15 (0.04, 0.33)	0.317 (0.77, 0.03)	0.21 (0.03, 0.50)
Environmental health engineering (n = 22)	0.02 (0.0002, 0.11)	0.09 (0.0006, 0.30)	0.03 (0.0003, 0.22)	0.05 (0.0003, 0.032)

Table 2. The ICC of PhD Students' Scores

Field	ICC (95% Confidence Interval)
Epidemiology (n = 11)	0.03 (0.0005, 0.15)
Biostatistics (n = 4)	0.062 (0.001, 0.39)

The low ICC may be also attributed to the students' low motivation during education, concerns about their future occupational status, and job saturation in the field of study, which may disperse some of the scores and lead to an unexpected ICC. In other words, low ICC may not be only related to the method of teaching and evaluation, and its value and interpretation may be distorted.

In a study by Danesh Kazemi et al., the correlation coefficient between the students' scores of theoretical and practical tooth restoration courses in the School of Dentistry, Yazd University of Medical Sciences, was determined during 1991 - 2012. There was a direct significant correlation between the evaluation score of all theoretical and practical courses (12). In another study by Haghdoost et al. (5), academic achievement of medical students in Kerman University of Medical Sciences was investigated during 1995 - 2003, and it was found that females are more successful in medical courses. Overall, the ICC between male students' scores was greater than that of females. Also, the students' scores in different courses had low consistency with the students' scores in comprehensive exams (5).

Moreover, in a study by Smits et al. (13), predictive factors of successful learning were investigated in postgraduate medical students. They observed that improvement of students' mental health problems significantly contributed to their academic achievement. It seems that in addition to teaching and evaluation methods, establishment of counseling centers and attempts to improve the mental health status of students can affect their academic achievement (13). In another study by Corell et al. (14), effects of competitive learning tools on the medical education of students were investigated. It was found that students, who used the competitive learning tool, had better academic performance and were more satisfied with this

type of learning. Therefore, besides teaching and assessment methods, a healthy competitive environment may contribute to the students' academic achievement (14).

The present study is different from similar studies in this area, as it only focused on the scores of postgraduate students. In this study, it was revealed that the ICC of MSc students' scores in the mentioned fields was low. As shown in Table 1, all ICCs were below 0.5, which represents poor consistency. According to the results, it seems that in some fields with low ICC, improvement of teaching and evaluation methods is essential. Generally, it is suggested to apply new and modern teaching and evaluation methods to increase the ICC and consistency of students' scores. Also, the students' motivation for a more effective learning experience should be promoted.

Clustering in the field of occupational health engineering, health education, and biostatistics was an effective method, which could successfully identify courses with similar scores or classify similar courses in separate clusters. In most fields, the ICC of clusters increased, which shows that some courses were different and that students should not necessarily obtain high scores in all courses. It is recommended to examine more educational groups and more entries in future studies. Also, further research, especially qualitative research, is recommended in departments with low ICC.

The present study had some limitations. The students' data and scores were included only during three years (2013 - 2015), and the sample size for each field of study was small. We used only three entries to remove the effect of faculty member (to have a single faculty member in the course). By including few entries (three entries in the present study), we hoped that the faculty member involved in each course would not change and that consistency and ICC would not be influenced.

Acknowledgments

We would like to thank the Educational Deputy and Office of Postgraduate Studies of Kerman University of Medi-

Table 3. Cluster of MSc Courses in Each Field in the School of Public Health

Field of Study	Course
Epidemiology	
Cluster 1	Epidemiology of non-contagious diseases, social epidemiology in health, seminar, and application of epidemiology in health systems
Cluster 2	Principles of epidemiology, research methodology, epidemiology of contagious diseases, computer-based analysis of health data, and epidemiological methods
Cluster 3	Biostatistics methods (1); biostatistics methods (2); statistical methods in epidemiology, and sampling methods
Biostatistics	
Cluster 1	Survival data analysis in medical research, applied multivariate analysis, design and analysis of clinical trials, statistical methods in epidemiology
Cluster 2	Biostatistics methods (2); biostatistics methods (3); and seminar
Cluster 3	Inference of biostatistics and analysis of categorical data
Health education	
Cluster 1	Health promotion and healthy lifestyle, school health education and health promotion, internship, and health education and communication (2)
Cluster 2	Health education and communication (1); group dynamics, technology and educational methods, communication in health education and health promotion, psychology of healthy behavior, and research methodology in health education and health
Cluster 3	Health sociology and biostatistics
Occupational health engineering	
Cluster 1	Design of lighting in workplace, occupational toxicology, applied occupational toxicology, and evaluation of air pollution
Cluster 2	Work-related diseases, radiation protection in the workplace, design of air pollution control systems in the workplace, and design of noise and vibration control systems
Cluster 3	Applied human-factors engineering (1); design of heat, cold, and humidity control systems, applied human factors engineering (2); modeling in occupational hygiene, and workplace safety (system safety)
Environmental health engineering	
Cluster 1	Management of radiation protection, soil pollution, wastewater treatment plant design, industrial wastewater management, and apprenticeship
Cluster 2	Water treatment plant design, management of water resources development, solid waste management, air pollution control, application of advanced methods in pollutant analysis, and evaluation of the effects of development on the environment

Table 4. Bayesian ICC for MSc Clusters in Each Field

Field of Study	ICC (95% Confidence Interval)		
	Cluster 1	Cluster 2	Cluster 3
Epidemiology	0.08 (0.0007, 0.28)	0.16 (0.004, 0.35)	0.11 (0.06, 0.49)
Health education	0.38 (0.11, 0.64)	0.56 (0.35, 0.75)	0.11 (0.0002, 0.50)
Biostatistics	0.64 (0.43, 0.82)	0.12 (0.0005, 0.44)	0.18 (0.0003, 0.61)
Occupational health engineering	0.32 (0.008, 0.65)	0.25 (0.001, 0.60)	0.33 (0.04, 0.63)
Environmental health engineering	0.03 (0.0005, 0.13)	0.01 (0.0001, 0.08)	-

cal Sciences, who helped us collect the data for this study.

Footnotes

Authors' Contribution: Study concept and design: Fatemeh Mohtasham and Yunes Jahani. Acquisition of data: Fatemeh Mohtasham, Yunes Jahani, Fariba Sharififar, and Abbas Aghaei Afshar. Analysis and interpretation of data:

Fatemeh Mohtasham, Yunes Jahani, Abbas Behrampour, Fariba Sharififar, and Abbas Aghaei Afshar. Drafting of the manuscript: Fatemeh Mohtasham, and Yunes Jahani. Critical revision of the manuscript for important intellectual content: Fatemeh Mohtasham, Yunes Jahani, Abbas Behrampour, Fariba Sharififar, and Abbas Aghaei Afshar. Statistical analysis: Fatemeh Mohtasham, Yunes Jahani, and Abbas Behrampour. Administrative, technical,

and material support: Fatemeh Mohtasham, Yunes Jahani, Abbas Behrampour, Fariba Sharififar, and Abbas Aghaei Afshar. Study supervision: Yunes Jahani.

Conflict of Interests: None declared by the author.

Ethical Approval: The study protocol was approved by the Ethics Committee of Kerman University of Medical Sciences. The Ethic approval code is IR.KMU.REC.1396.2198.

Funding/Support: The manuscript has been funded by Kerman University of Medical Sciences, Kerman, Iran.

References

1. Aghamolaei T, Zare S, Abedini S. [The quality gap of educational services from the point of view of students in Hormozgan University of Medical Sciences]. *Strides Dev Med Educ*. 2007;**3**(2):78-85. Persian.
2. Ghafourian Boroujerdnia M, Shakurnia AH, Elhampour H. [The opinions of academic members of Ahvaz University of Medical Sciences about the effective factors on their evaluation score variations]. *Strides Dev Med Educ*. 2006;**3**(1):19-25. Persian.
3. Fakharian E, Tagharrobi Z, Mirhoseini F, Rasoulinejad S, Akbari H, Ameli H. [Influential factors on results of comprehensive pre-internship exam in medical faculty of Kashan University of Medical Sciences: survey of an 18-year period]. *Hakim Res J*. 2012;**15**(3):203-12. Persian.
4. Mirzaei AR, Kawarizadeh F, Lohrabian V, Yegane Z. [Evaluation methods of the academic achievement of students Ilam University of Medical Sciences]. *Educ Strategy Med Sci*. 2015;**8**(2):91-7. Persian.
5. Haghdoost AA, Esmaeili A. [Educational Achievement in Medical Students Entered University between 1995 and 2003, Kerman University of Medical Sciences]. *Strides Dev Med Educ*. 2009;**5**(2):80-7. Persian.
6. Haghdoost A, Esmaeili A. Internal consistency of medical students' scores in general and basic science exams, Kerman University, Iran. *J Med Educ*. 2006;**9**(1):3-10.
7. Casella G, Berger RL. *Statistical inference*. Pacific Grove, California: Duxbury; 2002.
8. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;**15**(2):155-63. doi: [10.1016/j.jcm.2016.02.012](https://doi.org/10.1016/j.jcm.2016.02.012). [PubMed: [27330520](https://pubmed.ncbi.nlm.nih.gov/27330520/)]. [PubMed Central: [PMC4913118](https://pubmed.ncbi.nlm.nih.gov/PMC4913118/)].
9. Palmer JL, Broemeling LD. A comparison of bayes and maximum likelihood estimation of the intraclass correlation coefficient. *Commun Stat Theory Methods*. 1990;**19**(3):953-75. doi: [10.1080/03610929008830241](https://doi.org/10.1080/03610929008830241).
10. Brandimarte P. *Handbook in Monte Carlo simulation: Applications in financial engineering, risk management, and economics*. New Jersey: John Wiley & Sons; 2014. doi: [10.1002/9781118593264](https://doi.org/10.1002/9781118593264).
11. Rencher AC. *Methods of multivariate analysis*. New Jersey: John Wiley & Sons; 2003. doi: [10.1002/0471271357](https://doi.org/10.1002/0471271357).
12. Danesh Kazemi A, Davari AR, Momeni Sarvestani M. [Correlation between the scores of dental students in theory and practical restoration courses from 1991 till 2012]. *Med Educ Dev*. 2013;**8**(2):57-64. Persian.
13. Smits PBA, Verbeek JHAM, Nauta MCE, Ten Cate TJ, Metz JCM, van Dijk FJH. Factors predictive of successful learning in postgraduate medical education. *Medical Education*. 2004;**38**(7):758-66. doi: [10.1111/j.1365-2929.2004.01846.x](https://doi.org/10.1111/j.1365-2929.2004.01846.x).
14. Corell A, Regueras LM, Verdu E, Verdu MJ, de Castro JP. Effects of competitive learning tools on medical students: A case study. *PLoS One*. 2018;**13**(3). e0194096. doi: [10.1371/journal.pone.0194096](https://doi.org/10.1371/journal.pone.0194096). [PubMed: [29518123](https://pubmed.ncbi.nlm.nih.gov/29518123/)]. [PubMed Central: [PMC5843339](https://pubmed.ncbi.nlm.nih.gov/PMC5843339/)].