

Item Analysis an Effective Tool for Assessing Exam Quality, Designing Appropriate Exam and Determining Weakness in Teaching

Ghadam Ali Talebi¹, Reza Ghaffari², Eyoub Eskandarzadeh³, Ali E. Oskouei^{4*}

¹Department of Physiotherapy, Faculty of Rehabilitation, Tabriz University of Medical Sciences, Tabriz, Iran

²Medical Education Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

³Faculty of Rehabilitation, Tabriz University of Medical Sciences, Tabriz, Iran

⁴Department of Physiotherapy, Faculty of Rehabilitation, Tabriz University of Medical Sciences, Tabriz, Iran

ARTICLE INFO

Article Type:

Original Research

Article History:

Received: 12 Feb 2013

Revised: 7 Aug 2013

Accepted: 23 Sept 2013

ePublished: 15 Oct 2013

Keywords:

Item analysis
Students learning
Assessment

ABSTRACT

Introduction: Item analysis is an integrate component of course assessment which helps observe the item characteristics and improve the quality of the course exam. It also provides a guide for improving the teaching method to enhance the students' learning outcomes. However, item analysis results may not be applied to adjust the way teachers teach and improve the items characteristics. The purpose of this study was to determine the effect of item analysis in improving assessment and teaching quality. **Methods:** The Item characteristics of the final exam for kinesiology course for physiotherapy students in 2 semesters were studied. Improved and good multiple choice questions (MCQs) were then conducted for another semester, followed by application of both good MCQs and improved teaching for the other semester. The item characteristics were compared to observe any effect of good MCQs and teaching on educational performance. **Results:** The good MCQs along with the improved teaching were associated with the greater mean score and the students who passed the exam rather than those with only good MCQs. The percentage of easy questions (42.5%) in students who received good MCQs and improved teaching compared with those (15%) who only received good MCQs indicated that the improved pool of questions were shifted from medium to easier questions. **Conclusion:** We concluded that the item analysis should be followed by revised and improved teaching method. It appears that improved item characteristics are associated with improved teaching method and possibly with an improvement in students' learning.

Introduction

Assessment is a main step in the process of education by which the academic performance of students during a course attendance is tested.^{1,2} It can be considered as an educational tool which determines the competent of students in educational improvement as well as the gap between educational aims and the degree of learning.¹ Multi-choice questions (MCQs) are one of the written question types used usually in theoretically-based course activities.^{3,5} Although MCQs normally assess the low levels of knowledge, they can be improved to assess the high levels of knowledge, understanding, perception, applying, and problem solving provided that they constructed appropriately.⁶ There are some benefits for MCQs as follows: 1) MCQs are more flexible than the other question types and in addition to level of knowledge and justification

capabilities, they assess the judgements of students. 2) MCQs can assess a greater amount of educational aims and course contexts in a limited time period. 3) MCQs can be marked easily and scored electronically. 4) In compared with true/false question type, it is unlikely to mark the correct answer by chance. And 5) if the false answers are written properly, MCQs may diagnose the misunderstanding and educational problems of students.^{3,7}

Question designing and consequently exam conducting is often considered as the end stage of the course activity. However, in order to fulfill the process of education during a course, it is necessary to study and analyze the items (questions) quality. Therefore, item analysis is an integrate component of course assessment which helps to observe the item characteristics and to improve the quality of the overall exam.^{1,8} In other words, exam assessment is a

*Corresponding authors: Ali Eteraf Oskouei, Email: eterafoskouei@tbzmed.ac.ir

dynamic process aimed at improvement the questions and teaching.^{6,9}

Most studies have addressed the validity and reliability of an exam, questions taxonomy, difficulty and discrimination indices and, indeed suggested some advices to improve the exam level. Improvement of assessment quality based on item analysis may include the determination of proper questions, deletion of hard questions, deletion of questions with poor discrimination power, improvement of validity and reliability, and more importantly its application for the next semester. Furthermore, it provides a guide for improving the teaching style to enhance the students' learning outcomes, as the strengths and weakness and misunderstanding during teaching can be revealed by item analysis. Thus, item analysis seems to be crucial in the process of education.

There is no study, to our knowledge, to investigate the effect of item analysis and teaching improvement on the assessment of a course. Therefore, this study aimed at determining the effect of the results of item analysis in improving assessment and teaching quality. In the present study, MCQs questions of the course entitled "kinesiology" for physiotherapy students were analyzed in four semesters over three years. With the assumption that physiotherapy students of our faculty in different semesters are relatively at the same level, we hypothesized that item analysis and the corresponding intervention on the exam and teaching method would affect the students' learning

Materials and methods

The current experimental study was conducted in four stages as follows:

Stage 1

Stage 1 aimed at determining the descriptive statistics of final exam questions in two semesters. Final exam for the course was first performed following routine teaching for Bsc. physiotherapy students in winter semester 2010 (group 1) and then 2011 (group 2). Forty different, but equivalent, MCQs were prepared for the final exam of each group (totally 80 MCQs for both groups).

After conducting the exams and announcing the students' score, the exam questions were analyzed by using the item analysis software customized by the Education Development Centre, affiliated with the Tabriz University of Medical Sciences.

Stage 2

The goal of stage 2 was to construct 40 appropriate and good MCQs based on the descriptive Item analysis of 80 questions in stage 1. We considered the following criteria in constructing a new pool of questions: 1) questions with difficulty index less than 0.3 and more than 0.7 were deleted, 2) questions with poor discrimination (0.0–0.1) and negative discrimination indices were deleted, 3) questions with one or two marked choices, which is a reflection of inappropriateness of distracters choices, were removed from the pool of questions, and 4) questions with

deficiency in content were deleted or revised.

Stage 3

Stage 3 aimed at determining the possible strengths, weakness, misunderstanding about the content, and distortion during teaching based on the descriptive analysis in stage 1. The high difficulty and poor discrimination indices of items, for instances, were taken into account to determine the teacher's weakness in teaching which needed to be improved. In stage 3, the parts of the course that should be highly emphasized or taught differently were all revealed.

Stage 4

In stage 4, fall semester (group 3) 2011 and winter semester (group 4) 2012 were used to study the effect of good MCQs and improved teaching methods on the exam scores. Specifically, the goal of stage 4 was to determine whether descriptive statistics of the improved questions of the group 3 and 4 were different from those in group 1 and 2. Intervention included the routine teaching for group 3 and improved teaching for group 4 with the equivalent MCQs in final exam for both groups.

Final exam was conducted for group 3 and 4 and the questions were analyzed using the same software. Descriptive statistics were compared among 4 study groups.

To regard relatively similar validity and reliability of the exam in all groups, effective factors such as MCQs construction policy, different levels of learning, questions Taxonomy, period of the exam, exam administration, same numbers of questions, and same scoring were considered as much as possible.

Data collection and analysis

In stage 1 and 3, descriptive statistics (mean value, standard deviation, frequency, and percentage) of item analysis were calculated. In order to compare the variables among groups, ANOVA and Toki test were used. Nonparametric statistics (Crosstabs/Chi Square/Phi & Cramers V) were used to compare the nominal variables (pass/fail in exam, and appropriate/inappropriate questions according to the difficulty index) among the test groups. The level of significance was set at ($P \leq 0.05$) in all tests.

Results

Four groups of physiotherapy students at the same level were selected for this research in which the final exam of the course was analyzed. Descriptive analysis of the test groups was demonstrated in Table 1.

No difference was seen in the variables of the group 1 (routine teaching and MCQs) and group 2 (routine teaching and MCQs). However, the number of students who passed the exam was statistically increased in group 4 (good MCQs and improved teaching) compared with group

3 (good MCQs and routine teaching). Similarly, mean scores and the number of easy questions were significantly greater for the group 4 compared with the group 3.

Discussion

The present study assessed the item analysis outcomes of the final exam for kinesiology course for physiotherapy students in 4 different semesters over three years. Our results suggested that the good MCQs along with improved teaching, as done in group 4, were associated with the greater mean score and students who passed the exam rather than those with only good MCQs. Although teachers and some academic institutions perform item analysis computing the statistical variable and indicators, the results may not be applied to adjust the way teachers teach and improve the items indicators for better evaluation of the students.

The finding of the present study suggested an effective role of item analysis that needed to be considered during a course activity. Good MCQs and improved teaching method based on the item analysis variables were associated with the increased number of students who passed the exam with the greater mean score. This may indicate that the routine teaching should be adjusted in such a way that the possible weakness, misunderstanding and distortion during teaching can be corrected. This so called improved teaching may in turn affect the students' learning and lead to an educational progress. Our results indicated the improved pool of questions shifting (to a greater extent) from appropriate (medium) questions

to the easier ones by comparing the percentage of easy questions (42.5%) in group 4 (who received good MCQs and improved teaching) with the parentage (15%) in group 3 (who only received good MCQs) This finding together with the increased mean difficulty index in group 4 (0.65) compared with group 3 (0.54) provided support for the effectiveness of the improved teaching method. As far as we know there is not any evidence that would link the improved item characteristics to an adjusted teaching. Although the item characteristics did not statistically show any difference in group 1 and 2, group 1 exam appeared to be appropriate and its characteristics were also quite consistent with some findings.¹⁰ Needless to say that the idea that appropriateness of an exam is associated with item characteristics has its limitation. For example, students' characteristics, such as IQ, motivation, and learning style, are the facts that need to be taken into account in course activity and item analysis. However, with the assumption that all test groups are equivalent, and due to the lack of better explanation, it seems that improved item characteristics for the kinesiology course might be linked to improved teaching.

All we can say with certainty is that the improved teaching is associated with the appropriate results of the item analysis and possibly with the improved learning outcomes, but the students' characteristics can be associated with the educational progress.

Table 1. Descriptive statistics in four test groups.

Variable	group 1	group 2	group 3	group 4	P value
Number of students	34	23	20	28	
Number of the students who passed the exam (%)	25 (74%)	21 (91%)	11 (55%)*	23 (82%)*	0.03
Mean±SD scores (out of 20)	11.54±2.82	12.32±2.47	11±3.26*	13.28±3.12*	0.03
Number of easy questions (%)	8 (20%)	10 (25%)	6 (15%)*	17 (42.5%)*	
Number of medium questions (%)	30 (75%)	20 (50%)	30 (75%)*	20 (50%)*	0.01
Range of difficulty index	0.02-0.97	0.0-0.95	0.05-0.90	0.03-0.92	
Mean difficulty index	0.57±0.18	0.59±0.23	0.54±0.17	0.65±0.16	0.07

* P ≤ 0.05

Conclusion

Our results suggest that the item analysis should be followed by a revised and improved teaching. These would likely be associated with the students' learning.

Ethical issues

Participants' information was kept confidential.

Conflict of interests

No competing interests to be declared.

References

1. Popham WJ. Classroom Assessment: What Teachers Need to Know. Boston: Allyn and Bacon; 2008.
2. Trice AD. A handbook of classroom assessment. New York: Longman; 2002.
3. Siri A, Freddano M. The use of item analysis for the improvement of objective examinations. *Procedia - Social and Behavioral Sciences* 2011;29:188-197.
4. Muntinga JHH, Schuil HA. Effects of automatic item eliminations based on item test analysis. *Adv Physiol Educ* 2007;31:247-252.
5. Holsgrove G, Elzubeir M. Imprecise terms in UK medical multiplechoice questions: what examiners think they mean.

- Med Educ 1998;32:343-350.
6. Swanson DB, Holtzman KZ, Albee K, Clauser BE. Psychometric characteristics and response times for content-parallel extended-matching and one-best-answer items in relation to number of options. *Acad Med* 2006;81:52-55.
 7. Valiathan P. Overcoming the limitations of MCQs. Singapore: Knowledge Platform; 2009.
 8. Gronlund N E. Assessment of student achievement. Boston: Allyn and Bacon; 1998.
 9. Xu Y, Liu Y. Teacher assessment knowledge and practice: a narrative inquiry of a chinese college EFL teacher's experience. *TESOL Quarterly* 2009;43(3):493-351.
 10. Backhoff E, Larrazolo N, Rosas M. The Level of Difficulty and Discrimination Power of the Basic Knowledge and Skills Examination (EXHCOBA). *Revista Electrónica de Investigación Educativa* 2000;2(1):1-16.