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. 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 judgship 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
dynamic process aimed at improvement the questions and
teaching.\textsuperscript{a,b}

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≤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 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.

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.**

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

* 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**