RDME
Research and Development in Medical Education

TUOMS Publishing Group

CrossMark
click for updates

The relationship between negative stem and taxonomy of multiple-choice questions in residency pre-board and board exams

Mohammad Hassan Karegar Maher[1], Mohammad Barzegar[2*], Masoumeh Ghasempour[1]

[1]Pediatric Health Research Center, Tabriz University of Medical Sciences, Tabriz, Iran
[2]Medical Education Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

Abstract

Introduction: Multiple-choice question tests are considered as one of the most common assessment methods, frequently used in university tests. This study examined the relationship between question taxonomy and negative stem questions in university pre-board tests in Tabriz University of Medical Sciences and national boards tests in internal medicine, general surgery, pediatrics and obstetrics and gynecology residency examination from 2010-2011.

Methods: In this cross-sectional study, 2400 written multiple-choice questions related to the mentioned fields were studied in terms of the relationship between taxonomy levels of the questions and their stems. If there were a negative word or negative concept in the question body, it was considered a negative stem. Taxonomy was graded: taxonomy I, ability to remember facts, Taxonomy II, ability to interpret data and taxonomy III ability to solve a new problem. The data collected were analyzed by SPSS 18. P value <0.05 was considered significant.

Results: A total of 2400 questions from 8 tests (board, pre-board) in 4 fields were studied. In total, 23.1% of pre-board tests and 16.6% of national board tests had negative stems and the difference was statistically significant ($P=0.0001$). In this study 31.1% of questions were designed with positive stems and 63.9% with negative stems in taxonomy level I ($P=0.0001$). There is a correlation between negative stem questions and their taxonomy. This means that 63.9% of negative stem and 31.1% of positive stem questions have been designed in taxonomy level I ($P=0.0001$).

Conclusion: The use of negative stem questions considerably resulted in the design of low-level cognitive questions.

Introduction

Student assessment is one of the most important parts of any education program. A systematic process of collecting, analyzing and interpreting information to determine to what extent the educational objectives are being realized and fulfilled is called assessment or evaluation.[1] The assessment process should offer an appropriate picture of each student's academic progress during different time periods and identify problems and short comings of education. If there is no satisfactory assessment result, it may reflect the low activity of students, failure in planning and teaching or undesirable assessment methods.[1-4] Multiple-choice tests can assess a large number of students in a short period of time. This is one of the best type of objective tests in terms of uniformity of questions, less possibility of blind guesses there is in true-false questions and easy-to-correct answer sheets.

Multiple-choice questions are composed of a stem followed by four or five answers as options, only one of which is the best answer. In correct answers are provided as possible and distractor options. If the multiple-choice questions are designed with accuracy and in respect to testing principles, they could separate students with strong and weak abilities in the cognitive domain.[5,6] To validate multiple-choice questions, researchers developed Millman's principles on the structure of the stem and options, and experts agreed these principles lead to successful academic testing. Based on Millman's principals, if the stem question is negative, negative wording would be distinguished by underlining, italicizing and bolding.[5-7]

The results of studies indicated that students spent more time answering negative stem questions and made more

Corresponding author: Mohammad Barzegar, Email: mm_barzegar@yahoo.com

mistakes compared with positive stem questions.[8,9] However, the use of negative stem questions has been continued and they comprise 15%-36% of multiple choice questions in medical tests.[10,11]

Multiple choice tests are the most common objective tests used in cognitive domain evaluation related to the medical field, and they are frequently used as pre-board exams and board exams for clinical specialties in Iran.

It is necessary that a quantitative and qualitative analysis be carried out after providing test questions, running the test and determining the students' scores to guarantee the proper quality of the questions. Among the qualitative indicators of multiple-choice tests, the evaluation of taxonomy knowledge and respect of structural principles are examined most often.[12]

This study aimed at examining the relationship between negative stem questions and multiple choice question taxonomy in pre-board tests in Tabriz University of Medical Sciences and national board tests in internal medicine, general surgery, pediatrics and obstetrics and gynecology residency in 2010-2011.

Materials and Methods

In this cross-sectional study, Total of 2400 questions from 8 tests (board, pre-board) in 4 fields were studied. The relationship between the negative stem of multiple choice questions and their taxonomy were examined in pre-board and board exams among pediatrics, internal medicine, surgery and obstetrics and gynecology residency programs in 2010-2011. Taxonomy of each question (levels I, II, III) was determined by three experts in their own field and medical education.[13] Each question in every exam was reviewed by the project executive in terms of negative stems. If there were a negative word or negative concept in the question body such as not right, wrong, except, unless, but, least, not likely and forbidden, it was considered a negative stem.

The collected data were analyzed by SPSS 18. A chi-square test was used for the significant differences between positive and negative stems. P value <0.05 was considered significant.

Results

A total of 2400 questions from 8 tests (board, pre-board) in 4 fields (internal medicine, pediatrics, general surgery, obstetrics and gynecology) were studied.

Taxonomic distribution of various tests has been shown in Table 1.

Negative stems were distinguished by underlining negative words or concepts in all questions. The number and percentage of questions with negative stems in Tabriz University of Medical Sciences pre-board and board written exams in internal medicine, general surgery, pediatrics and obstetrics and gynecology for the residency period 2010-2011 are compared in Table 2.

As seen in Table 2, the highest percentage of questions with negative stems were related to the general surgery and pediatrics fields.

Of the total number of test questions, 23.1% in pre- board exams and 16.6% in board exams had negative stems. The difference was statistically significant comparing these two tests ($P = 0.0001$).

The relationship between question taxonomy and negative stems in Tabriz University of Medical Sciences pre-board and board written exams in internal medicine, general surgery, pediatrics and obstetrics and gynecology for the residency period 2010-2011 are shown in Table 3.

As shown in Table 3, there is a correlation between negative stem questions and their taxonomy. This means that 63.9% of negative stem and 31.1% of positive stem questions have been designed in taxonomy level I ($P = 0.0001$).

Discussion

One of the most common problems with multiple-choice questions is a structural form that often allows examinees to guess the correct answer when they are not knowledgeable enough. Not respecting testing principles in the preparation of questions can lead to structural forms that make it more or less difficult to answer questions and affect the proper functioning of examinees.[14] One of the difficulties in preparing multiple-choice questions is that it is sometimes too difficult to provide some distract or wrong options that potentially appear to be right. In such cases, if providing right options is easier, you can use negative stem questions in which all options except one of them are correct.[1] Based on the results of this study, the percentage of negative stem questions in the studied tests ranged from 11.6% up to 35%. Although no standard fixed percentage has been set for negative stem questions, it is cited as 15-36% in some studies.[10,11]

A study by Harasym et al showed that the use of negative words in stem questions made questions easier and increased scores.[15] However, in his study Tamir mentioned that negative stem questions with a high taxonomy level made questions difficult.[16] Other studies have noted that negative or positive words in the questions stem have no effect on the difficulty of questions or assessment of learners' education at higher levels.[17,18] In their study, Harasym et al emphasized the restricted use of negative stem questions and the replacement of simple response negatively-worded questions with multiple response positively-worded questions has been recommended.[10]

The use of negative verbs in the stem makes questions more difficult and confuses the examinee because this type of question forces the examinee to mentally change the question format from negative to positive and then look for the correct answer.[10,19,20] A study by Harasym et al showed that the validity and reliability of multiple response positively-worded questions are higher than simple response negatively-worded questions and those questions assess students' academic performance better. It also indicated that the use of negative words in the question stem should be limited to those questions in which it is important for the examinee to learn what must not be done. However, most learners need to learn what to do and positive words should be used in the question stem

Table 1. Taxonomic distribution of written tests for pre-board of Tabriz University of Medical Sciences and board exams in internal medicine, general surgery, pediatrics and obstetrics and gynecology for the residency period 2010-2011

Test	Pre-board		Board	
	Taxonomy I No. (%)	Taxonomy II-III No. (%)	Taxonomy I No. (%)	Taxonomy II-III No. (%)
General surgery	145 (48.3)	155 (51.7)	63 (21)	237 (79)
Obs and Gyn	107 (35.7)	193 (64.3)	124 (41.3)	176 (58.7)
Pediatrics	164 (54.7)	136 (45.3)	129 (43)	171 (57)
Internal medicine	82 (27.3)	218 (72.7)	90 (30)	210 (70)
Total	498 (41.5)	702 (58.5)	406 (33.8)	794 (66.2)

Table 2. Comparison of questions with negative stems in Tabriz University of Medical Sciences pre-board and board written exams in internal medicine, general surgery, pediatrics and obstetrics and gynecology for the residency period 2010-2011

Test	Pre-board		Board		P value
	Negative stem No. (%)	Positive stem No. (%)	Positive stem No. (%)	Positive stem No. (%)	
General surgery	105 (35%)	195 (65%)	46 (15.3%)	254 (64.7%)	$P < 0.001$
Obs & Gyn	35 (11.6%)	265 (88.4%)	43 (14.3%)	257 (85.7%)	$P = 0.098$
Pediatrics	88 (29.3%)	212 (70.7%)	(21.5%) 65	235 (78.5%)	$P = 0.028$
Internal medicine	49 (16.3%)	251 (83.7%)	49 (15%)	251 (85%)	$P = 0.828$

Table 3. The relationship between the number and percentage of questions with negative stems and the distribution of question taxonomy in internal medicine, general surgery, pediatrics and obstetrics and gynecology for the residency period 2010-2011

Test	Negative stem		Positive stem		OR	CI	P value
	Taxonomy I No. (%)	Taxonomy II-III No. (%)	Taxonomy I No. (%)	Taxonomy II-III No. (%)			
General surgery (pre-board)	72 (68.6)	33 (31.4)	73 (37.4)	122 (62.6)	3.65	2.20 to 6.04	$P < 0.001$
General surgery (board)	24 (52.2)	22 (47.8)	174 (36.1)	215 (84.6)	6.01	3.07 to 11.77	$P < 0.001$
Obs & Gyn (pre- board)	23 (65.7)	12 (34.3)	174 (36.1)	180 (67.9)	4.06	1.93 to 8.54	$P < 0.001$
Obs & Gyn (board)	31 (68.9)	12 (31.1)	174 (36.1	164 (63.8)	4.56	2.23 to 9.30	$P < 0.001$
Pediatrics (pre-board)	70 (79.5)	18 (20.5)	174 (36.1)	122 (57.5)	5.21	2.94 to 9.46	$P < 0.001$
Pediatrics (board)	40 (41.5)	25 (38.5)	174 (36.1)	146 (62.1)	5.27	1.49 to 4.61	$P = 0.001$
Internal medicine (pre-board)	20 (41.6)	29 (60.4)	174 (36.1)	189 (75.3)	2.10	1.11 to 3.98	$P = 0.029$
Internal medicine (board)	(48.9) 22	(51.1) 23	(36.1) 174	(73.3) 187	2.63	1.34to 5.02	$P < 0.001$
Total	302 (63.9)	174 (36.1)	174 (36.1)	1325 (68.9)			

of these questions. Choosing the negative stem option, on the other hand, does not mean that the examinee is aware of the right aspect of the action, which is the learning goal.[10]

Based on the results of this study, 31.1% of positive stem questions and 63.9% of negative stem questions were designed in taxonomy level I. ($P = 0.0001$). Although designing questions in taxonomy level I does not necessarily result in an easy question, it encourages residents to learn the content by memorizing the subject, while the main goal in clinical fields is interpreting and analyzing data and problem-solving. It appears the problem of overcoming memory based question (taxonomy I) in negative stem questions in this study not only taints test validity but also pushes students to superficial learning and memorizing. Although words with negative concepts are considered structurally acceptable based on the instructions of the ministry, the results of this study indicated that negative stem questions were significantly designed in taxonomy level I.

Study limitations
Reviewing the relationship between negative stem ques-

tions and quantitative results of questions was not possible due to the unavailability of difficulty index and discrimination index.

Conclusion
The negative stem questions significantly result in the design of low-level cognitive questions. It is recommended that the use of negative words in question stems should be limited to those that are important for the examinee to learn what must not be done and negative stem questions should be considered structurally undesirable.

References

1. Seif A. Educational measurement, Assessment and evaluation. 5th ed. Tehran: Doran; 2008. p. 30-41

2. Guilbert JJ. Education Handbooks for Health Personnel. 7th ed. Geneva: Word Health Organization;1998. p. 237-254

3. Azizi F. Medical education, mission, vision, and Challenges. Ist ed. Tehran: Shahid Beheshti University of Medical Sciences; 2003. p. 659-705.

4. Zolfaghari B, Asadollahi GH. Academic achievement tests in medical sciences. Isfahan: Isfahan University of Medical Sciences; 2000. p. 35-52

5. Burton SJ, Sudweeks RR, Merrill PF, Wood B. How to prepare better multiple choice tests items: Guideline for university faculty. Birgham Young University Testing Services; 1991.

6. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. Applied Measurement in Education 2002; 15(3):309-34. doi: 10.1207/S15324818AME1503_5.

7. Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences. Philadelphia: Pa National Board of Medical Examiners; 1998.

8. Colosi R. Negatively worded questions cause respondent confusion. In Proceedings of the Survey Research Methods Section. American Statistical Association; 2005. p. 2896-903

9. Weems GH, Onwuegbuzie AJ, Collins KMT. The role of reading comprehension in responses to positively and negatively worded items on rating scales. Evaluation & Research in Education 2006;9(1):3-20. doi: 10.1080/09500790608668322.

10. Harasym PH, Price PG, Brant R, Violato C, Lorscheider FL. Evaluation of negation in stems of multiple – choice items. Eval Health Prof 1992;15(2):198-220. doi: 10.1177/016327879201500205.

11. Barzegar M, Bilan N, Karegar Maher MH, Shiva S, Sayyah Melli M, Tabrizi A. Comparison of multiple-choice questions in quality parameters of pediatric residency tests between the pre-board examination of Tabriz University of Medical Sciences and national board examination in 2007 and 2011. Res Dev Med Edu. 2014;3(1):31-36. doi: 10.5681/rdme.2014.008.

12. Kehoe J. Basic item analysis for multiple-choice tests. Practical Assessment, Research & Evaluation 1995;4(10):121.

13. Sayyahmelli M, Barzegar M, Bilan N, Aslanabadi S, Khoshbaten M, Ghasemzadeh A, et al. Comparison multiple-choice questions quality parameters of pediatric, general surgery, internal medicine and genecology and obstetrics residency tests between preboard examination of Tabriz University of medical sciences and national board examination in 2010 and 2011. Journal of Medical Education Development 2015;8(18):43-53.

14. Brady AM. Assessment of learning with multiple-choice questions. Nurse Educ Pract 2005;5(4):238-42. doi: 10.1016/j.nepr.2004.12.005.

15. Harasym PH, Doran ML, Brant R, Lorscheider FL. Negation in stems of Single response multiple-choice items: An overestimation of student ability. Eval Health Prof 1993;16(3):342-57. doi: 10.1177/016327879301600307.

16. Tamir P. Positive and negative multiple choice items: How different are they? Studies in Educational Evaluation 1993;19(3):311-325. doi: 10.1016/S0191-491X(05)80013-6.

17. Downing SM, Dawson–Saunders B. The psychometric effects of negative stems, unfocused question, and heteregenous options on NBME part I and part. item characteristics. Chicago: The Annual Meeting of The National Council on Measurement on Education; 1991.

18. Rachor RE, Gray GT. Must all stema bee green ? A study of two guide lines for writing multiple choice stems. New York: The Annual Meeting of the American Educational Research Association; 1996.

19. Violato C, Marini AE. Effects of stem orientation and completeness of multiple choice items on item difficulty and discrimination. Educational and Psychological Measurement 1989;49(1):287-95. doi: 10.1177/0013164489491032.

20. Dudycha AL, Carpenter JB. Effects of item format on item discrimination and difficulty. J Appl Psychol 1973;58(1):116-121. doi: 10.1037/h00