

## **Analysis on the quality of test items on disaster management block**

**<sup>1</sup>Rosaria Indah, <sup>1</sup>Mulyadi, <sup>1</sup>Reza Maulana**

<sup>1</sup> Faculty of Medicine, Syiah Kuala University, Banda Aceh Indonesia

**Abstract.** Written exam is the format most often used to assess the extent of student learning achievement. The advantages of written exam are more easily prepared and relatively more economical when compared with clinical examination test. Block's written exam, especially on disaster management block is considered important because it will illustrate the students' competence at the end of this special block. This study is a qualitative research, aimed to evaluate the quality of test on disaster management block. The quality of the test were evaluated using qualitative method by deep interviewing 5 criteria of students. The result was analyzed using Miles and Huberman method and the themes emerged were: 1) The test item in disaster management block were fairly easy 2) Test items were in accordance with the competence of undergraduate, 3) The test item were congruent with the content material taught in this block. However, there were some weaknesses, including: 1) There were repeated items, 2) There were format of question that complicates matters without a clear purpose, 3) There were 'easy to predict' item questions, 4) Lack of variation of format. These problems might cause low discriminant items. Recommendation including continues improvement of item developers' capability by integrate it within curriculum for faculty development programs.

**Keywords:** disaster management, test item, medical competence, faculty development.

### **Introduction**

Disaster management block is the last of 21 blocks in Competency-Based Curriculum in Medical Science with Problem-Based Learning method at Faculty of Medicine, Syiah Kuala University. The duration of this block including 7 weeks including 1 week of evaluation, worth 5 credits points. This block's objective is to provide a thorough understanding and skills that appropriate as well as practical to our future doctors in the field of disaster management. This block also emphasizes the importance of good cooperation between the medical professions with entire community in preventing, responding and recovery many kinds of disasters. With the method of instructions to achieve the objectives, we expect the students will have the mindset that in disaster, medical profession may not work alone but instead we should be in a system that is able to work with anyone . Nevertheless, the ability of

medical professionalism must still be put forward, with constantly updating knowledge and skills through training so that the role of the doctor will soon be a major part in patient care and patient safety at every disaster that can happen anywhere and anytime, given the geographical location of Indonesia which makes it located on disaster-prone areas .

We have gone through various disasters. The earthquake that struck off the west coast of Indonesia on December 26, 2004 is the most powerful recorded in more than 40 years. It registered more than 9.1 on the Richter scale , lasted for a few minutes and create huge tsunami waves. Tsunami devastated coastal areas of 13 countries in the Indian Ocean , including Indonesia , Sri Lanka , South India , Thailand and others. Damage was reported as far as the east coast of Africa . More than 283,000 people were killed or reported missing. Approximately 1.7 million people lost their homes, with two million became refugees (1). Most of the region of Aceh and everything in it was destroyed in a matter of minutes , which left weeping bitterly because so many are dead, missing or injured and countless possessions lost, or damaged.

In particular, the cities of Meulaboh and Banda Aceh severely affected . One month after the tsunami , it is estimated that the death toll had reached 166,760: 127,749 people missing and 1,736 people were hospitalized with injuries. Damage to the village and building large structures also occur . At that time, Zainoel Abidin General Hospital in Banda Aceh was severely damaged by the tsunami (2). Some hospital staff were killed and others are no longer able to work.

It has been very obvious that the Aceh region are in need of health professionals who are competent on disaster management. Disaster management aims to reduce, or avoid, the potential losses from hazards, assure prompt and appropriate assistance to victims of disaster, and achieve rapid and effective recovery (3).

Discussion of disaster management in the curriculum should also include all phases of the disaster cycle. Students should also be introduced to the overall phase, both before, during and after a disaster (4).

Jones (2001) stated that improving the quality of health care, has implications for the design and development of medical education. As a result, the classical method which has been used to teach medical students need to be updated and improved in order to meet these expectations (5). In order to meet international standards, we need to educate clinicians to have an understanding of local culture and skills needed to conduct health-related natural disasters. Here are 2 things that particularly focus on this block .

More broadly, we could observe the global changes in the health care system, shifting from curative care to be more toward preventive. People demand health services with better quality (5). Surely these demands can be met one way to improve the depth and quality of medical education .

Awareness of disasters often reactive, resulting from the disaster experience. The impact of the earthquake in 1985 in a hospital in Mexico have accelerated the formulation of guidelines for the construction of health facilities in accordance with the mitigation measures against natural disasters around the world. Similarly, in the United States, during the attack on the World Trade Center, has accelerated the preparation of guidelines for hospital preparedness against terrorist threats (7). One

important effort in promoting human resource capacities in relation to disaster management is education-and in particular, the curriculum, either in the form of courses or blocks, which specifically addresses disaster management.

Motivation for designing a block on disaster management could come from many things. It could be a motivation comes from the desire to implement new teaching techniques, availability of facilities or new resources and the need to redesign the program to accommodate the specific character of the students who are now becoming more active than the previous generation (8). this time motivation came from the desicion maker of our faculty (senate and dean) to achieve goals and increase student achievement. Inspiration for improving the courses can also be derived from the evaluation of a course/block which considered achieved less satisfactory. In relation to curriculum design, we had several motives, including administrative duties of Dean, also the desire to improve and implement a better instructional method. These reasons supported by various stakeholders, including the government's health offices. They hope our medical faculty could train medical doctors on disaster management.

To measure the achievement of the students we use several method of students' assessment, one it them is written test. Written test is a format that should be optimized to ensure competency regarding it's objectivity. We should always remember that the test has several main objectives , namely : 1)Optimizing the ability of learners to provide motivation and direction for future learning, 2) Protect the public by identifying as early as possible those who are not worthy to be a graduate, e.g physicians, 3) Setting up the basic knowledge for learning to the next level (master or doctoral level) (9), and 4) Develop a measurement method that has high reliability, provide feedback on students' contribution (10).

Exams can be very influencing the behavior of students and faculty. Students will try to get the best grade, or at least they want to pass the courses they follow. Professors want students to master the subject well. Any exams tend to force students to increase the time and effort to learn the things that will be included in the material to be examined, and exams are also likely to reduce the time and effort to learn the things that have been known to not be included in the test. Consequence of the combination of the desire of students and lecturers are professors will be prioritized to teach things that are easy to understand in the test. Further due to a failure on teaching high-stakes competency, this kind of competencies rarely tested. More professors put more low-stakes competencies to be examined and underestimate the high one. This is also contribute to the test bias (11).

Written exam is the format most commonly used to assess the extent to which the achievement of student learning. The advantages include: it is more easily prepared and relatively cheaper/economical compared to the skill test. Yet it is understood that to examine most of medical competence will required both format, written and verbal format. It must be realized also that the decision of passing should not rely solely on the MCQ score (12)

There are various forms of written question: short-answer open-ended question, essay question, a simple true-false questions, multiple true-false question, extended matching question(EMQ), key-feature approach and the last is multiple -choice

question(MCQ) format that is most popular among other formats. Of course, each format has its advantages and disadvantages that the test designer must be aware of, as well as the indications and contraindications also rules about proper preparation

MCQ consists of a problem followed by a series of suggestions for solving them. Problem can be either direct question or an incomplete statement called the stem. A series of troubleshooting tips can be words, numbers, symbols, and phrases called choice answers. Students asked to read the selection of choices and then choose the one correct or most correct one. Correct answer called correct choice, while the wrong answer choices are called distracter (13). MCQ has the advantages that it is relatively economic, easy to answer, relatively easy to examine and excellent reliability. With this form we can test multiple samples from more extensive learning materials compared to other forms of the written test (14). The drawback is relatively more difficult to compile MCQ (taking into account the terms of the question either). It is more suitable to use if we have extensive learning material and large number of test participants (15). MCQ also should not be used to measure the performance /skills. OSCE (objective structured clinical examination) in this case is much more appropriate than written exams.

MCQ is usually used to test a simple memorization and factual knowledge, although now still can be used to test a higher level of knowledge, for example the understanding of a concept and the reasoning of the clinical condition (clinical reasoning). MCQ can be used to test the ability of a higher level and the answer choices designed to really consider the principles of good drafting to avoid questions about the completion of which is not able to distinguish(discriminate) which students have learned well, which one with less well studied and very little learning (16).

Another drawback of MCQ is 'cueing effect', the tendency of test item to show the reader the correct answer on the answer choices, among many wrong answers. There are tendency to give instructions on the matter which is not obtained in the form of open ended question(OEQ). It's hard to prove no difference between guessing the correct answer (cueing) to guess, scientifically. It must be recognized also that guessing also occur in the form of OEQ. MCQ still more often used because OEQ considered more ambiguous and has lower reliability. With all the advantages and drawbacks, MCQ exam remains a form most commonly used (17).

Students often think that the earliest answers they gave in answering MCQ questions is the correct answer. This causes the student does not want to change their answers, even if had thought of other options that might be true. The fact that many tests have proven that students think twice to change the answer option, then most of them obtain a higher value after changing the answers once. But the score does not change much after the exam a second improvement, a third, and so on (18).

Many institutions have a policy to establish a secret question bank and could not be accessed by students .

Schuwirth and van der Vleuten (2006) write the rules of MCQ good preparation, including:

1. It should have extrinsic ambiguity , not the intrinsic ambiguity . Problem should be able to distinguish between students who really understand the concept and the students who only understand some of the concepts (extrinsic ambiguity).

2. Question should not have tricky words, also did not invite different interpretations (intrinsic ambiguity ) (19).
3. It should have only one most correct answer. Other options should be a distraction that also equal and plausible, not completely different, too long or too short when compared to the correct answer (13)
4. Avoid the use of fillers e.g ABC right, none of the above) and others.
5. Avoid unnecessary additional complexity , e.g do not use any form of 1,2,3 are correct. Students often get stuck when selecting the combination of an answer, and there are not related with physician competence. Sometimes this form of giving "instructions answer" for instance, if number one is true then it is definitely number three is also true. It make us unable to distinguish between students who understand the concepts and the students who are good at utilizing tricks (12).

Haladyna, Downing and Rodriguez(2002) has developed a guidelines in preparation MCQ consists of 31 items in 5 main aspects that should be considered in formulating an of MCQ test. These aspects related to the management of course material, suppose each test point should examine the specific content, including mental attitude associated with it, put important content to be mastered and avoid content that is too easy/not important, using a paraphrase sentence to describe the things you want to test students who only remember the memorized content, arrange each question independently and not dependent on other item, avoid content that is too specific or too general, avoid questions that tend to lead to personal opinion, avoiding tricky question, keep the suitable language that appropriate for group of students who will be tested. Another aspect is the aspect about the format. Problem should use a question, or simply asked students to complete, or choose most correct answer, or true-false form, matching question, EMQ, context -dependent item/key feature but avoid complex MCQ format, compose about vertical sequentially, not horizontal.

Writing style should also be put into consideration, including appropriate language structures, placement of dots, capital letters and correct spelling, minimizing the number of words to be read on each question .

Aspects of the preparation of the body part (stem) : make sure the direction of stem questions to be very clear about, the main idea should be on the stem, avoiding words that give excessive description, using positive words, avoid negative words and minimize the use of "exception" unless it is in capital word ('EXCEPT' and not 'except').

Aspects of the writing of answer choices : develop as many as effective options, but research shows 3 pieces is the most appropriate option, make sure that the answer is only one, to distinguish the location of the correct answer, put it in a logical order, ensure choice must be independent, should not be overlapped, the words used among the answer choices should be the same as the structure, keeping the length of the sentence by the same option , be aware of the option "none of the above" , avoid "All of the above", collate option in the positive, not in the negative form, avoid provide guidance in the selection of the correct answer, the wrong answer choices should make sense, identify the causes of errors students in choosing the wrong answer choices, use humor if it fits the style of learning situations (20).

Quality problem can be judged from the degree of difficulty (p) and about the ability to distinguish between students who truly understand the lesson with students who are not yet perfectly understood (faulty grasp) is called the level of discrimination (D). Determination of the level of difficulty and the level of discrimination is done in a matter of item Analysis (19). Test-item analysis should be conducted on an ongoing basis and should be performed by a designated assessment teams in the faculty structure .

## **Materials and Methods**

This research used a qualitative approach that includes students selected to represent the different views. Data collection through structured interviews were equipped with the interview guide addressing the research question. This study conducted on June to December 2011 on medical education courses at the faculty of medicine that has been implementing a PBL curriculum since 2006 .

The sample student class of 2007 involved in disaster management block. The sampling technique used was snowball sampling. Data collected by triangulation technique which means collecting data from different sources, but the same technique until data saturated. The data collected through structured interviews using open-ended question to guide the interview. The questions focused on students' perceptions about test items in block's exam. Interviews were recorded with a tape recorder with the permission of the participants, and then transcribed verbatimly and then analyzed and ultimately generate themes.

Researchers act as active participants in data collection, intends to understand in detail how students think and how they develop their views. Interview guide contains topics to help researchers explore, see and ask questions to get information and confirmed it. Therefore, researchers are always free to build and establish a conversational style, albeit with a particular focus on pre-determined.

Analysis of the data in this study using the model of Miles and Huberman revealed that activity in qualitative data analysis is done interactively and continues over time until complete, until the data is already saturated. Activity in the analysis of data including data reduction, data display, and drawing conclusion.

There are 15 interviews in 5 groups of students (high-scorer, low-scorer, enthusiastic, unenthusiastic and unique students).

## **Results**

The themes that emerged were :

1. The test items in disaster management block were fairly easy. Most of the interviewees argued that the problems in the block are easier to answer than the questions in the previous block. This is due to the many questions that simply test factual knowledge, only memorization phase, yet to test comprehension.  
*"Test item were easy, too many items on basic theory, testing our ability to memorize things, repeated items and less variations.." (U2 :8 - 9 ) .*
2. Test items were in accordance with the competence of undergraduate. Students think that the questions that came out in the test had been in accordance with the material being taught. The material taught also in accordance with the

competence of medical undergraduate, not of higher level (e.g master). As expressed by one of the student groups 'low scorer':

*"..I think the questions were appropriate to test the undergraduate competence" (K1:12-15).*

3. The test item tested the content material taught in this block. Students explained that the answer to the problems they see during final exams actually can be learned from a variety of learning formats , either lecture , practical sessions and tutorial discussions.

*"....The test item tested the subject discussed in tutorial discussion, or in the lectures. I also had the chance to implement the knowledge I had within the practical sessions so it made me easier to remember and understand..."(U1:9-12).*

4. There were repeated items.

From several interviews revealed that there is a recurring items as mentioned by some students :

*" ... I think they should fix some recurring items. I hope it can be replaced with better items ... " (A2 :191 - 192) .*

5. There were format of question that complicates matters without a clear purpose. The questions were prepared by using the filler such as 1and 3 are true, ABC are correct, ABC are incorrect or none of the above. They were considered complicate matters, also impressed as tricky question or gave impression that the test constructor could not make another option.

*"....those kind of question format made it more difficult to answer, I prefer an ABCDE format" ( U2 :29 - 30 ).*

There are also some problems which become more difficult to answer because it uses English rather than Indonesian language. Students find it difficult to answer not because they do not master the material but because it was not sure of their understanding of a foreign language.

*" ..... The drawback is that there are a few questions in English... " ( A2 : 162-163 )*

6. There is no question that mis-keyed

*" ... It looks like there was an answer in every questions. No matter one does not know the answer ... " ( A1 :28 - 29 )*

7. There were 'easy to predict' item questions.

*"Some questions were easy to guess what is the answer, such as the item tested the definition, meaning of an terms. Relatively easy for me. I can directly answer them except items that story-based triage, this kind force us to prioritize which one to be handled first"(A2:82-87).*

*"I think question number 89 can be answered by logical thinking, even without learning within this block. Some questions are so easy we can just guess" (K1:65-67) .*

8. Lack of variation of format. It is talking about considering the variation form of MCQ

*" ... I prefer an item using cases, closer to the reality. I believe on the field, we deal with cases, not theories. Theory could be built within a case" (A2 :90 - 93)*

## **Discussion**

Most of the content of test item on disaster management block in 2011 seemed have not been able to discriminate between students who learned a lot and student that doing less studying. There are a large number of questions that were too easy, asking the students recall and endors rote memory about things that exist in the low competence . If this happens to the materials that are ranged on first level competency (knowing, identify) they do not need to be fixed. But if it happens to the materials that were considered to be very important to master, they will need serious improvement (20) even to the extent of level 4 competence (able to apply independently without supervisors) so it will need repairs to items such questions before it is given again in the coming years to improve its discriminant level (19).

The forms of questions in the test need to be improved, not merely continued the tradition of exams in other blocks. The test item constructor should consider the objective, goals and outcome of this block, which were to equip students with disaster management knowledge, skills and attitude. Skills certainly can not be tested by written examination, and require performance test. The block designer should also design performance test in this block that are congruent with the objectives of the block (21).

Given this block is the final block of the undergraduate medical curriculum, an assessment of the overall competence of the students should be integrated within this block . One of the famous and assessment instruments recommended to include many facets within the competence is portfolio (22). Training of teaching staff to assess through a new test method would also be required if the decision-makers at the level of faculty agreed to impose a new method (23).

The students' comment on some 'easy to gess' items should be considered an important point of improvement for test item costructor. They should guard against the deterioration of the quality of test item in this block. Cueing effect that makes students tend to 'guess ' the correct answer easily , should be minimized in the preparation of the next block test (17), as well as items that tend to complicate without a clear purpose as fillers like "ABC right", none of the above or even items with the use of foreign language (English) made the items became more difficult to answer without any correlation with interest to assess the competence of students (12).

Preparation of stem often ignore the principles of language and psychological effects. Sentences without subjects often found. This tends to confuse students (20). It also often ignore the psychological effects when using the word 'except' or 'not' . These words will trigger negligence if not written in capital. Should be 'EXCEPT' or 'statement below is NOT included .... ' (20).

Exam will be deemed easy if congruent with the learning materials . Tutorial discussions, lectures, lab work, practical sessions should be very clear that the



material will allow students to understand the content and achieve learning goals (24).

Increase the validity of the content is also required. The test item is valid if it test the desired competency. There are still some items that tend to test the competence of the lower stakes than desired (e.g only test ability to memorize, when it should test the comprehension) or even another competency test , e.g English language. Lack of validity items will lead to a lack of validity of the competency test , bring on the low validity of the overall assessment system (13). The change in learning should be followed by changes in assessment (21).

Improved reliability of the test items is necessary too, by measuring the difficulty index (CI), the index of discrimination (ID) and the correlation (r) following the examination (19). For the measurement of three parameters quickly we could use a tool such as a computerized item scanner. Procurement of these machines is a big cost , but it is very necessary to improve the quality of the question.

Overall curriculum improvements require support from human resources. Teachers as center variable of the quality of education should be trained in a continuous faculty development/training program, has a specialized curriculum to develop optimal method of instruction that be interesting and not boring, also professionally managed by its organization of structure at education office (23).

## **Conclusion**

From this study it can be concluded that the test item in disaster management block were fairly easy and in accordance with the competence of undergraduate. The test item were congruent with the content material taught in this block. However, there were some weaknesses, including: existance of repeated items, format of question that complicates matters without a clear purpose, there were 'easy to predict' item questions, and lack of variation of format. These problems might cause low discriminant items. Recommendation including continues improvement of item developer capability by developing curriculum on faculty development programs.

This shows the importance of continuous training for block and item designers on ways to design a good block, also the quality of assessment. Staff Development Curriculum is a document that should be drafted as a start to begin an improvement. After the special staff training on the preparation of a items it could be use as a material to help developing a standard operating procedure for development of test items

## **Acknowledgements**

We thank HPEQ project that support this study financially and the participants of this study who contribute their best time and effort.

## References

1. Levy JK, Gopalakrishnan C. Promoting disaster-resilient communities: the great Sumatra-Andaman earthquake of 26 december 2004 and the resulting Indian ocean tsunami. *Water Resources Development*. 2005; 21(4): p. 543-559.
2. Riddez L, Kruck M, Gardarsdottir H, Redwood-Campbell L. The surgical and obstetrical activity at the ICRC Field Hospital in Banda Aceh in the aftermath of the tsunami. *International Journal of Disaster Medicine*. 2006; 3(1): p. 55-61.
3. Jamal F. *Buku Tutor Blok 21 Disaster Management Banda Aceh: Fakultas Kedokteran Universitas Syiah Kuala*; 2011.
4. Ireland M, Ea E, Kontzamanis E, Michel C. Integrating Disaster Preparedness into a Community Health Nursing Course. *Disaster Management & Response*. 2006; 4(3): p. 72-76.
5. Jones R,HR,dAC,PD. Changing face of medical curricula. *The Lancet*. 2001; 357.
6. Prideaux D. The Emperor's new cloths: From objectives to outcomes. *Medical Education*. 2000; 34: p. 168-169.
7. Bradt DA, Abraham K, Franks R. A strategic plan for disaster medicine in Australasia. *Emergency Medicine*. 2003; 1: p. 271-282.
8. Posner GJ, Rudnitsky AN. *Course Design: A Guide to Curriculum Development for Teachers* Boston: Pearson; 2006.
9. Epstein RM. Assessment in Medical Education. *The New England Journal of Medicine*. 2007; 356(4): p. 387-396.
10. Wass V, Vleuten Cvd, Shatzer J, Jones R. Assessment of clinical competence. *The Lancet*. 2001; 357: p. 945-949.
11. Frederiksen N. The real test bias: influences of testing on teaching and learning. *American Psychologist*. 1984; 39: p. 193-202.
12. Schuwirth LWT, van der Vleuten CPM. Written assessments. In Dent JA, Harden RM. *A practical guide for medical teachers*. Edinburg: Elsevier; 2006. p. 311-322.
13. Linn RL, Miller MD. *Measurement and assessment in teaching New Jersey*: Pearson Education; 2005.
14. McCoubrie P. Improving the fairness of multiple-choice question. *Medical Teacher*. 2004; 26(8): p. 709-712.
15. van der Vleuten C. The assessment of professional competence: Development, research and practical implications. *Advances in Health Sciences Education*. 1996; 1: p. 41-47.
16. Coderre SP, Harasym P, Mandin H, Fick G. The impact of two multiple-choice question formats on the problem-solving strategies used by novices and experts. *BioMed Central Medical Education*. 2004; 4(23): p. 1-9.
17. Schuwirth LWT, van der Vleuten CPM, Donkers HJLM. A closer look at cueing effects in multiple-choice question. *Medical Education*. 1996; 30: p. 44-49.
18. Fischer MR, Herrmann S, Kopp V. Answering multiple-choice question in high-stakes medical examinations. *Medical Education*. 2005; 39: p. 890-894.
19. Hopkins KD, Stanley JC, Hopkins BR. *Educational and Psychological Measurement*

and Evaluation Massachusetts: Allyn and Bacon; 1990.

20. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*. 2002; 15(3): p. 309-334.
21. Schuwirth LWT, van der Vleuten CPM. Changing education, changing assessment, changing research? *Medical Education*. 2004; 38: p. 805-812.
22. Schuwirth LWT, van der Vleuten CPM. A Plea for New Psychometric models in Educational Assessment. *Medical Teacher*. 2006; 40: p. 296-300.
23. Steinert Y, Mann KV. Faculty development: Principles and practices. *Journal of Veterinary Medical Education*. 2006; 33(3): p. 317-324.
24. Dent JA, Harden RM. *A Practical Guide for Medical Teachers* Elsevier Churchill Livingstone: Elsevier Churchill Livingstone; 2005.