

# Development and Testing of an Assessment Tool for Integrated Learning

Z. Naqvi ( Departments for Educational, The Aga Khan University, Karachi. )

R. Ahmed ( Development Pathology, The Aga Khan University, Karachi. )

B. Jamil ( Departments of Microbiology and Medicine, The Aga Khan University, Karachi. )

## Abstract

**Objective:** Development of integrated questions for undergraduate medical students and assessing students' capability to retain and apply basic concepts in an integrated way.

**Method:** Multiple choice questions, attempting to probe understanding and application of different concepts, affiliated with different subjects and topics, were developed. As a pilot these were administered to the students of third year and final year at the end of their certifying examinations as a mock test. The following year these questions were introduced in the continuous assessments and subsequently the certifying examinations (in third year) as 33% of the whole examination. Difficult indices, item analysis and internal consistency (alpha) were calculated for all the questions.

**Results:** The difficulty indices of integrated questions indicated that students found the integrated questions difficult.

**Conclusion:** There was a significant change of distribution in terms of scoring for the integrated questions. Students' power to integrate showed reliable improvement when tested in the certifying examinations (JPMA 51:384,2001).

## Introduction

A number of curricular changes have been instituted through multiple reviews of the content and instructional strategies at the Aga Khan University Medical College (AKU), Karachi, Pakistan. The principle reason for these changes has been to aspire students away from memorising (rote learning) a wide variety of facts about each discipline. Nevertheless, examinations are still held for each subject separately and mainly consist of single best questions, which do not necessarily assess understanding<sup>1</sup>. This subject orientation of examination questions with intent of testing recall, tends to drive students to work for summative examinations<sup>2</sup>. It has also been reported that students could not retain the memorised information of basic sciences even after fifteen days of the certifying examinations<sup>3</sup>.

In similar curricular environments, Patel and Evans<sup>3</sup> found that medical students were unable to solve patient problems even when specific basic information was provided to them. Similarly residents of a large medical hospital were in 50% of the cases unable to perform critical clinical procedures, yet when tested on this subject by virtue of multiple choice questions, performed quite well (mean 82%<sup>4</sup>). These and many other incidents indicate that although students score more or less brilliantly in examinations, they possess little ability to think and apply the acquired knowledge<sup>5-7</sup> suggesting that we do not necessarily test students on what we expect them to perform or apply in their practice.

Since students' learning activities are directed towards the content of examinations, interventions in assessment content and format can be used as a positive force both for teaching and learning<sup>8</sup>.

However, successful interventions to change learning will more or less depend on the student's perception of the demands of learning context in relation to the intervention.

The development and use of multiple choice (questions) items capable of testing integration of basic and clinical sciences concepts, higher level reasoning skills and problem solving has been reported<sup>9-12</sup>. On the other hand, it seems that in most of such questions the context has been used to add a flavour of

clinical situations to the items. Analysis of these test items depicts that they are independent of the contexts portrayed in the scenarios. An exception is the Case Cluster questions developed by Case and Swanson<sup>11</sup>. The integrated questions developed at the Aga Khan University were tested with the hypothesis that students' scores in integrated questions would correlate positively with their certifying examination scores.

## Materials and Methods

### The pilot

This was conducted in 1997 at the time of the third and final year examinations, held in June and September 1997 respectively. Integrated single-best questions (Appendix) were formulated to probe the thinking process and understanding of integrated concepts. These questions were administered after the respective certifying examinations. Scores obtained in integrated questions were correlated with the scores of certifying examination.

### The study

#### Materials

A generalisable format for construction of questions was developed. The questions were based on case scenarios, representing common problems and required reasoned responses. Five to six multiple choices (single best) item sets followed each case.

#### b) Construction of the questions

Initially, concepts considered important and related to the specific presentations and problems were highlighted. Cases were then developed to provide the relevant data to accentuate the identified objectives. Questions were subsequently formulated to assess students' ability to integrate and think by analysing the provided information. Each case was followed by three to six related questions probing different aspects.

#### c) Format of the questions (Figure)

question stems short.

- Each question had five choices.

## Question Format

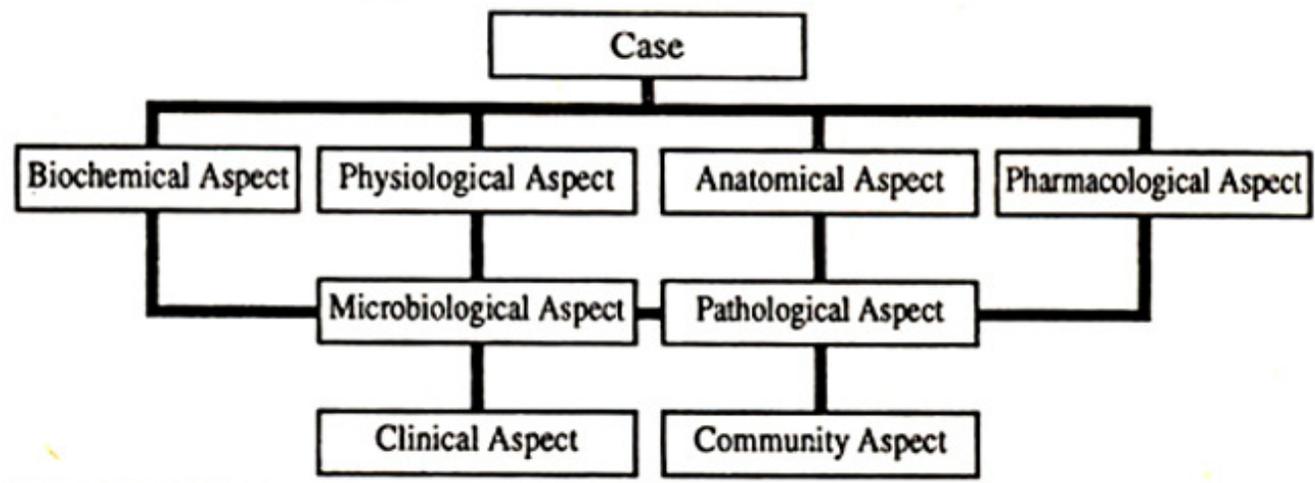


Figure.

Whilst scenarios were used, efforts were made to keep

**question stems short**

Each question had five choices.

**Procedure**

**Pilote**

a) The integrated questions were administered to students immediately after each of their certifying examinations at the end of the third and final years.

b) The study: The questions were introduced in the continuous assessment examinations of third year from the academic year 1998-1999 and subsequently in the final certifying examination as 33% of the whole paper.

**Problems**

Although Pamphlet and Famill13 report that anxiety is not associated with performance in multiple choice examinations, faculty resisted incorporation of these questions into the main body of the paper to avoid extra stress for students.

**Analysis**

I. Scores of the two examinations (certifying and integrated) were correlated (Pearson's t-test).

II. Cross tabulations were made to determine the number of students who had performed differently in the two examinations (Chi-Square test and paired t-test).

III. Reliability of the integrated questions and the certifying examination questions was determined by calculation of internal consistency of the whole papers (Cronbach's Alpha).

**Results**

**Response Rate**

One hundred and twenty-six (complete populations of the two classes), undergraduate medical students of the third year and final (fifth) year, were invited to participate, after informed consent. All sixty-eight students of third year (n=68) responded to integrated questions, but answers of sixty (n = 60) students were complete (88%) and were included in the study. Though all fifty-eight students of the final year participated, 54 (91%) filled the response sheet validly for inclusion in the study. Hence the total number of students included was 113 (90%) of the total population of the two classes.

**Comparison of Cohorts of Scores (pilot)**

A large difference in the means and standard deviations of the two classes in both assessments was depicted. However ranking according to the means remained approximately the same. The overall correlation coefficient (r) for the individual scores of exams and integrated questions within the same classes of third year and final year was 0.23, 0.33 ( $P=>0.01$ ) respectively. For final year t-test for correlated means was significant ( $p=>0.01$ ) whereas ANOVA for final year also depicted significance for final year correlation ( $F=6.377$ ,  $p=>0.01$ ). Distribution of students class ranking in the two examinations, showed a statistical significance for final year. (Chi square  $p=0.049$ ).

**Comparison of Cohorts of Scores (Study)**

When the scores of traditional MCQs and the integrated MCQs were correlated, statistically significant associations ( $p=>0.001$ ) of 0.71 (Pathology), 0.68 (Pharmacology) and 0.71 (Microbiology) were found. The changes in scores were also statistically significant (paired T-test  $p=>0.001$ ). Frequency distribution of students according to grouped scores also showed statistically significant changes (Table 1 and 2)

**Table 1. Distribution of scores in integrated and pathology question.**

Pathology Certifying Examination Scores			
Groups	41-65	66-76	77 and above
Integrated	22*	15*	3*
Questions	5*	14*	20*
Scores	0*	0*	3*

\* Frequency of students within the group.

**Table 2. Distribution of scores in Integrated and Pharmacology Questions.**

Pharmacology Certifying Examination Scores			
Groups	41-65	66-76	77 and above
Integrated	14*	21*	5*
Questions	7*	11*	21*
Scores	0*	0*	5*

\*Frequency of students within the group.

(chi square  $p > 0.001$  with  $df=2$ ). The middle group depicted a most stable performance.

## Discussion

This study indicates that development and testing of questions having a story structure<sup>14</sup> and clinical context stimulates students to think logically in terms of the various aspects of the problem without fragmentation of concepts into different subjects (anatomy, biochemistry and physiology etc.)

The quantitative results and feedback (like Schwartz and Loten's study<sup>16</sup>) show that these innovative MCQs can be used to test higher level cognitive skills viz, integration, interpretation and thinking capabilities about clinical relevance of basic medical sciences. However, we can not claim that these questions also assess the higher cognitive problem-solving skills. To minimise the effect of confounding factors like Case Specificity Phenomenon<sup>6</sup> we tried to select the most common problems and oft-repeated topics in the curriculum, and the items framed probed broad aspects as far as possible. The process of development, testing and post-examination analysis of the new questions further highlighted the importance of multi-disciplinary input in question formation. This input helped to make the questions contextually relevant to the overall curricular objectives. This recent experience also strengthened Schuwirth, Vieuten and Donkers<sup>15</sup> report that the cueing effects of multiple choice questions can decrease with increasing level of expertise in construction and further review of MCQs.

The present analysis showed that the standardised usage of five options in MCQs may reduce the chances of guessing but has little or no effect on the overall behaviour of the questions. Contrarily, Fray and Robert<sup>17</sup> and Haladyna<sup>18</sup> advocate that some valid and important questions demand only two or three options and the major generalisation from these studies is that three options maximise test reliability and efficiency. However, Hodson<sup>19</sup> noted that there was virtually no change in test reliability on reducing the number of options.

The pilot study pointed out that students' performance in terms of individual scores and class means in the integrated paper versus the traditional certifying examinations were considerably lower and unreliable, for the third year students. This indicated that at that time students of the junior classes were less tuned for application of basic concepts in clinical scenarios<sup>20</sup>. This is also similar to Simpson et al<sup>9</sup> and Blumberg<sup>14,21</sup> claim that analysis of the standard item analysis showed that thinking items were significantly more difficult than the pure knowledge items, specially for unprepared students.

The third year students started performing better with stronger correlations to their performances in the traditional examinations when these questions became part of their continuous assessment and final examinations which are summative. These changes could be attributed to the increasing familiarity of the students with question format and perhaps also to a change in the learning approaches of the students, however these factors should be explored further to determine the real effect. The overall reliability of the integrated question also improved significantly when these questions were incorporated into the summative examinations.

Assessment tools can be designed to guide student learning in the right direction. As no assessment method or tool claim to test all the desired objectives, it is recommended that a blend of methods should be incorporated in the examinations to suit all the objectives. The current traditional assessment system does not appear to motivate students towards integrated learning. Ways and means of testing should be formulated and investigated to make the students learn in an integrated manner.

## **Appendix**

### **Question set 1**

On a specially chilly night Uncle Mustafa, 56 years old and his nephews were having their usual fiesta. Uncle Mustafa loved to eat and smoked heavily. Around midnight, Uncle felt a slight heaviness in the chest. He attributed this to overwork and retired to bed.

The next morning, he didn't seem well. He admitted that he had had difficulty in sleeping. His wife insisted that he should immediately go to the doctor, as he had been a known hypertensive for the past several years. While he was lifting himself out of chair, he experienced excruciating pain across his chest and rapidly became breathless. The ambulance was immediately called.

At the hospital, he was restless and confused and his blood pressure was 70/40 mm of Hg with a pulse rate of 100. per minute and his JVP was not raised and he had bilateral basal crepitations in the chest. The doctor immediately shifted him to the CCU where he was placed on the Cardiac Monitor.

**Pick out the single best answer in the following questions.**

**His Blood Pressure indicates:**

- a) Insufficient venous return
- b) Decreased contractility of the Left Ventricle
- c) Reduced peripheral vascular resistance
- d) Reduced diastolic interval
- e) Decreased arterial compliance

**The physiological basis for his sweating is:**

- a) Vasoconstriction
- b) Stimulation of baroreceptors
- c) Reduced tissue perfusion
- d) Increased circulating endogenous catecholamines

e) Increased vagal tone

**The basic reason for his breathlessness is:**

a) Decreased arterial P<sub>O2</sub>

b) Pulmonary venous pressure of 35 mm Hg

c) Increased arterial PCO<sub>2</sub>

d) Low arterial pH

e) Activation of stretch receptors

For the following questions you may pick out as many choices as you find appropriate. On the response sheet Darken "A" of the corresponding number for each of your choice:

**Q. In order to improve tissue perfusion the following**

**drugs would be useful:**

6) Streptokinase

7) Glyceryltrinitrate

8) Normal saline infusion

9) Dobutamine

10) Noradrenaline

11) Frusemide

12) Pethidine

**Question Set 2**

Mubashir, a lean eight years old boy, was taken to a doctor for an increased frequency of urine and bed wetting. The doctor had boiled his urine with a chemical and declared him a diabetic.

Match the numbered colour changes the doctor is expected to observe while performing the test with the lettered inferences.

14. Yellow a. ++

15. Green b. ++

16. Orange c. +

17. Blue d. No Sugar

18. Red e. ++++ and above

**Select the single best answer for the following questions:**

**The doctor should have requested for the following tests to confirm his diagnosis:**

a) Lipid profile

b) Renal profile.

c) Glucose Tolerance Test.\*\* \*

d) Glucose Challenge Test.

e) Serial Fasting Blood Sugar levels.

**20. Mubashir's condition is due to:**

a) Decreased glucose transport across renal tubular cells. b) Increased utilisation of glucose in the body.

c) Decreased absorption of glucose from the intestinal mucosa.

d) Decreased metabolism of glucose\* \*

e) Increased intracellular glucose levels.

**21. Mubashir had a past history of diarrhoea as a child. This could be a significant etiological factor for development of Diabetes Mellitus (DM) because:**

a) Electrolyte imbalance is a precipitating factor for DM

b) Anaemia causes DM.

c) Bacterial diarrhoea may induce DM

d) Viral agents responsible for diarrhoea induce DM\*\*\*

e) DM causes inflammation of small intestines.

**22. His diabetes is most likely due to:**

a) Insulinitis

- b) Hereditary cause
- c) Pancreatitis
- d) Hypopituitarism.
- e) Malnutrition

**Treatment for Mubashir's Diabetes Mellitus is:**

Mubashir should be monitored to prevent the

- a) Oral Hypoglycemic Agents.
- b) Strict diet control.
- c) Intravenous insulin
- d) Subcutaneous insulin\* \* \*
- e) Intramuscular insulin.

**development of:**

- a) Hyperosmolar non-ketotic diabetic coma\* \*\*
- b) Ketonuria
- c) Decreased gluconeogenesis
- d) Lactic acidosis
- e) Increased utilisation of glucose

**References**

1. Marton F. Describing and improving learning. In: R.R. Schmeck (ed.) Perspectives on individual differences, learning strategies and learning styles. New York: Plenum Press, 1988. pp. 64-89.
2. Siddiqui, AA, Ahmed, HR, Talati J, et al. Problem based learning in basic health sciences : An AKU Experience. Paper presented at 6th Biannual Conference. Pakistan Physiological Society. Pakistan, 1997.
3. Patel VL, Evans DA, Kaufman DR. Reasoning strategies and the use of biomedical knowledge by medical students. Med. Educ., 1990; 24: 129-36.
4. Gonella JS, Goran MJ, Williamson JW, et al. Evaluation of patient cares, an approach. J. Am. Med. Assoc., 1970; 214:2040-3.
5. Van Berkel HJM, Nuy HJP, Geerligs, T. The influence of progress test and block tests on study behaviour. Instructional Sci., 1995 22: 317-33.
6. Norman GR, Tugwel P, Feightner FW, et al. Knowledge and clinical problem solving. Med. Educ., 1985; 19: 344-56.
7. Swanson DB, Case SM. Assessment in basic science instruction: Directions for practice and research. Adv Health. Sci. Res., 1997; 2: 71-84.
8. Ramsden P. Approaches To Learning. In: Learning to teach in higher education. London: Routledge, 1992, pp. 38-61.
9. Simpson DE. Problem solving questions for multiple choice tests, Paper presented at American Educational Research Association; Chicago, 1995
10. Szeberenyi J, Boglari G, Komaromy I, et al. Problem oriented teaching of molecular biology. Med. Educ., 1996; 30: 232-4.
11. Case SM, Swanson DB. Problem based learning and use of case clusters. In: Constructing written tests questions for basic and clinical sciences. Philadelphia. National board of Medical Examiners 1996. p. 43.
12. Martenson D, Hakan A, Kerstin G. An integrated final examination in preclinical subjects for medical students: 10 Years experience. Test. Reach. Med., 1999; 11, 26-33.
13. Pamphlet R, Farnill, D. Effect of anxiety on performance in multiple choice questions. Med. Educ., 1995; 29: 297-03.

14. Bowman MG. The effects of story structure sequencing upon reading comprehension. Paper presented at the American Educational Research Association. Los Angeles, 1981.
15. Schuwirth LWF, Vleuten VD, Donkers, HJLM. A closer look at cueing effects in multiple choice question. *Med. Educ.*, 1996; 30: 44-9.
16. Schwartz PL, Loten EG, Bowman L. Brief problem solving questions in medical school examinations: Is It necessary for students to explain their answers. *Med. Educ.*, 1999; 33:823-7
17. Frary RB. more multiple-choice item writing do's and don'ts. Office of Educational Research and Improvement (ed), Washington, DC, 1995.
18. Haladyna TM, Downing SMA. Quantitative review of research on multiple-choice item writing. Paper presented at the 69th Annual Meeting of the American Educational Research Association, Chicago, III, 1985.
19. Hodson D. Some effects of changes in question structure and sequence on performance in a multiple choice chemistry test, research in science and technological education. 1985; 2: 77-85.
20. Kaufman DM, Mann VK. Basics sciences in problem-based learning and conventional curricula: Students' attitudes. *Med. Educ.*, 1997; 31: 177-80.
21. Blwnberg P. Should Taxonomic levels be considered in developing examinations. *Educ. Psychology. Measurement.*, 1982; 42: 7.