

The Forgotten Dimension: The Information Content of Objective Questions

Alan Dugdale

Griffith University, Gold Coast, Australia

Email: dugdalea@gmail.com

Received 4 August 2015; accepted 13 September 2015; published 16 September 2015

Copyright © 2015 by author and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Objective examination questions are used for assessment in many disciplines and levels. They allow extensive coverage, reliability, re-usability and administrative ease, but the amount of information they gather about student knowledge is not widely known. In this paper, I show that small changes in question format make large differences to the information gained, and therefore to the efficiency of the assessment process. The basic objective question (BOQ) can be extended beyond the true/false response to test probabilities in complex situations. When objective questions are used to assess students' abilities, the information gained should be sufficient in quantity and coverage and at the appropriate DOK (Depth of Knowledge) to give a robust and valid assessment. Objective questions take time and expertise to set and validate, so we should seek methods that meet these criteria with the least administrative load. The basic objective question is a statement followed by two or more responses; one response must be marked as correct. The Type 1 MCQ (Multiple Choice Question), multiple true/false question (MTFQ), extended matched question (EMQ) and others are all variants of the BOQ, but they differ widely in the amount of information they gain. For a given input of statements and responses, the Type 1 MCQ delivers the least information about the student's knowledge, the MTFQ and the BOQ deliver almost four times as much and the EMQ about six times as much. The expected level of student knowledge also affects the information gained. If questions are set so that most (say 90%) of students get the correct answers, then the information gained, and the discriminatory power of the test is much less than when the test is set for lower expected levels. The format used for objective questions and the expected level of correct responses govern their efficiency in collecting pass/fail information. The quality of that information such as depth of knowledge (DOK) depends on the wording and situations modeled for the questions and this is independent of their format.

Keywords

Objective Questions, Information Level, Depth of Knowledge

1. Introduction

Objective questions are a major form of testing in education, including medical education. There is a large and sometimes contradictory literature on the validity, reliability, discriminatory power and the depth of knowledge tested using different types of objective questions (Palmer & Devitt, 2007; Rothhoff, Baehring et al., 2006; Schuwirth & van der Vleutin, 2004). There are data on the number of questions needed to gain a reliable assessment (Pamphlett, 2005), but little analysis of the information given by each question (Dugdale, 2013). Objective questions are difficult and time consuming to set, validate and standardize. Provided we do not compromise other features, we should prefer the question that tells us most about the student's knowledge for a given input of examiner time. In objective questions, the student must select one or more listed answers. Intuitively, the most informative questions are those where the student has to give many answers and has the widest range of choices for each answer. We can now measure the number of "bits" of information in each type of question using information theory and then compare the cost/benefit levels of the different types.

Medical teachers have been imaginative and prolific in designing new formats of objective questions to meet different needs and to test a wide range of clinical situations and knowledge. These techniques can be used in other areas. The formats are all variations of the basic objective question (BOQ). We can compare their properties, and particularly their information contents to select the most efficient and valid formats.

In this paper, I shall demonstrate the formats of various formats of objective questions, calculate their information content and administrative efficiency and suggest appropriate formats for specific purposes.

2. The Structure of Objective Questions

2.1. Basic Objective Question (BOQ)

All objective questions have the same basic structure. There is a STATEMENT followed by two or more RESPONSES. The candidate must choose the most appropriate response. There is usually a CONDITION which may be implicit. Most conditions limit the student's responses and reduce the information gained. The examples below illustrate some alternatives

Examples

- a An apple is a fruit True False
Condition: If no response marked this indicates Don't Know
- b An apple is a fruit True False Don't Know
Condition: One response must be marked
- c An apple usually weighs <20 g 20 - 500 g >500 g DK
Condition: one response must be marked

Each of these is a complete and valid objective question. In normal usage, the Condition is implied but is stated here for clarity and completeness. The student must answer each statement, which gives a large output of information for a small input of effort by examiners. All complex types of objective questions use this BOQ.

2.2. Multiple Statement Questions

Type 1 MCQ

The Type 1 MCQ is a Basic Objective Question in a different format. If expressed on a single line it is a BOQ or one line of a multiple true/false question (MTFQ). When formatted differently (see Example below) it becomes a Type 1 MCQ.

Example

- A An apple usually weighs <20 g 20 - 500 g >500 g DK
Condition: one response must be marked
- B An apple usually weighs
 <20 g
 20 - 500 g
 >500 g
 DK
Condition: one response must be marked

In this format, most of the possible responses are not shown, usually only a box for True is present for each

statement. Because of the condition imposed, the statements are not independent, but relate to one another. The student can give only a single answer, this reduces the information we get about the student's knowledge.

2.3. Multiple True/False Questions (MTFQ)

This allows the examiner to ask several questions on the same topic. The introductory statement is divided into a STEM followed by two or more STATEMENTS all relating to that stem. Each statement is followed by two or more RESPONSES, either expressed or implicit: one to be marked. It is essential that each statement is independent of the others, so the response to one statement does not influence responses to other statements.

A multiple statement question (also known as a MTFQ) is a series of BOQs on the same general topic. The usual condition is that the student must select a single response for each statement.

2.4. Other Types of Objective Questions

All objective questions use the basic form of statement followed by a selection of responses, but these can be assembled in different ways and conditions may be used to limit the choice of responses. There are advantages and disadvantages from each format. Several new formats have been developed. They include:

2.5. Extended Matched Question (EMQ)

This is a Type 1 MCQ where the stem and statement are sentences or clinical descriptions. In each EMQ there is a column with six or more statements and a second parallel column with ten or more possible responses. The student must select an answer for each statement from the list of responses. Responses may be appropriate for one statement, several statements or to no statement. The student must therefore select one response from a list of 10 or more response options instead of the 4 - 6 possible responses in the standard form Type 1 MCQ. Each extended matched question is therefore equivalent to several (depending on the number of independent statements) Type 1 MCQs and the increased number of options for each statement gives the student more choices than a single Type 1 MCQs. Although the extended matched question is basically a variant of the Type 1 MCQ it has the potential to give much more information than a series of single Type 1 MCQs for the same time and effort by examiners. Although it is not a necessary feature of the EMQ, its use of longer phrases and even sentences in both the statements and responses encourages the testing of deeper knowledge.

Example

A 7-year-old girl developed a fever and sore throat 2 days ago. Now has very sore throat with pain on swallowing, fever and general malaise. Had similar attacks 6 months ago. Examination showed Temp 38.6 degrees, white coated tongue, pharynx mildly red, tonsils very red with yellow flecks, a large tender lymph node at angle of jaw. Other systems clear

The following conditions can cause sore throat			The following agents may lead to the diagnosis you have chosen		
A	Tonsillitis	[]	1	Measles virus	[]
B	Pharyngitis	[]	2	Herpes simplex	[]
C	Laryngitis	[]	3	EB virus	[]
D	Epiglottitis	[]	4	Streptococcal infection	[]
E	Stomatitis	[]	5	Haemophilus influenza	[]
F	Generalised upper respiratory infection	[]	6	Influenza virus	[]
			7	Staphylococcus aureus	[]
G	Measles	[]	8	Adenovirus	[]
H	allergic rhinitis	[]	9	None of the above	[]
X	Don't know	[]	10	Don't know	[]

Choose the most likely principal diagnosis and causal agent for the clinical condition,

Note: There were several clinical scenarios requiring responses to the same lists of possible answers Script Concordance Test (SCT)

This is an example of a series of BOQs with a complex clinical stem and statement (van Bruggen, van Wou-denbergh et al., 2012).

Example

 A 15 year old girl admits to taking a toxic dose of acetaminophen 30 hours before the consultation following a break up with her boyfriend. She has no important medical history and her physical examination shows only mild tenderness in the upper abdomen

If you were Thinking of	And then you find	This investigation becomes
Checking levels of salicylates and other drugs	That the patient denies taking any other drug	-2, -1, 0, +1, +2

 -2 = totally contraindicated -1 = fairly useless or possibly harmful 0 = neither useful or harmful
 +1 = useful +2 = indicated or absolutely necessary

 Each statement is independent of the others, which makes it equivalent to a MTFQ, but the responses are not True/False but grades of likelihood.

Although the structure is basic and simple, the use of probabilities rather than simple True/False increases the number of choices, and hence the amount of information gained. Probabilities also increase the depth of knowledge needed for correct answers, and so raises the quality as well of the quantity of information gained.

2.6. Matched Question List

This format is similar to Extended Matched Questions (EMQs) but the statements and responses are usually single words

Each of the formats is a variation of the BOQ and is designed to test various aspects of the student's knowledge and his/her ability to use that knowledge, but they vary widely in the amount of information they yield for a given input of question material. To avoid random errors in assessment we need to test several aspects of knowledge and how it is used; from each aspect we must get enough information for a robust and reliable assessment. Setting objective questions is difficult, so we should, where possible, use the format that yields the most information for each hour of setting time. Although it is not the main subject of this paper, we should also select questions which tests greater depth of knowledge. I shall demonstrate the wide variability in the efficiency of objective questions.

3. Information Content of Objective Questions

3.1. Information Theory—A Very Brief Overview

Advanced information theory is complex and highly mathematical, but the basic ideas are simple and need only high school algebra. The basis was published in a trade journal by Shannon, 1948 and a brief simple description is given by Moulton (2012). Many of the results derived from information theory are numerical results for ideas that are intuitively obvious. Information is measured in bits, one bit is the amount of information gained by choosing one of two equally likely alternatives such as Yes/No or True/False. If there are three equally likely choices, such as True/False/Don't Know then we gain more information. The basic equation is 2

$$\text{Bits of information} = \sum -p \cdot \log_2(1/p)$$

where p = probability of each event and all the p values add up to 1, and $\log_2(1/p)$ is the logarithm to the base 2 of $1/p$.

Most assessments are designed so that the average student will score about 70% correct answers. This built in

bias lessens the amount of information we gain. An extreme example is to start a questionnaire by asking a sample of people “Are you male or female”. If the subjects were random users of a shopping mall, then the response would help in the analysis of later answers so provides useful information. However if we took that same questionnaire to a nunnery, it would be no surprise that 100% answered “Female”. We knew this in advance, so gained no new information from the question. The formula above gives us a numerical measure of the information gained, allowing for prior knowledge.

I shall use this basic formula to calculate the information in the different types of objective question used in assessment of medical education. To make the different types comparable, I shall give each type a choice of four items, but this is not necessary in practice. The data used in the example are designed to show the process and may not be important or useful items of student knowledge.

3.2. Basic Objective Question

Two main factors govern the amount of information gained from a BOQ, 1) the expected percentage of students giving each response and 2) the number of alternative responses.

The simplest condition for the BOQ is where there are three choices True/False/Don’t Know (the Don’t Know option may not be given, but the students may Abstain). If each of these responses is equally likely, then the answer contains 1.58 bits of information. However, if we expect from design or past usage that 70% of the students will give the correct answer and that 5% will indicate Don’t Know, then the information content is 1.08 bits. With 90% Correct and 5% Don’t Know we gain 0.57 bits. A higher expectation drastically lowers the information content.

When we increase the number of possible responses, we increase the student choice and the yield of information. In the Script Concordance Question above, which is a BOQ with six choices plus the implicit Don’t Know). If we expect 70% correct answers, 5% Don’t Know and remainder equally distributed, then the information gained is 1.58 bits. The extra two options increases the information from 1.08 bits to 1.58 bits.

3.3. Multiple True/False Questions

The MTFQ is a compound form consisting of two or more BOQs using the same stem. Provided the statements are independent one from the other, then this format is a series of BOQs or Type 1 MCQs with a single stem and multiple statements. The information gained from an MTFQ with five statements is the same as five separate BOQs, or five separate Type 1 MCQs. With 70% expected correct and 5% expected Don’t Know, each statement yields 1.08 bits, so the total information gain for the MTFQ is 5.40 bits. When the expected scoring is 90% correct and 5% Don’t Know, the total gain is 2.85 bits.

Type 1 MCQ

In a standard MCQ with a stem and five alternatives—choose the correct alternative. The student makes one choice. There is the hidden alternative—Don’t Know—which the student selects by not marking any response.

This form of objective question is a multiple true/false question with the condition—ONLY ONE OF THE FOLLOWING ANSWERS IS CORRECT. This condition reduces the number of student responses from five to one and so reduces the information gained. For a five statement question, with 70% expected correct and 5% expected Don’t Know, then the information yield is 1.58 bits. If 90% are expected correct and 5% Don’t Know, the yield is 0.67 bits.

3.4. Extended Matching Questions (EMQs)

The EMQ is an extension of the Type 1 MCQ. If there are five independent statements (words or clinical scenarios) in the first column and twelve possible responses in the second column, 70% of responses expected correct and 5% expected Don’t Know, then the yield for each statement is 1.94 bits, and for the whole question is 9.70 bits.

3.5. Script Concordance Test

This is a series of BOQs where the stem and the statement are clinical events or scenarios, and for each statement there are several responses. The information yield is the same as a BOQ. For a Script Concordance test with five statements and six Responses for each statement (5 declared plus Don’t Know), each statement yields

1.66 bits, giving a total of 8.30 bits.

3.6. Matched Question Lists

This is a variant of the EMQ and the information yield is the same.

4. A Measure of Educational Efficiency of Objective Questions

The aim of an assessment is to collect enough relevant information about the student's knowledge and its applications to form a valid opinion on his/her future pathway. This information will always be a sample of the total range of knowledge. A larger and diverse sample gives a more accurate picture less subject to random errors. In free-form questions where the student produces essay type answers, the information content cannot be measured numerically. With objective questions and a controlled format, we can calculate the amount of information gained. If we need 100 Type 1 MCQs to get a workable set of data we can show the equivalent numbers of other objective questions. Using the data above for standard sizes of each type, 100 Type 1 MCQs yields the same information as 100 BOQs, 29 MTFQs, 16 extended MCQs, 19 Script Concordance Questions and 16 Multiple Matching MCQs. We must consider other factors, such as the suitability of each format for its specific task and the bias that occurs with MTFQs (Kelly & Dennick, 2009), but the information gained is worth considering.

5. A Measure of the Organisational Efficiency of Objective Questions

Setting and validating objective questions uses valuable staff time and resources. The most demanding part of setting any type of an objective question is selecting alternative answers that are definitely true but not too obvious or definitely false but still plausible. The ratio

$$\text{Efficiency} = \frac{\text{Bits of information gained}}{\text{Number of alternative answers}} * 100$$

gives an rough indication of the information gained for a given input of time and resources. A higher efficiency score shows a more productive use of resources (Table 1).

6. Discussion

The aim of any student assessment is to collect information that will make a reliable decision on his/her competence and so determine his/her future pathway. Competence has many facets, so there must be varied assessments to test these. Objective questions have been used to test factual knowledge for many years. More recently, they have been used to test clinical reasoning (van Bruggen et al., 2012). This is a multi-step process using relevant facts in a logical sequence to reach a reasoned outcome. The present standard formats of objective questions will not do this, but they can be adapted to test clinical logic one step at a time (Carrière et al., 2009). The existing criteria of validity, reliability, discriminatory power and coverage still apply, but it will take both time and resources to develop and test these new applications. We must therefore consider the additional criterion of efficiency. The more information about the student's knowledge (in its widest sense) that each new question gives, the fewer questions will be needed and the quicker the process of development and testing.

Table 1. The organizational efficiency of various formats of objective questions.

TYPE OF QUESTION	NO OF ALTERNATIVE RESPONSES (INCLUDING DON'T KNOW)	RELATIVE EFFICIENCY IN PREPARATION
Basic objective	3	158
Multiple true/false	3	158
Type 1 MCQ	6	32
Extended multiple choice	13	97
Script concordance	6	166
Matching lists	13	97

The Type 1 MCQ has been the mainstay of objective testing for more than 50 years. Its validity, reliability and discriminatory power are known and banks of questions are available. However, we get a single student response from each question; compared with other formats this is very inefficient. For the same effort in setting, the same question in MTFQ format and without the condition, gives nearly four times the information. Marks gained by the standard MTFQ seem to introduce gender bias into the scoring (Kelly & Dennick, 2009), so this format is now not used. However, the Type 1 MCQ and the MTFQ are both examples of BOQs in different formats, which makes the gender difference hard to explain. Some of the newer types of objective question, such as the Script Concordance Test have the same structure as the MTFQ but have clinical scenarios instead of simple statements, and the responses are clinical probabilities rather than True/False/Don't Know. We await analyses to see whether this new version of the MTFQ has the same biases as the original MTFQ.

The loss of information and efficiency when high scores are expected raises interesting issues. We expect most (say 90% - 95%) of medical students to pass their assessments. If we set the pass mark at 70% then this is must be about the -2SD level of the distribution. The expected mean score must be 80% - 90% assuming a normal distribution. Information theory shows that at this expected level the information content of all objective questions is low.

Information content is only one of the criteria to judge the suitability of objective questions. If other qualities, such reliability and coverage are the same for all formats, then more information makes the assessment less liable to random errors. It seems likely that scores from questions with high information content will accord with the overall outcome, although the argument is circular as they contribute more to that outcome. Administrative efficiency is another factor. Constructing objective questions is expensive of skilled resources, getting the same outcome from fewer questions is highly desirable. I suggest that when decisions are made on new and new types of objective questions, the information content be considered along with validity, reliability and coverage.

The other important factor is the depth of knowledge required to answer the questions. Questions on superficial knowledge (DOK1) are simple to set, test and score. The level of the student's DOK1 knowledge may parallel his/her deeper knowledge, but if examinations concentrate on this level, professional learning will inevitably suffer. The tests which demand deeper knowledge, such as the Script Concordance Test are still BOQs. The difference does not lie in the format, but the way that it is used. Like most real life situations, the problems involved in clinical medicine are complex. Initial statements of the background must be reasonably comprehensive and the selection of alternative answers wide. More important is the use of probabilities in answering objective questions. Few situations are 100% true or 100% false and most of these fall into the recall of facts (DOK1) category. The major advance in the Script Concordance Test is not the format but allowing answers to be given as probabilities rather than True/False. This immediately widens the range and complexity of questions that can be asked and, at the same time, increases the depth of knowledge required from DOK1 or DOK2 to DOK3 and DOK4.

7. Conclusion

Objective questions will remain a major assessment tool in the foreseeable future. We can select the most useful and efficient type for the particular task. However, most of the common formats, such as Type 1 MCQ, MTFQ and EMQ have a fundamental limitation. The answer must be True or False. This restricts testable knowledge to established facts and excludes current less-than-certain knowledge or decisions recalling established facts involve only superficial knowledge (DOK1). In the real world, answers to questions are seldom black or white, but are shades of grey depending not only on the facts, but also on the circumstances. For example, in clinical medicine, if I am asked whether a child with a mild illness and fever should have paracetamol, I know that both the risks and benefits are small, my level of judgment might be superficial. However, if I am deciding whether to give a highly toxic, but potentially life saving, anti-cancer drug, I would give much more time, thought and study to the decision. The risks and benefits are not trivial and my decision would be based on DOK4 knowledge. If assessment is to be more than an academic exercise, then we must go beyond the standard objective questions. We should use questions that demand more than true/false answers and we should specify the gains from correct answers and the losses from error. The script concordance question or the extended BOQ allows for uncertainty; the potential gains and losses should be stated. These are poorly charted waters (Confidence Marking, 2007); little has been published on the setting, marking and reliability of such formats and their effect on the learning process, but it is the next logical step in objective assessment.

References

- (2007). *Confidence Marking and Degrees of Certitude*. <https://certainty.wordpress.com/category/tariff>
- Carrière, B., Gagnon, R., Charlin, B., Downing, S., & Bordage, G. (2009). Assessing Clinical Reasoning in Pediatric Emergency Medicine: Validity Evidence for a Script Concordance Test. *Annals of Emergency Medicine*, 53, 647-652. <http://dx.doi.org/10.1016/j.annemergmed.2008.07.024>
- Dugdale, A. E. (2013). Towards More Efficient Assessments: Increasing Information from Objective Questions. *Creative Education*, 4, 39-41.
- Kelly, S., & Dennick, R. (2009). Evidence of Gender Bias in True-False-Abstain Medical Examinations. *BMC Medical Education*, 9, 32. <http://dx.doi.org/10.1186/1472-6920-9-32>
- Moulton, R. (2012). *A Short, Simple Introduction to Information Theory*. <http://knol.google.com/k/ryan-moulton/a-short-simple-introduction-to/3kbzhxxyg4467/7#>
- Palmer, E. J., & Devitt, P. G. (2007). Assessment of Higher Order Cognitive Skills in Undergraduate Education: Modified Essay or Multiple Choice Questions? *BMC Medical Education*, 7, 49. <http://dx.doi.org/10.1186/1472-6920-7-49>
- Pamphlett, R. (2005). It Takes Only 100 True-False Items to Test Medical Students: True or False? *Med Teach*, 25, 468-470.
- Rotthoff, T., Baehring, T., Dicken, H.-D., Fahren, U., Richter, B., Fischer, M. R., & Scherbaum, W. A. (2006). Comparison between Long-Menu and Open-Ended Questions in Computerized Medical Assessments. A Randomized Controlled Trial. *BMC Medical Education*, 6, 50. <http://dx.doi.org/10.1186/1472-6920-6-50>
- Schuwirth, L. W. T., & van der Vleutin, C. P. M. (2004). Different Written Assessment Methods: What Can Be Said about Their Strengths and Weaknesses. *Medical Education*, 38, 974-979.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379-423. <http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- van Bruggen, L., Manrique-van Woudenberg, M., Spierenburg, E., & Vos, J. (2012). Preferred Question Types for Computer Based Assessment of Clinical Reasoning: A Literature Survey. *Perspectives on Medical Education*, 1, 162-171. <http://dx.doi.org/10.1007/s40037-012-0024-1>