

Multiple Choice Questions (MCQs) vs. Short Answer Questions (SAQs) for Inferential Comprehension

Bartholomeus Budiyo

budiyo@ukwms.ac.id

Universitas Katolik Widya Mandala Surabaya

Abstract

Among a variety of item types or test formats, multiple-choice tests are much more associated with many advantages. However, short-answer questions as a type of open-ended questions are proposed to be more appropriate for measuring inferential skills. This study investigated this particular issue to provide evidence whether it was the case in college reading for inferences of implied main ideas. The findings indicate that the performance in MCQs was significantly much better than that in SAQs. A recommendation would be a blend of MCQs and SAQs in measurement of inferential comprehension of main ideas with greater weighting on SAQs.

Keywords: multiple-choice, short-answer, inference, significant

I. Introduction

There is a theoretical framework in assessment as regards a variety of item types or test formats that can be administered. They are multiple-choice questions (MCQs), true-false (TF), fill-in-the-blank, short-answer questions (SAQs), problem-solving exercises, and essay questions. Outside to the MC and TF questions are those formats that may be described as constructed-response (CR) questions or the open-ended questions that require the development of a rubric. A relevant question about these two major formats in reference to item type selection to measure instructional objectives is whether they provide different information of assessment.

A closed-ended question is different from an open-ended question in terms of the type of response that can be given. The first renders a limited number of options to choose whereas the second allows the students to write their own answers to make up a wide variety of answers. Another difference is that it is much faster to rate the answers to closed-ended questions than the answers to the open-ended questions because the questions allow a variety of responses. From the

point of view of test development, the use of the open-ended type is likely to increase the error in scoring or to lower the reliability level. It is expensive in that it requires multiple raters.

The MC formats, as closed-ended questions, are also distinguished from and the CR formats in terms of expectation. Martinez (1999) and Traub & McRury (1990), as quoted in Simkin and Kuechler (2005) argue that test takers who anticipate the MC formats favor memorization of details whereas those who anticipate the CR formats struggle for concepts. Dufresne et al. (2002) in Simkin and Kuechler (2005) respond that the answers to the MC questions –more often than not, a false indicator of deep conceptual understanding.

II. Review

A. Multiple-Choice

Many advantages are associated with multiple-choice tests and make them widespread as an objective means of testing. They can be used for diagnostic or formative tests, cover a broad range of knowledge. They can be scored easily and objectively as a classroom achievement test or proficiency test.

In open-ended and performance items, the students must construct their own answers and these answers provide the assessor with the opportunity to analyze the students' higher thinking skills. On the other hand, the multiple-choice items are considered to be unable to measure higher order thinking skills. These items, however, can also yield some information of the students' misconceptions about the content subject. MCQs direct the students to one-right answer thinking. Mullen & Schultz (2012) argue that MCQs "usually lead to a surface approach to learning, and do not allow students to demonstrate the depth of their knowledge or understanding". In addition, Tychonievich (2012) notes that in MCQs "Any skilled test-taker quickly learns that most multiple choice tests contain in their wording most of what you need to know to get a good score." In this case, test-wise students are likely to favor MCQs. Furthermore, Funk & Dickson (2011) make their comment that "performance on multiple-choice exams may provide inaccurate information to instructors".

There is a preference for the MC format as an assessment tool across

disciplines or course subjects such as those in the national school-leaving exam where the MC format is employed for the compositional section. It dominates the entirety of the ETS tests such as the IELTS and the TOEFL. This preference relates to the formulation of guidelines in the construction of MC questions, e.g., Haladyna et al. 2002) for example, mention guidelines about multiple-choice item writing. Furthermore, Simkin and Kuechler (2005) mention the following advantages of MCQs:

Facilitates timely feedback for test takers in classes, and immediate feedback in web-based systems; Enables instructors to ask a large number of questions on a wider range of subject materials; Helps students avoid losing points for poor spelling, grammar, or poor writing ability; Easier preparation by test takers; Does not require deep understanding of tested material (student advantage)

They classify the advantages from the points of view of the instructor (ease, accuracy, coverage), of the test takers (perceptions of possible guessing and the absence of bias).

Different MCQs test different levels of understanding, e.g., a question of definitions tests superficial knowledge or rote learning. For this purpose, Bloom's taxonomy can be used to perform a detailed knowledge level analysis of both constructed response and MCQs questions. It is theoretically possible to construct MCQs that measure many of the same cognitive abilities as CR items.

The MCQs can, however, be unfavorable. The development of MCQs can be a burden due to the difficulty in the development and the time allocation. It is also criticized for giving no opportunity to the test takers to express their understanding in their own words. Another concern is the possibility that the MC test results are influenced by the test takers' test wiseness. The MC format may also be criticized for discouraging critical thinking, failing to attract students to critical areas of the course subject (Simkin and Kuechler, 2005).

Test takers may readily realize that the wording in MCQs leads them to get a score. They may recognize some options are beyond their memorization but are available for them to choose. They may also wisely discard bad options. In this case, the MCQs facilitate learning, i.e., providing concepts that they probably have not internalized and, therefore, they have to learn as new concepts. It does not happen in open-ended question when "concepts are not yet acquired and the results are –vague or confused answers" (Tychonievich, 2012)

Gay (1980) found that students who were trained in MCQs performed the same as those trained in SA testing in a final exam and those trained in SA had significantly better performance in SA. Funk & Dickson (2011) also confirmed that performance on MC items was significantly higher than performance on the same items in short answer test. They reflected on their finding that not all SAQs are harder than multiple-choice questions and not all MCQs are easier than SAQs.

An issue in the MC questions is one about the number of options. First, there has been a recommendation for developing three options. Straton and Catts (1980) determined that the level of reliability of results of the three-option was higher than those of the four-option test but the difference was not significant. The item difficulty rose from the two-option to the three-option and to the four-option items. They also revealed that the time allocation increased with the number of options. They concluded that the three-option test was the optimal for classroom purposes.

Aamodt and McShane (1992) investigated the interaction of the number of options, test scores and the time allocation. They found that the three-option was a bit easier than the four-option test and significantly took less time for completion. They concluded that the number of the three-options could be increased instead of the small number of four-option items. Landrum, Cashin, and Theis (1993) compared three-choice and four-choice test forms that were more or less of the same level of difficulty. The mean score as well as the paired t-test result indicated that there was some improvement with the three-option test. Baghaei & Amrahi (2011) recommended the presence of three options for multiple-choice vocabulary test because the three-option, four-option, and five-option vocabulary tests resulted in almost identical properties. With the three-option items, one can save time, energy, and money. This recommendation may also reduce the risk of formulating implausible distracters and providing clues to test-wise students. Time saving would make it possible to construct more items.

Furthermore, the number of options relates to the quality of the options. Crehan, Haladyna, and Brewer (1993) compared the reliability of three-option and four-option test and confirmed that the three-option test was easier than, but less

reliable than, the four-option test. There was no important difference in the item discrimination indices between these two forms. Delgado and Prieto (1998) also compared two test versions (three-option and four-option tests) administered in two successive years. The three-option version was constructed by deleting the least frequently selected options. They found no significant change in item difficulty, item discrimination, and test reliability. In addition, Rodriguez (2005) analyzed 27 empirical studies and come to the following findings: (1) item difficulty decreased from four-option to three-option and to two-option items, (2) there was a slight increase in item discrimination and reliability in reducing the options from four to three. For this reason, Rodriguez concluded that three-option items are optimal.

Bridgeman & Morgan (1996) clarified that success in college courses belong to both patterns: first, strength in essay format and weakness in MCQs, and second, weakness in essay format and strength in MCQs. Parmenter (2009) argues that for assessing higher thinking, open-ended questions are more appropriate and the difficulty in administration and scoring deserve this high appropriateness. This is in contrast to MCQs that suggest ease in scoring and are usually recall-oriented. However, he also defends that MCQs may be used to assess higher order thinking skills.

The review so far has recommended the presence of three options as the optimal strategy in MCQs and the comparability of MCQs and SAQs. It has not come to particular reading skills such recognition and inference as provided Barrett taxonomy. The present study is intended to explore whether the recommendation of three options works well in comparison with the SAQs in inferential reading comparison.

B. Short-answer questions

Short-answer questions belong to the type of constructed-response questions or open ended-questions that require tests takers to give answers. In practice, they may take the formats of sentence completion, supply of missing words. The answers may need one word or a few lines or bulleted. They are commonly used for low cognitive level measurement (Chan, 2009). It is easy to develop short- answer questions and to grade the answers. Some of the advantages of SAQs are as follows:

Allows students to demonstrate complex, in-depth understanding;
Less likelihood of guessing correct answer; Motivates students to learn in a way that stresses the organization of information, principles, and application

Source: Digital Chalkboard (2015)

Weimer (2015) mentions the advantages and disadvantages of MCQs and SAQs . The answers to MC questions are easy and quick scoring, for a test of a wide range of content areas or skills. They take much time to construct, let the students guess especially when they have insufficient knowledge of the item. On the contrary, it is easy to develop short-answer questions to grade the answers. It may encourage memorization of superficial contents. Short-answer questions require students to give answers and, thereby, avoid cueing. Multiple-choice questions may make students feel less anxious and it is easy to score the answers. Kim (n.d) has the opinion that short-answer questions are easy to construct, reduce guessing correctly, but are unsuitable for measuring complex learning outcomes. Correct answers may vary depending on how clearly the questions are written.

There are other studies on multiple-choice and short answer tests. Ramraje and Sable (2011) administered classroom MC and SA tests on two separate batches with a delay of three weeks. The MC group came up with a high mean delayed retention score of 10.97 and the short answer test yielded 8.42. This finding leads to a conclusion that MCQs and SAQs provided a good delayed retention. Funk & Dickson (2011) explored the retention the same questions in the MC and SA formats. They administered MC and SA tests to 50 students: 25 students to complete SA then 50 MC test and 25 students to do MC tests then 10 SA test. The performance on MCQs was significantly higher ($p < 0.001$) than the performance on the SA test. Additionally, the students who were unable to answer several SAQs were able to answer significantly more of the same items when presented in a multiple-choice format. Tychonievich (2012) provides a relevant comment about the intellectual impact of the MC format in that a skilled test-taker will easily recognize the correct answer among the options already provided.

III. Methods

One-sample posttest design was adopted to investigate the difference of the performance on the MC test and the SA test of inferring main ideas. The sample consisted of 20 university students who attended the low-advanced reading class at the even semester of the academic year of 2016-2017. They took the multiple-

choice test one week before the mid-test and the short-answer test one week before the final test.

The instrument was a pair of tests. The first test was the multiple-choice test that consisted of thirty short paragraphs and thirty questions, one question for one paragraph. The students had simply to choose the right answer within 100 minutes. The second test included the same thirty paragraphs and the students had to work a bit harder to write down the inference of the main idea of each paragraph within the same time length. These tests were derived from Elder (2017a), Elder (2017b) and Dalton State (2014). Additional items were selected from Vocational Preparatory Instruction (n. d) that includes Main Idea Part 1 and Main Idea Part 2.

The data were the scores of the students’ performance on those two tests, the multiple-choice and the short-answer tests. The answers to the multiple-choice questions were scored as either right or wrong whereas the answers to the short-answer questions were scored according the short-answer question rubric with the inter-rater coefficient of 0.7. The score per item in the rubric ranges from 0 to 1 for each answer. Each score was converted to the 1-to-100 scale.

IV. Results and Discussion

The excel operation of the paired-test has given p-value $2.52252E-11 < 0.05$ and the Minitab calculation p-value $0.000 < 0.05$.

Table 3: Results

Test Format	Mean	Inter-rater	Paired-t test	
			Excel	Minitab
Multiple-Choice	74.04	0.7	p-value $2.52252E-11 < 0.05$	p-value $0.000 < 0.05$
Short-Answer	34.74			

The results confirm that there was a significant difference between the students’ performance on the multiple-choice and the short-answer tests. This finding is in line with the finding by Sam et al. (2016) that scores in SBA (single best answer) were found to be higher than the scores in VSA (very short answer). This finding also supports the finding by Funk & Dickson (2011) that “Students

performed significantly better on items presented in a multiple-choice format”. However, the finding is not in line with the result of the study by Kastner & Stangla (2011) that “that CR tests are equal to MC tests with multiple responses if Number Correct (NC) scoring is used.” Another finding is posed by Mullen & Schultz (2012) that “a similar percentage mark for their MC as for their SA sections.

There are possible reasons to account for that finding. First, the poor performance on S-A test may have been caused by the limited time (100 minutes) to comprehend the paragraphs and to write the answers in their own sentences whereas in the MC test the students didn’t have to write sentences. This time limitation was indicated by the number of sentences. Almost all of the students could finish writing 26 answers out of 30 within 100 minutes.

Second, there was probably lack of exercises in SAQs that led to failure in summarizing one paragraph into one sentence, as supported by individual lengthy sentences that much exceeded the space provided for writing the answers on the answer sheet. Moreover, the scarcity of exercises in inferring main ideas in the course book may probably account for the poor performance. There were very few questions about main ideas. This happened although they were trained to infer main ideas in the classroom interaction. This colloquial interaction certainly was not enough.

Finally, the test takers were probably much more dealing with factual details in reading than concepts (Simkin and Kuechler, 2005) in such a way that SAQs became much demanding.

V. Recommendation

There are three recommendations. The first is about test developer and test administration. One can not rely much MCQs for measurement of inferential comprehension. A blend of MCQs and SAQs would be a fair instrument with higher weighting on scoring the SA items because SAQs require more struggling for accuracy in ideas and words to be the answers. For this purpose, test administration should go beyond practicality and reliability. University lecturers should be encouraged to spend much time and energy in scoring SAQs by implementing a rubric. This will provide the students with the opportunity to develop their conceptual framework in reading for comprehension.

The second is about reading classes. Additional exercises should be provided to train the students in drawing inferences especially in the formulation of main ideas.

The last recommendation is for further research. This study did not investigate the different performances of the male and female students because the number of the test takers was much dominated by female students. Further research is expected to address this issue by a proportional sample of female and male subjects.

References

- Aamodt, M. G., & McShane, T. (1992). A meta-analytic investigation of the effect various test item characteristics on test scores. *Public Personnel Management*, 21(2), 151–160. Retrieved from https://www.researchgate.net/publication/266737879_The_effects_of_the_number_of_options_on_the_psychometric_characteristics_of_multiple_choice_items
- Baghaei, P. & 1 & Amrahi, N. 2011. The effects of the number of options on the psychometric characteristics of multiple-choice items. *Psychological Test and Assessment Modeling*, Volume 53, 2011 (2), 192-211. Retrieved from <https://pdfs.semanticscholar.org/9991/fb47cc50ff7f1d9a694d5ec72b9b34f26f69.pdf>
- Bridgeman, B; Morgan, R. 1996. Success in college for students with discrepancies between performance on multiple-choice and essay tests. *Journal of Educational Psychology*, Vol 88(2), Jun 1996, 333-340. Retrieved from <http://dx.doi.org/10.1037/0022-0663.88.2.333>
- Chan C. (2009) Assessment: Short Answer Questions, Assessment Resources @HKU, University of Hong Kong. Retrieved from http://ar.cetl.hku.hk/am_saq.htm
- [Crehan](#), K. D, [Haladyna](#), T. M, & [Brewer](#), B. W. 1993. Use of an Inclusive Option and the Optimal Number of Options for Multiple-Choice Items. Retrieved from <http://journals.sagepub.com/doi/10.1177/0013164493053001027>
- Dalton State. (2014). Implied Main Idea Practice. Retrieved from https://dscweb.daltonstate.edu/writing-lab/practice-exercises-for-reading-classes/implied_main_idea1.htm
- Delgado, A.R. and Prieto, G. 1998: Further evidence favoring three-option items in multiple-choice tests. *European Journal of Psychological Assessment* 14, 197–201. Retrieved from https://www.researchgate.net/publication/232600748_Further_Evidence_Favoring_Three-Option_Items_in_Multiple-Choice_Tests
- Digital Chalkboard. 2015. Selected-Response vs. Constructed-Response Assessment Methods. Retrieved from <https://www.mydigitalchalkboard.org/portal/default/Content/Viewer/Co>

- [tent?action=2&scId=505706&sciId=15387](#)
- Elder, Janet (2017a) Implied Main Idea, Exercise 1 Retrieved from http://highered.mheducation.com/sites/0073123587/student_view0/chapter10/implied_main_idea_exercise_1.html
- Elder, Janet (2017b) Implied Main Idea, Exercise 2. Retrieved from http://highered.mheducation.com/sites/0073123587/student_view0/chapter10/implied_main_idea_exercise_2.html
- Funk, S.C. & Dickson, K.L. 2011. Multiple-choice and short-answer exam performance in a college classroom. *Teaching of Psychology*, 38 (4), 273-277. <http://www.open.ac.uk/blogs/SallyJordan/?p=809>
- Gay, L. R. 1980. The Comparative Effects of Multiple-Choice versus Short-Answer Tests on Retention. *Journal of Educational Measurement*. Vol. 17, No. 1 (Spring, 1980), pp. 45-50. Retrieved from https://www.jstor.org/stable/1434671?seq=1#page_scan_tab_contents
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–333. Retrieved from http://dx.doi.org/10.1207/S15324818AME1503_5
- Kastnera, M & Stangla, B. 2011. Multiple Choice and Constructed Response Tests: Do Test Format and Scoring Matter? *Procedia Social and Behavioral Sciences* 12 (2011) 263–273. Retrieved from https://ac.els-cdn.com/S187704281100125X/1-s2.0-S187704281100125X-main.pdf?tid=3401ac54-dc84-483e-a44c-1f7cb6ce4865&acdnat=1534231832_69593c6ba011ce2ed95687bf663489b3
- Kim, Kyung-Hee. (n.d). Shorter Answer Questions. Eastern Michigan University Retrieved from <https://kkim.wmwikis.net/file/view/Lecture9ppt+Short+Answer+Questions.pdf>
- Landrum, R. E., Cashin, J. R., & Theis, K. S. (1993). More evidence in favor of three-option multiple-choice tests. *Educational and Psychological Measurement*, 53 (3), 771–778. doi:10.1177/0013164493053003021. Retrieved from

https://scholarworks.boisestate.edu/psych_facpubs/119/

Mullen, K & Schultz, M. 2012. Short Answer versus Multiple Choice Examination Questions for First Year Chemistry. *International Journal of Innovation in Science and Mathematics Education IJISME* [Vol 20, No 3 \(2012\)](#). Retrieved from <https://openjournals.library.sydney.edu.au/index.php/CAL/article/view/6441/7096>

Parmenter, David A. 2009. Essay versus multiple-choice: student preferences and the underlying rationale with implications for test construction. Retrieved from <http://www.freepatentsonline.com/article/Academy-Educational-Leadership-Journal/208747754.html>

Ramraje, S. N & Sable, P. L. 2011. Comparison of the Effect of Post-Instruction Multiple-Choice and Short Answer Tests on Delayed Retention Learning. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3562952/pdf/AMJ-04-332.pdf>

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24 (2), 3–13. doi:10.1111/j.1745-3992.2005.00006.x. Retrieved from <http://psycnet.apa.org/record/2005-07701-002>

Sam, A. H., Hameed, S., Harris, J. & Meeran, K. 2016. Validity of very short answer versus single best answer questions for undergraduate assessment. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5064885/pdf/12909_2016_Article_793.pdf

Simkin, M. G. & Kuechler, W. L. 2005. Multiple-Choice Tests and Student Understanding: What Is the Connection? Retrieved from <https://pdfs.semanticscholar.org/754d/21c5f0863e36f1620d862c4e65dbfdcba849.pdf>

Straton, R. G., & Catts, R. M. (1980). A Comparison of Two, Three and Four-Choice Item Tests Given a Fixed Total Number of Choices. *Educational and Psychological Measurement*, 40(2), 357-365. Retrieved from <http://psycnet.apa.org/record/1980-31468-001>

Tychonievich, L. 2012. Multiple choice or short answer? The intellectual impact of question format. Retrieved from

<http://www.cs.virginia.edu/~lat7h/blog/posts/206.html>

Vocational Preparatory Instruction – Construct Meaning: Main Idea Part 1

Retrieved from

<http://www.fldoe.org/core/fileparse.php/3/urlt/cmml1.doc>

Vocational Preparatory Instruction – Construct Meaning: Main Idea Part 2

Retrieved from

www.fldoe.org/core/fileparse.php/3/urlt/cmml2.doc

Weimer, M. 2015. Advantages and Disadvantages of Different Types of Test Questions. Retrieved from

<https://www.facultyfocus.com/articles/educational-assessment/advantages-and-disadvantages-of-different-types-of-test-questions/>