# Short vs. Extended Answer Questions in Computer Science Exams

Alejandro Salinger
David R. Cheriton School of Computer Science
University of Waterloo
ajsalinger@cs.uwaterloo.ca

**Abstract**

Short-answer questions such as multiple-choice, true-or-false, or fill-in-the-blank, are popular among instructors when designing written exams in different fields. These types of questions, especially multiple-choice, are suitable for testing a broad range of course topics, while enabling efficient marking and timely feedback for test takers. However, detractors argue that short-answer questions fail to measure high level cognitive skills and encourage surface learning approaches.

In this work, we study the suitability of short and extended answer questions in written exams in University-level Computer Science courses. We present an overview of relevant literature on the topic. We discuss whether both formats can be used to assess the same set of skills; we review Computer Science instructors' perspectives about the use of multiple choice items in introductory programming courses; and we discuss the influence of assessment types on students' learning approaches.

We finally present our own perspectives on the issue. We argue that, in general, Computer Science courses and jobs require a set of skills that is difficult to assess with short answer-questions, and that extended answer questions encourage the development of such skills. We suggest that even if short-answer questions can be used for practical motivations, these should be complemented with extended-answer questions in written exams and other types of assessments, even for introductory courses.

# Contents

# Introduction

It was the first midterm of my undergraduate exchange program at the University of California, San Diego (UCSD). The course was Image Processing. The exam had 9 questions and we had only 50 minutes to write it. The questions seemed to require fairly short answers, but they were by no means trivial. Three of these questions involved sketching the Fourier Transform of given functions. It was clear that there was no time to calculate the transform and then do the sketch. It seemed that one had to be able to recognize the type of function, know what the transform of such function should look like, and sketch it. I realized that I had not studied properly for such type of question, and the result was fairly disappointing. What I was used to from my undergraduate program at the University of Chile was 3-hour long midterms, with 3 or 4 questions, which required proving a theorem, designing an algorithm, or working through an application problem. I always had time to figure details out during the exam. After that first midterm at UCSD, I realized that if I wanted to succeed in the courses during my exchange, I had to adjust the way I studied, focusing on the details and learning them by heart. Not all, but many of the written exams of the courses I took had many short-answer questions. If anything, they seemed easier than what I was used to, but I just had to study differently. It wasn't hard, but the experience raised some questions in my mind, and my desire to find answers grew later when marking exams of undergraduate courses at the University of Waterloo: Why were the written exams at UCSD and the ones of many courses at UW so different from the ones I had at the University of Chile? They seem easier, but is that really the case? Is it really expected from students to memorize so much? What about testing students' creativity and problem solving skills?

Motivated by the above questions, in this project we study the suitability and implications of short-answer questions (e.g., multiple-choice, true-or-false, fill-in-the-blank) and extended-answer questions (e.g., problem solving, code writing) in University-level Computer Science written summative assessments. We first define the types of questions we consider and give a general comparison of both types. We then discuss whether both formats can be used to assess the same set of skills; we review Computer Science instructors' perspectives on the use of multiple choice in introductory programming courses; and we discuss the influence of both types of assessment methods on students' learning approaches. Finally, we provide our own perspectives and argue that, in general, Computer Science courses and jobs require a set of skills that is difficult to assess with short-answer questions, and that extended problem solving questions encourage the development of such skills. Therefore, an effort should be made to use extended-answer questions in Computer Science courses, even if in addition to short-answer questions.

## Question Formats in Written Examinations

In this section we briefly describe various types of question formats that are commonly used in written exams[1] in Computer Science. The distinction between short and extended-answer questions is not unique. The distinction we make in this work is related to the length and depth of the process involved in answering the question, which includes the length of the description of the answer itself.

### *Short Answer Questions*

We regard as a short-answer question any question whose answer is chosen among a set of available options, and questions whose answers and thought process

[1] We use the term *question* to refer to any item that requires an answer, although the item might not be posed in the form of a question. In the literature the term *item* is sometimes used to avoid this inaccuracy.

can be specified in no more than one or two sentences. Some examples are:

1. **Multiple-choice (MC)**: consists of a statement or question (the stem), followed by a list of options representing possible answers. These consist of one correct option and a set of *distractors*.

   **Example**. What is the worst-case time to sort $n$ items using quicksort?

   (a) $\Omega(n)$ (b) $\Omega(n \log n)$ (c) $\Omega(n(\log n)^2)$ (d) $\Omega(n^2)$.

2. **True-or-false**: a claim is presented and the test taker must indicate if the claim is true or false. Sometimes a justification is required.

   **Example** The worst-case time to sort $n$ items using quicksort is $\Theta(n \log n)$.

3. **Fill-in-the-blank**: a statement is given with one or more missing words or expressions. The test taker must write the correct expression in each blank. A set of options for possible expressions might or might not be provided.

   **Example** The worst-case time to sort $n$ items using quicksort is _____.

4. **Brief-answer**: a question requiring an answer that can be specified in at most one or two sentences[2].

   **Example** What are the best-, average-, and worst-case times to sort $n$ items using quicksort?

*Extended Answer Questions*

We consider a question to be an extended-answer question if the answer consists of several sentences or paragraphs, or it requires to show the process involved in obtaining the answer. This terminology is not a standard one. The terms *constructed response* and *free response* are used in the literature to refer to questions that require answers that must be produced by the test taker, as opposed to being chosen from

[2]These are commonly called short-answer questions, but we use the name brief-answer here and reserve short-answer for the more general type.

a set of options. These might include fill-in-the-blank and brief-answer questions, but usually refer to essay and problem solving questions. Thus, extended-answer questions are all free-response and constructed-response questions. We describe below the extended-answer questions that are common in Computer Science exams.

1. **Code writing**: it requires to write a program in a programming language or pseudo-code to solve a given problem.

   **Example** Given an array $A$ of $n$ integers, sorted in increasing order, and an integer $x$, write a program that returns the position of $x$ in $A$ or -1 if $x$ is not in $A$.
2. **Problem solving**: given a problem, such questions require answers that may range from a simple mathematical calculation to applying a multiple-step process to obtain a solution from the given information.

   **Example** Solve the following recurrence $T(n) = 2T(n/2) + O(n \log n)$.
3. **Proof**: A property or mathematical statement is given and the test taker is required to show that the statement follows logically from the assumptions of the system or model.

   **Example**. Prove that in the comparison model any sorting algorithm requires at least $\Omega(n \log n)$ comparisons in the worst case when sorting an input of size $n$.

*General Comparison*

We now mention some of the general characteristics that make short-answer questions (SAQ) preferable over extended-answer questions (EAQ) and vice versa. Among SAQ we focus mainly on multiple-choice and true-or-false questions, since they seem to be the most commonly used and studied types of short-answer questions. Among the advantages of SAQ over EAQ we can mention (Simkin & Kuechler, 2005; Industrial & Organizational Psychology, n.d.; Kuechler & Simkin, 2010):

1. SAQ enable an efficient way of collecting and grading examinations from a large number of test takers, while grading of EAQ require significant resources (e.g., time and trained graders).
2. The grading process of SAQ is objective and is perceived as such. EAQ are prone to subjective grading and special effort must be made to ensure objectivity.
3. SAQ can provide more timely feedback to test takers.
4. SAQ allow testing students on a wider range of topics than EAQ, since SAQ take much less time to answer than EAQ.
5. SAQ avoid the problem of favouring students with superior writing skills compared to students who might have superior knowledge of an answer but write it poorly.
6. SAQ are suitable for the creation of multiple versions of the same test, thus helping to minimize cheating.
7. SAQ provide multiple statistics and tests themselves can be evaluated on a question-by-question basis.

On the contrary, EAQ might be preferable to SQ for the following reasons, among others (Simkin & Kuechler, 2005; Industrial & Organizational Psychology, n.d.; Kuechler & Simkin, 2010; Livingston, 2009):

1. Effective SAQ are difficult to design, while EAQ usually take less time to construct.
2. It is easier to design EAQ that test high-order skills than to do this with SAQ.
3. In multiple-choice questions, distractors that are too close to the correct answer can make a test taker who knows the correct answer (without it being given as an option) to choose a nearly-correct answer instead.
4. EAQ allow for partial credit for partially correct or incomplete answers.
5. EAQ make more difficult for students to obtain marks by guessing correct answers.
6. EAQ are suitable for designing tests that are close to actual work situations.

7. EAQ enable assessing abilities to organize thoughts into a cohesive written response.

## Do Multiple-Choice and Constructed-Response Exams Measure the Same?

The multiple advantages of short-answer questions, especially of multiple-choice questions (MCQ), make them a popular choice when designing tests in Computer Science and other fields. Several studies have tried to provide evidence to show that multiple-choice tests measure the same set of skills as tests with constructed- and free-response questions, and thus the former could completely replace the latter. If this was indeed the case, then much effort in administering and grading tests with free-response questions could be saved. Other studies present evidence of the contrary, arguing that free-response questions can provide information about test takers that is impossible to obtain with multiple-choice questions. The issue is an ongoing debate.

Many of these studies have focused on Advanced Placement (AP) tests[3]. All except one of the AP tests are composed of a multiple-choice section and a free-response section, either essay or problem solving. Lukhele, Thissen, and Wainer (1994) analyze the results of the multiple-choice and free-response sections of AP exams in Chemistry and United States history, concluding that "constructed-response items provide less information in more time at greater cost than do multiple choice items", and that there is no evidence that the two kind of questions measure fundamentally different things. They argue as well, based on an analysis of the average information provided by each item, that for a middle level examinee one requires about 20 times more examination time to obtain the same amount of information with an essay as one

[3] Advanced Placement courses are standardized courses offered to high school students in the USA and Canada and are regarded as equivalent to college-level introductory courses. See `http://apcentral.collegeboard.com/apc/Controller.jpf`.

would obtain with MC items. They argue that multiple-choice questions can therefore replace constructed-response questions. A similar analysis was performed on the AP Economics exam of 1991 (Walstad & Becker, 1994), with results suggesting that the score of the essay adds minimal information about the student for determining the AP grade beyond what was is already given by the multiple-choice score.

Bennett, Rock, and Wang (1991) assessed the equivalence of multiple-choice and free-response items in the AP exams in Computer Science of 1998. This test is composed of 50 MC items and 5 free-response questions requiring the test taker to write or design programs or data structures and to analyze the efficiency of certain operations. The study poses a two-factor model to test the relationship of the skills measured by multiple-choice and free-response items. The results of the analysis suggest that a single factor provided the most parsimonious fit, leading the to the conclusion that there was “little support for the stereotype of multiple-choice and free-response formats as measuring substantially different constructs (i.e., trivial factual recognition vs. higher order process)”. They argue that a reason for this conclusion might be that although no single multiple-choice question is likely to test more than one of the various processes required to solve a free-response questions (e.g. to decompose specifications into goals, formulate a plan for each goal, translate plans to code, and simulate and debug this code), in combination, 50 of such questions might cover in some depth many of these processes. An informal analysis of the contents of the test supports this suggestion.

While many empirical studies such as the ones described above have led researchers to conclude that MCQ and free-response questions can measure the same set of skills, several other studies present arguments against this idea, either finding no or small relationship between student performance on MC and free-response portions of examinations in various fields (Kuechler & Simkin, 2010).

Livingston (2009) argues that in general, studies exhibiting evidence for MC and constructed-response (CR) questions providing the same information have been carried out with a single group of test takers, which can mask important differences between groups of test takers. For example, studies have shown that differences between results of male and female test takers on CR do not parallel differences between these groups on MC questions. Typically, when men and women have similar performance on MC questions, women outperform men in CR questions. On the other hand, when both groups perform equally well on CR questions, men perform better than women on MC questions. Another argument against MC replacing CR questions is that even though studies show correlations between performance of students in questions in both formats, that does not mean that the two measures will change in the same way over time. In addition, if only MC questions are used, teachers have incentives to emphasize the kinds of skills and knowledge tested by MC tests, giving lower priority to activities that teach the skills better measure by CR tests.

Simkin and Kuechler (2005; 2010) claim that previous studies yielded mixed results because they treat MC questions as homogeneous, and that a factor that might explain the differences is the relative difficulty of questions. Thus, focusing on basic computer language programming tests, they study the relation between MC and CR, but at different cognitive levels. For this sake, the classify MC and true-or-false questions from classes in Internet Programming and Visual Basic, according to the cognitive level they test in Bloom's taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). They suggest that only the first three levels of the taxonomy (Knowledge, Comprehension, Application) apply to undergraduate, introductory programming courses. Interestingly, they found extremely difficult to construct an MC question at the Application level, since an MC question precludes the student's ability to create a solution. The closest approximation they could find was a "find and fix error" type of question: given a piece of code and its intended purpose, the question asks for a way to fix it so that the purpose is achieved. The options give the possible fixes. While it is possible to answer this question by mentally programming the English language description of the code and then choosing an answer, it is certainly possible (and likely faster) to find the right answer by implementing each change given in the options and tracing the resulting code. The latter, however, is a comprehension-level skill, not an application one.

The analysis suggests that no combination of MC questions can exactly match the skills required to solve certain CR questions. However, it also shows that there are substantial differences between MC questions. The authors therefore suggest the hypothesis that if questions at the same level of understanding are asked in different formats, there should be a high correlation between scores, and furthermore, given a mixture of MC questions at different levels, the correlations should increase with MC question difficulty. However, this refined analysis yielded weak correlations, thus providing evidence that the cognitive process involved in answering a CR question appears to be much richer than the process to answer an MC question, even if both questions are at the same knowledge level.

In short, there is no consensus about the extent to which both formats can measure the same skills. After many years and many studies in this respect, the evidence is inconclusive (Martinez, 1999).

## Instructors' Perspectives on Multiple-Choice Questions for Programming Courses

Multiple-choice questions appear to be the preferred type of test used by many instructors. For example, it has been estimated that the usage of MC tests is between 45% and 67% in economics courses (Simkin & Kuechler, 2005). The literature on testing pedagogy suggests that usually both instructors and students prefer MC test to other formats.

To investigate the perspectives on MCQ of instructors in introductory Computer Science courses, Shuhidan, Hamilton, and D'Souza (2010) gathered views and opinions of programming courses instructors about the appropriateness of MCQ in summative assessments, and their evaluations of the skills tested by each of a given sample of questions. The majority of instructors believe that MCQ test students' understanding and encourage students to think carefully before choosing the right answer. Nevertheless the majority of instructors agreed that MCQ assess low levels of understanding of concepts. It is also interesting to note the reasons why instructors in the group use MCQ. Examples of these reasons are: to get an idea of the breadth of students' understanding; to give weaker students confidence to answer questions; to test understanding of fundamental programming terminology; because they provide a shorter feedback time; to gain analysis skills; to keep students happy; to constrain students' creativity. About 10% of the respondents indicated that they do not support the use of MCQ in summative assessments, one instructor saying, "I feel that multiple choices is a completely inappropriate tool for judging deep understanding and comprehension of programming concepts", and another one stating, "You need to include essay questions because Computer Science students need to know how to write". Another responded that although he used MCQ, he deems them as free points and uses them to keep weaker students on track.

The results of instructors' analyses of the given questions in terms of the skills they measure (language syntax knowledge, language semantic knowledge, problem solving skills) and overall difficulty, suggest that there are no significant correlations between the amount of experience in teaching or devising exams and any of the measures. This, the authors say, answers a popular misconception of students that more experienced instructors evaluate questions harder than less experienced ones.

## Influence of Question Format on Students' Approach to Learning

The primary goal of summative assessments in a course is to measure the extent to which the learning outcomes have been attained by students. However, assessments are not isolated from the learning process. An assessment is naturally also a learning instance that requires preparation from students. Consequently, their type, format, expectations, and perception, have a significant influence on the way students approach the preparation process.

Students take different approaches when studying. A *surface* approach is associated with a focus on recall and reproduction, routine and unreflective memorization, and procedural problem solving. This approach is associated with an intention to complete a learning task with little personal engagement. In contrast, a *deep* approach to learning stems from an intention to understand and it is associated with active conceptual analysis (Scouller, 1998; Struyven, Dochy, & Janssens, 2005).

There is a large body of research that associates MCQ exams with the employment of surface learning approaches. Scouller (1998) studied the relationship between students' learning approaches, their perceptions of the intellectual levels that assessments measure, their preference of type of methods and their performance, for an MCQ exam and an assignment essay in a second year Education course at the University of Sydney. Results revealed the students were significantly more likely to employ surface learning approaches when preparing for the MCQ test and deep learning approaches when preparing for the assignment essay. In addition, MCQ was perceived as assessing knowledge-based or lower levels of cognitive skills, while the essay was perceived as assessing higher levels such as analysis, application and comprehension. It was found that a perception of MCQ as assessing low levels cognitive skills was associated with successful outcome in that kind of exam. Similarly, the employment of deep learning approaches and a perception of essays as assessing higher levels was associated with better marks in the assignment essays, while marks where negatively related to surface strategies. This supports the idea that students' perception of an assessment plays an important role in students' choices for a type of learning approach, and consequently in their outcomes.

Performance in each type of assessment was also related to students' preferences: students who preferred essays reported using deep approaches when preparing their essays and performed better in their assignment than students who preferred the MCQ exam, who reported using surface strategies for their assignment. It is interesting that students that preferred MCQ exams also regarded these exams as assessing higher level cognitive processes, and this perception was associated with worse performance in this exam. Another interesting finding was that the employment of deep strategies when preparing for the MCQ exam was predictive of poor results in this exam. This should be a concern, considering that most courses have learning outcomes related to skills associated with deep approaches to learning (Scouller, 1998).

Struyven et al. (2005) reviews further the influence of students' perception of assessments and its influence on their learning approaches. They claim that assessment methods that are perceived to be inappropriate encourage surface learning approaches, but that appropriate evaluations are not enough to achieve deep approaches. There are several factors that may influence the perceived appropriateness of an assessment by students. One of them is their preference. Students favour multiple-choice types exam in terms of perceived difficulty, anxiety, complexity, and success expectancy. However, not all students prefer MC tests; it has been reported that students regard essay exams more favourable for the purpose of representing their knowledge in the subject being tested, and that students with good learning skills and with low test anxiety tend to prefer essay type of assessments. Assessment expectations also influence students' behaviours. For example, written notes from students expecting a multiple-choice exam differ from notes of students that expected an essay test. The former focused on facts and details, while the latter concentrated on information of higher structural importance. A study that considered alternative types of assessments (e.g., portfolios, self and peer evaluations, simulations) found that traditional assessment tasks are perceived as arbitrary and irrelevant by students, and that they aimed to learn only for the particular assessment, which was seen as a means to acquire marks (Sambell, Brown, & McDowell, 1997). On the other hand, alternative assessments were perceived to measure qualities and skills that would be valuable in contexts other than the particular assessment. For example, an assessment that attempts to produce an activity that simulates a real-world situation would make students perceive the relevance of their work beyond the particular examination.

Although assessment methods seen as "appropriate" by students do not ensure deep learning approaches, understanding students' perception of assessments and their influence on their learning approaches can guide instructors in the design of assessment methods to motivate students and encourage deeper learning approaches.

## Discussion

We have described studies addressing various aspects of question formats in written examinations. Based on these and our own perceptions, in this section we argue that extended-answer questions should be used in Computer Science assessments, although not necessarily precluding the use of short-answer questions. We base our argument in that extended-answer questions are more suitable for assessing the skills involved in most Computer Science courses, that they constitute a valuable learning experience, and that they are closer to real-world scenarios in the field.

*Assessments and Intended Learning Outcomes*

The majority, if not all courses in a Computer Science curriculum should have Intended Learning Outcomes (ILOs) that go beyond the Knowledge and Comprehension levels of Bloom's taxonomy. Even in introductory programming courses, students are expected to be able to create programs for problems that were not covered in class, skills which are at the Application level. Constructing multiple-choice questions at the Application level has been found a challenging task, and nearly impossible at higher levels of learning. We assume that this extends to other types of short-answer questions such as fill-in-the-blank and true-or-false, and to brief-answer questions perhaps to a lesser extent. Therefore, a written test that relies mainly on short-answer questions is not suitable for assessing many of the ILOs of most Computer Science courses. In addition, since short-answer questions encourage surface learning approaches, assessing ILOs involving higher learning levels with short-answer tests will likely result in these ILOs not being attained by the majority of students.

Naturally, written exams with short-answer questions are not the only form of assessment. Summative assessments in Computer Science courses typically include various assignments, usually having extended-answer questions, including programming tasks. Assignments and other kinds of assessments could be used to both measure and encourage learning at higher cognitive levels. However, it is usually written examinations which carry the largest weight among the assessments in a course, presumably because they are better at reflecting students' actual individual skills than assignments in which the use of external aid could distort the relation between students' skills and outcome. In this case, an important part of a student's grade will be determined by the short-answer questions of written exams, minimizing the influence of higher level skills. Thus, in our opinion, if a course has ILOs related to high order skills, the use of short-answer questions in written examination should be complemented by either extended-answer questions in the same exams or by other types of assessments that are more likely to assess higher level skills. The latter should be carefully designed so that they effectively test students' individual skills, and should have an appropriate weight in a student's final mark.

*Assessment as a Learning Instance*

Any type of summative assessment is both a measure of students' performance and a learning activity. In this sense, a test that has a few extended-answer questions and plenty of time to be answered is an instance in which a student can reflect, derive ideas from principles, make connections between various ideas, evaluate their use in a particular scenario, be creative, and in general engage in a deep thinking process to solve challenging problems. All in all, a student will learn a great deal while writing the exam, while at the same time putting skills spanning various levels of Bloom's taxonomy to a test. Not only such tests can be useful for learning skills related to the particular course, but they are also effective at training students to organize and express ideas in written form by extracting the important arguments from a development of a solution and presenting them in a convincing manner in the final answer. These are important skills that are transversal to the courses of any University-level curriculum.

While it is true that marking such extended-answer questions involves costly resources, the feedback obtained both by instructors and students can be very valuable. By seeing a sample of the thinking process of a student while solving a problem instructors can identify misconceptions and reinforce the corresponding concepts in class. In contrast, a wrong-answer in a multiple-choice question does not provide much information to an instructor about the reason why a student did not pick the right answer. Similarly, a properly marked extended-answer question that includes written messages by the marker can be very useful for students for recognizing their shortcomings and for learning from expected processes and answers.

*Structural Fidelity*

Structural fidelity is the congruence between the performance called upon by a test and performance in the field (Simkin & Kuechler, 2005). Assessments with high structural fidelity enable assessing skills that are required in a real-world situation. In addition, they are also perceived as more appropriate by students, since they seem to provide value beyond the examinations themselves (Sambell et al., 1997). It should not be hard to argue that it is easier to achieve higher structural fidelity with extended-answer questions than with short-answer ones. Although not exclusively, many Computer Scientists deal with programming environments in their jobs. Clearly a question that requires writing a program is a much better approximation to a real-world activity in the field than any form of short-answer question. Moreover, although short-answer questions can be used to test programming skills, these skills are mostly limited to Knowledge and Comprehension levels. However, memorizing facts and programming language syntax is actually not particularly useful in a programming environment. In a real programming environment method documentation, sample usage, and libraries with implemented basic solutions are readily available for use, and thus the challenges are related to combining the small pieces into a larger solution. In this sense, programming questions in written exams are necessarily close to reality either. However, exams can be made to mimic a more real environment by, for example, including a reference sheet of basic methods, allowing students to bring a self-prepared cheat sheet, or by making the exam open-book, while at the same time modifying the nature of questions appropriately. Taking structural fidelity further, instructors could design evaluations that include open-ended, poorly defined problems, as it is usually the case in many real-world situations, in the spirit of the Generic Quiz of Felder (1985).

Finally, it could be argued that not all courses in the curriculum can aim to have assessments mimic real-world situations, and that short-answer questions are perfectly suitable for first-year courses. While this is true, if this kind of assessment is used exclusively in these courses, a gap will be created between the nature of assessment in first and upper year courses, and students will have a harder time adjusting to the requirements of the latter. Thus, we believe that extended-answer questions should be included in Computer Science courses from the very beginning of the program.

## Conclusions

In this work we presented a general comparison of the use of short-answer and extended-answer questions in written exams in Computer Science. We presented an overview of studies discussing the possible differences in skills that the two formats measure, their influence on students' learning approaches, and the perspectives of Computer Science instructors on multiple-choice questions. We provided our own points of view about the advantages of extended-answer questions in Computer Science courses in terms of their suitability to assess and encourage high learning levels, their structural fidelity, and their use as a learning instance. Although the consensus seems to be that in general extended-answer questions are a better assessment tool from the point of view of learning, we must yield to the fact that short-answer questions are more practical and are even necessary in some situations. Evidence for this is the large body of studies attempting to show that they can replace other formats. Nevertheless, we think that even if short-answer questions are used in Computer Science tests, assessments with extended-answer questions should be used, and should be given an important place in a course in terms of mark weight and students' effort.

## References

Bennett, R. E., Rock, D. A., & Wang, M. (1991). Equivalence of free-response and multiple-choice items. *Journal of Educational Measurement*, *28*(1), pp. 77-92.

Bloom, B. S., Engelhart, M. B., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives. the classification of educational goals. handbook 1: Cognitive domain.* New York: Longmans Green.

Felder, R. (1985). The Generic Quiz: A Device to Stimulate Creativity and Higher-Level Thinking Skills. *Chemical Engineering Education*, *19*(4), 176-181.

Industrial & Organizational Psychology, I. Society for. (n.d.). *Employment testing.* Available from `http://www.siop.org/Workplace/employment%20testing/employment_testing_toc.aspx` (03/19/12)

Kuechler, W. L., & Simkin, M. G. (2010). Why is performance on multiple-choice tests and constructed-response tests not more closely related? theory and an empirical test*. *Decision Sciences Journal of Innovative Education*, *8*(1), 55–73.

Livingston, S. A. (2009). *Constructed-response test questions: Why we use them; how we score them.* ETS, R&D Connections 11. Available from `http://www.ets.org/Media/Research/pdf/RD_Connections11.pdf` (02/20/12)

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement*, *31*(3), pp. 234-250.

Martinez, M. (1999). Cognition and the question of test item format. *Educational Psychologist*, *34*(4), 207-218.

Sambell, K., Brown, S., & McDowell, L. (1997, 0). "But Is It Fair?": An Exploratory Study of Student Perceptions of the Consequential Validity of Assessment. *Studies in Educational Evaluation*, *23*(4), 349–71.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, *35*, 453-472.

Shuhidan, S., Hamilton, M., & D'Souza, D. (2010). Instructor perspectives of multiple-choice questions in summative assessment for novice programmers. *Computer Science Education*, *20*(3), 229-259.

Simkin, M. G., & Kuechler, W. L. (2005). Multiple-choice tests and student understanding: What is the connection? *Decision Sciences Journal of Innovative Education*, *3*(1), 73–98.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: a review. *Assessment & Evaluation in Higher Education*, *30*(4), 325-341.

Walstad, W. B., & Becker, W. E. (1994, May). Achievement differences on multiple-choice and essay tests in economics. *American Economic Review*, *84*(2), 193-96.