QuesGen: Clarifying the Dimensions of Multiple-Choice Question Quality

Morgan C. Benton
James Madison University
Marilyn M. Tremaine
Rutgers University

Please address all correspondence to:
Morgan Benton
ISAT Department, MSC 4310
James Madison University
Harrisonburg, VA 22807
p/f: +1-540-568-6876/2955
bentonmc@jmu.edu

Paper presented at the
Annual Meeting of the American Educational Research Association
New York City, March 24-28, 2008

QuesGen: Clarifying the Dimensions of Multiple-Choice Question Quality

Morgan C. Benton
James Madison University
Marilyn M. Tremaine
Rutgers University

Many subtle factors impact multiple-choice question (MCQ) quality. Developing instructors' understanding of these factors remains a challenge. Yet, classroom instructors must understand the dimensions of question quality, else how will they know whether or not they have constructed good questions? An experiment was conducted to test the effectiveness of QuesGen—a web-based system designed to help instructors write better MCQs. Whereas in the professional testing industry advanced metrics based on item response theory (IRT) exist to measure question quality, in classroom settings, IRT-based metrics are impractical. This paper reports the results of the experiment that have an impact on the understanding of the dimensions of MCQ quality, and suggests designs for software that simultaneously support teaching and deepen practice.

"Is this going to be on the test?" This is a common question in today's classrooms—a question that reveals a lot about students' approach to learning and about the way teachers have structured the learning process. This question is code for the unasked question: "Is this material important (enough for me to commit it to memory)?" It reveals students' uncertainty and lack of confidence in their own ability to decide what is really important to take away from a class. It also reveals a certain level of apathy, or perhaps a hard-nosed practicality about how students have come to perceive their education: the goal of studying is to do well on the test. The way that reward structures have been established in the educational system has done little to dissuade students of this notion. Some universities go so far as to stipulate the approximate percentage of a semester grade that should be decided by midterm or final examinations. For better or worse, this situation makes the quality of the questions that appear on tests an important issue.

Writing test questions is not a skill in which many instructors in higher education receive explicit training. The QuesGen research project purposed to build an online tool that allowed teachers to enter and deliver multiple-choice quizzes. The key innovation of QuesGen was that it embedded training for teachers in how to construct of high quality questions. One of the key challenges of the QuesGen project was arriving at a reasonable definition of what constitutes a "high quality" question. The goal was to operationalize this definition in such a way that the tool could guide and provide feedback to instructors on the quality of their questions. This paper documents some of the discussion and analysis of what it means for an MCQ to be "high quality."

Theoretical Framework

What makes a good test question? Put simply, a good question is defined as one that helps an instructor learn the greatest amount about the current state of his or her students' understanding both at the individual and aggregate levels. High quality questions inform the decision making processes of both teachers and students as they choose what and how to teach and learn. Unfortunately, however, the questions asked on tests in classrooms frequently do not meet this standard of quality (Stiggins, 2001). In other words, the test questions do not provide a very reliable or accurate measure of what students understand, and therefore are not useful in guiding the learning decisions made by both teachers and students.

This definition conforms to the belief that formative assessment, as opposed to summative assessment, is the most appropriate form of assessment for fostering learning. Summative assessment tends to serve needs or actors outside of the classroom such as credentialing, entrance examination, program auditing, or program performance evaluation. The results of summative assessments, by definition, are not used to identify, plan, or otherwise inform the instruction that goes on in a classroom following the assessment. Summative assessments may improve learning inasmuch as they are designed for overall program or curricular improvement; however they seldom if ever are directly beneficial to the students who take the actual tests (Bloom, Hastings, & Madaus, 1971).

On the other hand, formative assessment, also known as assessment for learning (Black, Harrison, Lee, Marshall, & Wiliam, 2003), is primarily student focused. Formative assessments provide information to the instructor, the students, or both that can be used immediately to indicate whether a given topic has been mastered, or whether remediation is in order. At the class level, formative assessments can help an instructor make strategic choices about the formation of student teams, either pairing students with higher and lower levels of mastery, or grouping students according to mastery so that specific topics can be targeted directly to those students for whom they are most needed. This research project attempted to build a web-based software system that would support the creation of formative assessments, more specifically, assessments using MCQs.

A key challenge was devising metrics for question quality that would be accessible and understandable to classroom instructors. In the "high stakes" testing industry, complex statistical models of question quality are available. One of the most common is this 3-parameter logistic model (Hambleton, Swaminathan, & Rogers, 1991):

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \quad i = 1, 2, \ldots, n$$

Where:

$P_i(\theta)$ = the probability that a randomly chosen examinee with ability θ answers item i correctly

D = a scaling factor introduced to make the logistic function as close as possible to the normal ogive

a_i = the item i discrimination parameter

b_i = the item i difficulty parameter

c_i = the item i pseudo-chance level parameter

n = the number of items in the test

e = a transcendental number (like π) whose value is 2.718 (correct to three decimal places)

$P_i(\theta)$ = an S-shaped curve with values between 0 and 1 over the ability scale

When graphed, the item-characteristic curve (ICC) for the 3-parameter model looks as follows:

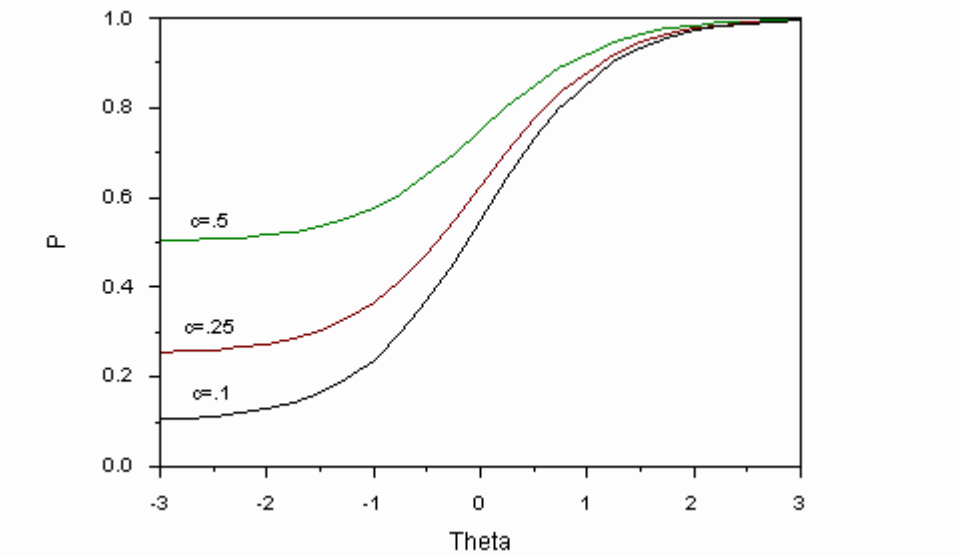


FIGURE 1. Sample ICC for the 3-parameter logistic model. Image from http://luna.cas.usf.edu/~mbrannic/files/pmet/irt.htm

For the 3-parameter model to be used, the values of the three parameters must be estimated. This is typically done using a procedure known as maximum-likelihood estimation (MLE). To successfully perform MLE for a 3-parameter model requires somewhere in the range of 500-1000 responses to each test item. Without going further into this description, suffice it to say that the time, energy, resources, and sophistication needed to do an IRT-based analysis are beyond the average classroom instructor. Furthermore, this level of analysis is probably not necessary to accomplish the goals of the test questions. For this reason, the QuesGen study chose to make use of some less robust, yet more easily accessible metrics for question quality: namely Bloom's taxonomy, the discrimination index, expert and student ratings.

Making use of these metrics to give useful guidance to instructors proved to be a challenge. The main reason for this challenge is that these metrics do not result in scores amenable to black and white interpretation. The rest of this paper will describe how the question of quality was clarified. The guiding question in the determination of measures of question quality is "Does this measure provide insight into aspects of a question that have an immediate impact on the teacher-student relationship, and the student learning that results thereof?"

Methodology and Data Collected

Three independent variables were manipulated in this study: course topic, instructor experience at writing MCQs, and the user interface of the web based system used for writing the MCQs. The primary dependent variable was broadly described as "question quality," and more specifically operationalized using a number of different metrics described below. The main hypothesis was that using an advanced web-based interface for writing MCQs, ie. QuesGen, would improve question quality. It was hypothesized that course topic would not have a significant impact upon question quality, but instructor experience was expected to play a significant role in the quality of questions written.

The QuesGen System

Two versions of a web-based system for writing multiple choice questions (MCQs) were created (see FIGURE 2 and FIGURE 3). The only difference between the two systems was in the interface via which instructors input their questions. In the "traditional" system, which was designed to mirror the functionality common in online learning management systems (LMSs), e.g. WebCT and Blackboard, the interface consisted of a textbox into which the question was typed, and several textboxes into which the answer choices were typed. In the experimental system, named QuesGen, additional scaffolding was added to aid instructors in the construction of questions that conform to empirically established best practices for writing MCQs (Haladyna, Downing, & Rodrigues, 2002). The scaffolding took four forms:

1. online video tutorials in how to write questions,
2. a dropdown menu for aligning educational objectives with questions,
3. online and printable versions of question templates that could be used to write MCQs targeting higher-order thinking
4. a question-quality checklist that guided instructors in reviewing their questions

Subjects

Twenty instructors from a large, public, liberal arts university were recruited for this study. All of the instructors taught between one and four sections of either an introductory kinesiology or an introductory communications course. These courses were part of the required general education program followed by all students at the university, and therefore the students enrolled in the courses represented a relatively random sample of freshmen and sophomores from the entire student body. Each of the twenty instructors was randomly assigned to use either the QuesGen system or the traditional system. In each of the two courses an online video lecture was prepared covering one of the scheduled topics of the semester.

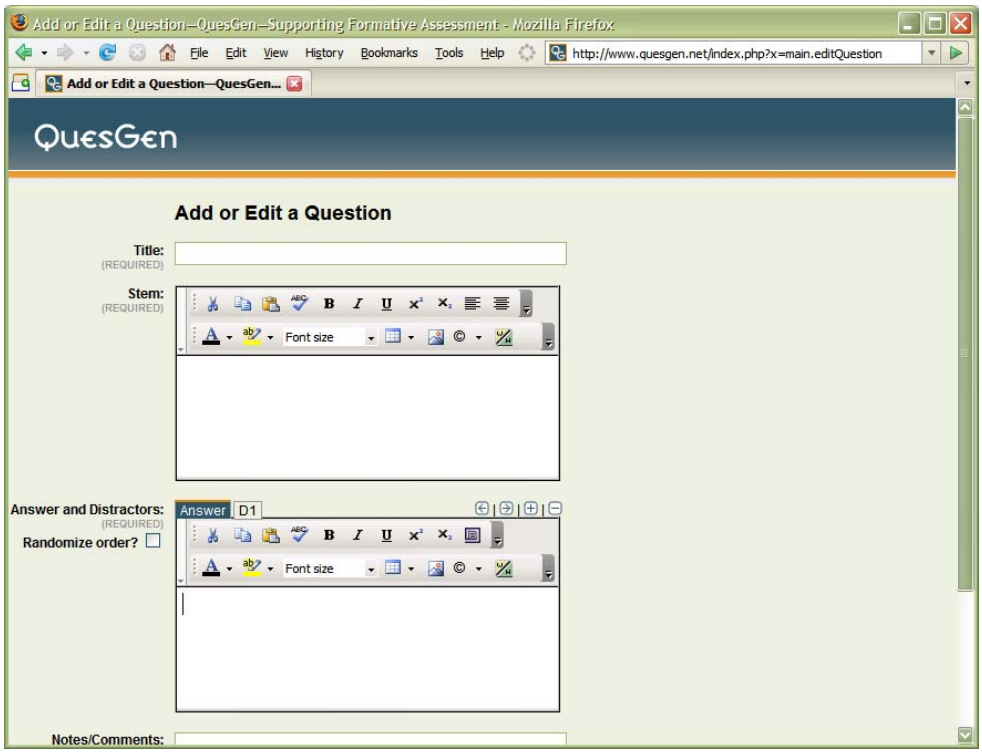


FIGURE 2. *The "Traditional" Interface only allowed entering the stem and answer choices. No other scaffolding or aids were provided.*

Week One—Instructors Write Questions

During week one of the study, each of the instructors watched the video lecture and wrote ten multiple-choice questions using the online system. Instructors were told that they were participating in an experiment to test a system for writing MCQs, but were not told whether or not the version of the system that they were using was the experimental version or not. The instructors had agreed to give these ten questions to their students as a quiz. Instructors were asked to make the quiz count for some small amount of course credit, such as a minor quiz or extra credit assignment. The goal of this reward structure was to provide some motivation for students to complete the quiz, but not so much motivation that they would be tempted to cheat on the quizzes. System user logs were collected which could be used to identify what features the instructors actually used. Furthermore, upon completing their ten questions, instructors completed a short, online, user-satisfaction survey.

Week Two—Students Take Quizzes and Survey

During week two of the study, an email was sent to the 1236 students enrolled in the two courses. Students were provided a link to the website where they could take their quiz, and a username and password. Upon logging in, students were instructed to watch the video lecture that their instructor had watched before creating the quiz questions. Once they watched the video they were given access to the online quiz made up of the questions written by their instructor. After completing the quiz, the students were asked to take a follow-up survey in

which they evaluated the quiz questions. By the end of week two, 820 students had completed the quiz, and 636 had completed the follow up survey.

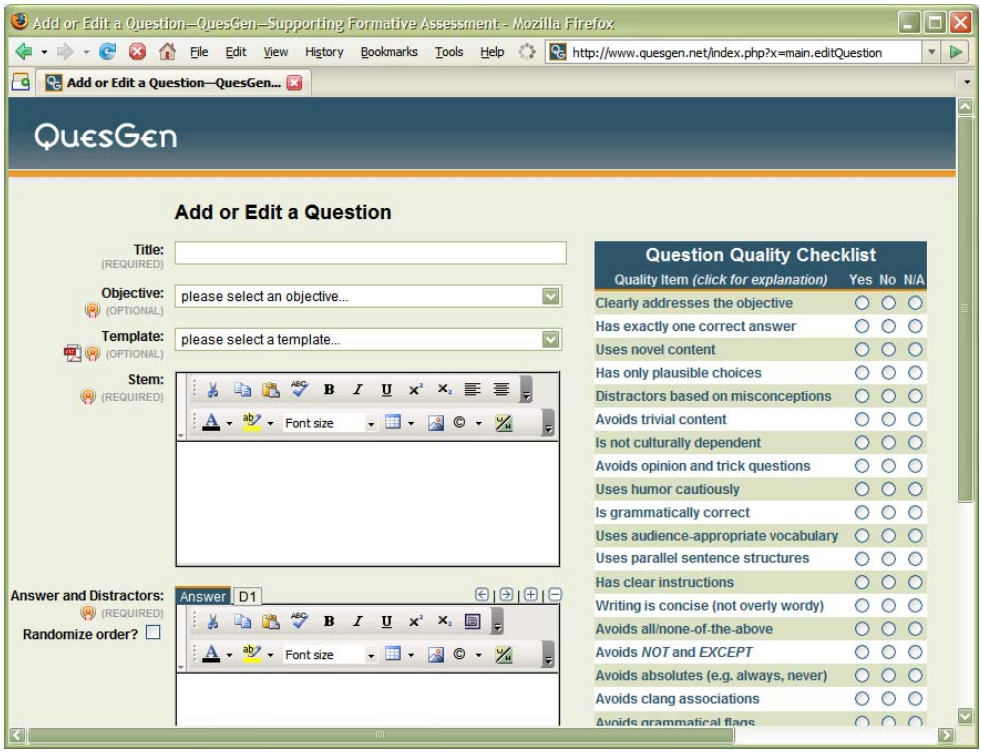


FIGURE 3. *The QuesGen Interface allowed questions to be aligned with course objectives, based on semantic templates, and evaluated with a "quality checklist." In addition, instructors had access to video tutorials on how to write good questions.*

Week Three—Expert Judges Review Questions

During week three of the study, two expert judges, who were both graduate students in an educational measurement and assessment program, evaluated each of the 200 questions that had been written by the participating instructors using a review instrument consisting of approximately 25 items. The instrument was based upon Bloom's taxonomy (Anderson & Krathwohl, 2001), and upon empirical work done by Haladyna et al. (2002). Once complete, the data from the expert judges was compiled along with data from the students' questionnaires, and a question analysis report was produced for each instructor consisting of summary data about each of the questions they had written.

Weeks Four to Six—Follow-up Interviews with Instructors

Finally, during weeks four through six, each of the participating instructors was contacted and a follow-up interview was conducted. Follow-up interviews lasted roughly an hour, about half of which was spent eliciting the instructor's attitudes towards MCQs, students, the course they taught, and the system they used. The other half of the interview time was spent debriefing the instructor and discussing the item analysis reports (see example Item Analysis Report in

Appendix A). The Item Analysis Reports were compiled for each of the ten questions written by each of the participating instructors and contained:

- The title of the question
- The question that was asked and the answers
- The number and percentage of students who selected each of the answer choices
- Students' evaluations of the question
- Expert judges' evaluations of the question
- The Discrimination Index

The Item Analysis Reports were generated automatically by the QuesGen system. These reports were received enthusiastically by the instructors. While it is not practical to have assessment experts rate every question of every professor, it is plausible that using the web a network of peers could be employed to generate ratings. Otherwise the information in the reports can be incorporated into every system. Information gleaned in the interviews was used to help build explanations about the meaning of the other data that was collected.

Study Limitations

Some limitations were encountered with the study. The most important limitation was the duration of the study. Since this study was conducted for a single quiz on a single unit, instructors did not have time to explore and make use of all of QuesGen's functionality. A more longitudinal design would have given them more opportunity to do so. A second limitation was that the item review instrument used by the expert judges yielded a floor effect, and as such did not provide as much resolution into which questions were actually of high and low quality as was hoped. This instrument will need to be revised before being used further. Third, there was an unforeseen overlap between course taught and instructor experience—two of the main independent variables. Eight out of ten of the instructors in the kinesiology course were graduate students with little or no experience writing MCQs, whereas all of the communications instructors had been teaching and writing MCQs for at least several semesters. As such, it is difficult if not impossible to disentangle the considerable effects of instructor experience and the courses that were taught. Despite these limitations, useful results were obtained.

Results

The QuesGen system turned out to be only very marginally effective at improving the quality of MCQs generated by instructors. The main reason that the system failed to have more of an impact appears to be that the instructors did not use much, if any, of the additional functionality. Almost none of the instructors using the QuesGen interface watched the videos. Almost none of them used the question templates. Almost none used the question quality checklist. In the few cases that instructors did use the functionality, their questions showed promising but not statistically significant results in terms of measurably higher quality. Future system designs will need to give better thought to how to entice instructors to use the available features.

Discussion—What is a "high quality" question?

Despite the disappointing results with respect to QuesGen's effectiveness, the process of evaluating the questions shed a great deal of light on the question of what a "high quality" question is. The theoretical framework section of this paper ended with the following guiding question for thinking about question quality:

Does this measure provide insight into aspects of a question that have an immediate impact on the teacher-student relationship, and the student learning that results thereof?

This question was written to guide the interpretation of all of the measures of question quality to be gathered in this study. The primary measures used to evaluate the quality of questions were: Bloom's taxonomy level, expert judges' evaluations, the DI, and student evaluations. This section will further deepen the discussion of what it means for a question to be of high quality, and, in light of the data, make a determination of whether or not QuesGen was effective at achieving its goal of helping teachers write better multiple-choice questions.

Does knowledge of where a question falls in Bloom's taxonomy have an immediate impact on the teacher-student relationship, and the learning that results thereof? Low-level questions only assess whether or not a student remembers a given fact or figure at a given point in time. Such questions do not give any indication of whether or not the student understands the significance, value, or application of that bit of information. Low-level questions allow students to regurgitate knowledge unreflectively. One could train a parrot to respond appropriately when asked for the name of the first president of the United States, but that doesn't indicate that the parrot knows the significance of this information. If the parrot can identify the name of the first president, does it mean that the parrot is now ready to move on to learning the significance of George Washington's decision to step down after only two terms in office? If this question seems nonsensical, it is meant to illustrate that low-level questions may not provide enough information for teachers to make pedagogical decisions about how and what topics to cover next with students.

Higher-order questions, on the other hand, require not only that students recall facts and information, but also that they be able to apply that information in new and different contexts. Does students' knowledge of Washington's precedent-setting two terms allow them to see the significance of FDR's being elected four times? Does it allow them to understand the nature of power and why we now only allow our presidents to serve at most two terms? This is the type of information that higher-order questions can elicit and it is the kind of knowledge that teachers need to make pedagogical decisions. The answer then is yes—knowledge of where a question falls in Bloom's taxonomy does have an impact on the teacher-student relationship. Therefore a measure of QuesGen's success is the degree to which it helped teachers write questions that were higher order.

Did use of QuesGen lead to the writing of higher-order questions? Teachers who used the question templates built into QuesGen were significantly more likely to write questions that assessed higher-order thinking. ***If*** teachers use this feature, then it appears that QuesGen is

effective at improving question quality in this dimension. A problem is that most of the teachers who had access to the question templates chose not to use them.

Do expert judges' ratings have an immediate impact on the teacher-student relationship, and the learning that results thereof? A person who has been trained to spot question flaws can spot questions that are likely to be unhelpful in identifying whether or not students have attained sufficient mastery in the subjects being studied. Catching flaws before questions have been delivered can give instructors time to adjust their questions for maximum effectiveness. Again, good questions inform sound pedagogical decisions on the part of the instructor. The answer to the question is yes—expert judges' ratings can have an immediate and positive impact on the teacher-student relationship. If QuesGen were to provide a mechanism that allowed the questions to be evaluated in the same way an expert judge would, it could improve question quality.

While it is unrealistic to have judges waiting to review every question that a teacher writes, it is not terribly difficult for a teacher to be trained to spot the flaws that an expert looks for. This was the motivation behind including the question quality checklist in QuesGen. The items in the item-review instrument were essentially the same items in the question quality checklist. Unfortunately, in the form that they were presented, instructors who used QuesGen did not use the checklist, and therefore the use of QuesGen was not related to a reduction in question flaws. As such, QuesGen was not effective in this dimension of question quality.

Does knowledge of the discrimination index (DI) have an immediate impact on the teacher-student relationship, and the learning that results thereof? Analysis of the DI results from the study indicated that the DI can be useful in identifying problems with item distractors. Since the DI can only be calculated *after* students have taken a quiz or test, the DI itself cannot be used by a teacher to avoid delivering subpar questions to the students. However, since it can be easily and automatically calculated based on the results of a quiz, the DI can be used by instructors to avoid a misdiagnosis of misconceptions held by the students about the material being studied. Since the DI can be used prior to returning feedback to students on the results of a quiz, the DI can have an impact on pedagogical decisions and the teacher-student relationship. Therefore DI can be used as one indicator of question quality. Results from the study show that DI scores that are either too high or too low indicate problems with distractors. As such a measure of QuesGen's effectiveness would be that it led to the production of questions with mid-range DI scores.

There was no difference in the DI on questions between those developed with and without QuesGen. In that respect, QuesGen was not effective at increasing question quality. The item-review instrument, particularly Factor 2, indicated a set of questions that are related to distractor quality. Enough information was gained from the study to make some concrete recommendations about the redesign of QuesGen. The question quality checklist was either too intimidating because of its length, or too easily ignored, or perhaps both. A way needs to be found to make it less intimidating, and more assertive about the way that it suggests changes to questions. Also since a large number of the checklist guidelines didn't seem to apply in many cases, some sort of pattern recognition that would enable the interface to "intelligently" notify the instructor of potential question flaws seems desirable.

Finally, does knowledge of students' evaluations of questions have an immediate impact on the teacher-student relationship, and the learning that results thereof? As with the DI, student's evaluations of questions can't occur until *after* the questions have been delivered, but student feedback is clearly relevant to the interpretation of question quality. In the QuesGen study, one instructor wrote a question that asked students to identify a "hedge" as a verbal indicator of power. About eight of the students indicated that they didn't know what the word "hedge" meant in the context of the question, and the instructor realized that this term had not been covered in the lecture. This is a serious flaw that is very unlikely to be found in any other way than through student feedback. However, given the trend identified in the follow-up interviews of neither students nor instructors reviewing the questions after a quiz, it is very likely that such errors are going unnoticed on a regular basis. Also in the study, there was a significant relationship between the questions judges rated as verbose and questions students marked as unclear. A third result from the student feedback was the occurrence of the selection of "careless error" as the reason students missed questions. Careless error indicates that students actually knew the correct answer, but for some reason clicked on the wrong button. This causes problems for interpretation of question results. All of these results indicate that yes, students' evaluations of questions are meaningful indicators of quality. As such an indicator of QuesGen's effectiveness would be if use of QuesGen was associated with students rating questions to be clearer and fairer. This was indeed the case. Use of QuesGen was significantly related to students' rating questions as clear and fair.

In summary, QuesGen was effective in some ways, and ineffective in others. QuesGen's template feature is associated with the writing of higher order questions, which in turn have a greater potential to inform pedagogical decisions. Questions written with QuesGen were more likely to be rated as clear and fair by students, which is another indicator that the system helped improve question quality. On the other hand, the use of QuesGen was not associated with a distribution of DI scores that would indicate that use of the system improved the quality of the question's distractors. Furthermore, QuesGen was not effective at enticing instructors to use the question quality checklist in a way that would allow them to avoid making technical errors in their questions. The checklist functionality seems to have the right idea, but its implementation in QuesGen was flawed.

Conclusion—Implications for System Design

All of these metrics have the potential to have an immediate impact on the teacher-student relationship and therefore can be used by instructors to evaluate their questions. It is practical to gather most or all of the data needed to report on these metrics via the QuesGen system. Furthermore, it does not take long for instructors to be able to understand and interpret these metrics. The following is a list of implications and goals for the next version of QuesGen that follow on from the discussion of the results above. The use of a participatory design strategy for incorporating these features is strongly indicated.

More obtrusive features

Perhaps "obtrusive" is not the correct word, but a major problem with the system was that the teachers did not use the features that were designed to help them. Given the history of IS

implementations, this is not really a surprising result, but it is disappointing nonetheless and future versions of the tool need to work harder to convince users to take advantage of what it has to offer. In particular, the question templates, given their association with higher-order questions, should be a focus.

"Smarter" incorporation of the checklist

While it is clear that the question quality checklist has lots of useful suggestions for avoiding flaws, not all of them apply to every question. A way needs to be found to prompt the user intelligently to check for flaws. This might mean the incorporation of grammar-checking software and/or natural language processing tools for flagging potential problems. Wisdom on how to go about this might come from the work of other researchers such as Joanna McGrenere's work on "bloat" (McGrenere, 2002).

More responsiveness

While the usability studies done with the tool clearly indicated that it had to contain a WYSIWYG editor for writing questions and incorporating images into questions, when implemented this greatly decreased the responsiveness of the system. The editor also introduced bugs which made the interface basically unusable to at least one instructor. In the months since QuesGen was implemented the state of the art of web-based applications has advanced tremendously, and solutions to the problems encountered here have been developed.

Explicit incorporation of the student surveys and Question Analysis Reports

When presented with their students' feedback on the question analysis reports, nearly all of the participant instructors said that they would really like to be able to get this type of feedback from their students on a regular basis. Originally it was included in QuesGen as a way to generate data for the study, but it seems that it should become a feature in future versions of the tool.

Incorporation into existing systems, i.e. Blackboard or WebCT

QuesGen does have the ability to export questions in the text format needed for both Blackboard and WebCT, but with the new APIs being released by these companies, even tighter integration is possible. It may be possible to replace or supplant the existing tools in these systems with higher-quality ones.

It is hoped to be able to add this functionality in future versions of QuesGen.

REFERENCE LIST

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives.* New York: Longman.

Black, P., & Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan , 80* (2), 139-149.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning: Putting it into practice.* Berkshire, England: Open University Press.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on the Formative and Summative Evaluation of Student Learning.* New York: McGraw Hill.

Haladyna, T., Downing, S. M., & Rodrigues, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education , 15* (3), 309-334.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory.* Newbury Park: Sage Publications.

McGrenere, J. (2002). An Evaluation of a Multiple Interface Design Solution for Bloated Software. *ACM CHI Letters , 4* (1), 163-170.

Stiggins, R. J. (2001). The Unfulfilled Promise of Classroom Assessment. *Educational Measurement: Issues and Practice , 20* (3), 5-15.

APPENDIX A—SAMPLE QUESTION ANALYSIS REPORT

Title: Power Form- Question 3

Objective: Define and identify the 3 forms of power.

Template: Troubleshooting

Question: Diagnose the problem in the following scenario. Joey tells his roomate John, "Clean up this messy room or else." John says "Sure" but proceeds to avoid cleaning up his room even though he had planned on cleaning it up before speaking to Joey. In fact, he makes his room messier on purpose after their conversation.

65% (15/23)	***A. Joey is operating from a dominance perspective and John is opeating from a prevention perspective.***
17% (4/23)	B. John is operating from a dominance perspective and Joey is operating from a prevention perspective.
17% (4/23)	C. Joey and John are both operating from a dominance perspective.
0% (0/23)	D. Joey and John are both operating from a prevention perspective.

Students' Evaluation of the Question

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The question was difficult	13% (2/16)	56% (9/16)	19% (3/16)	13% (2/16)	0% (0/16)
The question was fair	0% (0/16)	0% (0/16)	25% (4/16)	38% (6/16)	38% (6/16)
The wording was clear	6% (1/16)	13% (2/16)	25% (4/16)	25% (4/16)	31% (5/16)

Words students said they didn't know:

Making it messier made it seem like he wanted dominace to prove a point, prevention perspective

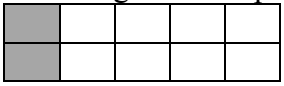
Why students said they missed the question:

The question was confusing: 25% (1/4)
My interpretation was different than the instructor's: 50% (2/4)
Careless error: 25% (1/4)

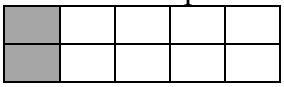
Item-review Panel Evaluation of the Question

The panel rated the degree to which this question assessed understanding of the following educational objectives, which were taken from the lecture slides:

Objective #1: Define power and give an original example of power in their own lives.

Panelist #1: not at all  completely
Panelist #2: not at all completely

Objective #2: Define and identify the 3 forms of power.

Panelist #1: not at all  completely
Panelist #2: not at all completely

Objective #3: Define and identify the 5 types of power resources.

Panelist #1: not at all	■					completely
Panelist #2: not at all	■					completely

Objective #4: Use the terms to label which power forms and resources are present when presented within a communication scenario.

Panelist #1: not at all					■	completely
Panelist #2: not at all					■	completely

The panel agreed that this question adhered to the following best practices:

- The answer marked as the key above is the best answer.
- This question clearly addresses a stated educational objective.
- The concept examples used in this question are novel.
- This question clearly addresses exactly one educational objective.
- This question has exactly one correct answer.
- All of the answer choices for this question are plausible.
- This question's distractors are based upon likely student misconceptions.
- The question addresses salient knowledge.
- The content in this question is NOT based upon opinion.
- This is a NOT trick question.
- This question does NOT depend upon cultural knowledge.
- This question is grammatically correct.
- The vocabulary in this question is appropriate to the student level.
- The answer to this question is not obvious.
- This question follows best practices in the use of the words NOT and EXCEPT.
- This question uses humor appropriately.
- Students can NOT answer this question correctly purely from memory.
- This question avoids using "all" or "none of the above."
- This question avoids using absolutes such as "every" or "never."

The panel agreed that this question violated the following best practices:

- The panel didn't agree that your question violated any of the best practices.

Level of Bloom's Taxonomy targeted by this question:

- Panelist #1: application
- Panelist #2: analysis

Discrimination Index: 0.333333333333