Int. J. Engng Ed. Vol. 20, No. 6, pp. 958–965, 2004
Printed in Great Britain.

0949-149X/91 $3.00+0.00

A Spreadsheet Method for Promoting Learning in a Non-Supervised Testing Environment*

SEBASTIAN A. MAURICE and ROBERT L. DAY
University of Calgary, Calgary, Alberta T2N 1N4, Canada. E-mail: smaurice@ucalgary.ca

We propose a spreadsheet method that promotes learning in a non-supervised testing environment (NSTE). In a survey response from 157 students in an undergraduate university course consisting of 609 students, we present survey results that show our spreadsheet method was effective in helping students learn the course material and helped them prepare for the exams in NSTEs. Our results suggest that the majority of students can learn in NSTEs without being distracted by different forms of cheating.

INTRODUCTION

IN TIMES when class enrollments are rising, university budgets are shrinking, and tuition is increasing, students are busier with activities other than school, such as work and family. For this reason, NSTEs are becoming increasingly critical because they give students the flexibility to be evaluated on their understanding of course material—at their convenience. As a result, NSTEs help to minimize consumption of lecture time and institutional resources.

The main disadvantage of NSTEs is the increased likelihood of inaccurately evaluating student understanding of course material, which can be caused by cheating behavior. Our spreadsheet method provides deterrents to cheating, such as preventing sharing of answers between quiz takers, that could otherwise lead to an inaccurate evaluation of student understanding of course material. By using a Microsoft Excel spreadsheet, and Visual Basic Code (VBC) embedded in the spreadsheet, we are able to generate random quizzes or tests. Specifically, our method allows choices to be a function of the question. Questions are parametric, and choices can utilize any Excel formula, be it constant numbers, text, or a graphic image. The parameters in the questions are allowed to range in a closed interval, [a, b], where a, b, chosen by the quiz maker (QM), are real numbers. The VBC randomly chooses a number between [a, b], inclusive, thus changing the questions, and the choices. The VBC further *shuffles* the choices with the objective of changing the location of the correct answer; the correct answer is chosen by the QM. By utilizing any Excel formula, graphic image, constant number, or text in the questions and choices, makes our method both flexible and effective.

The automation of our method on the Internet enables one to generate any number of random quizzes, in a short time-frame, which can then be distributed to the quiz takers (QTs). Our method also lends itself to automatic marking of quizzes, thus greatly minimizing the administrative effort needed to mark potentially hundreds of unique quizzes for a single class. In NSTEs our spreadsheet method is, we believe, an improvement over giving all students the same quiz or test.

In the Faculty of Engineering, University of Calgary, some of our instructors are using our spreadsheet methods to deliver tests and quizzes to undergraduate engineering students in non-supervised testing environments (NSTE). (Used in ENGG 319: Probability and Statistics for Engineers, and ENGG 317: Mechanics of Solids.) While our system is currently multiple-choice, we feel it presents improvements over existing multiple-choice systems in the way it deters cheating. (It should be obvious that part marks cannot be given on multiple choice tests—while this can be viewed as a disadvantage of our method it is not a major concern. However, we are considering implementing written answers, i.e. from essay type of questions.) Our spreadsheet method allows students to be evaluated on their understanding of course material, while assuring instructors that their students are being *encouraged* to understand the course material leading to a more accurate evaluation of their understanding. Spreadsheets also enable instructors to use complicated mathematical and statistical formulae, and cell referencing in their questions and choices, which is parsed by the Visual Basic Code (VBC) embedded in the spreadsheet, and output as quizzes [1–2].

* Accepted 10 June 2004.

THE SPREADSHEET METHOD: A JUSTIFICATION

The objective of the spreadsheet method is to provide remedies to behavior that could lead to an inaccurate evaluation of students' understanding of course material. This behavior can be categorized as cheating, in NSTEs, through several means: 1) memorization of answers– which can happen if another quiz taker (QT) has taken the same test and shares his/her answers with the QT yet to take the test and, 2) getting help from others as QT takes the test.

The advantages of NSTEs:

- does not take away from scheduled lecture times
- does not require an invigilator
- does not impinge on instructor time
- does not impinge on student time
- is a cost-effective way to evaluate student understanding of course material

The disadvantages of NSTEs:

1. Raises the potential for students to collaborate, and may allow students to do well even if they don't understand the material—therefore, could be inaccurately evaluating student understanding of course material.
2. They could unintentionally penalize students who work individually on the test rather those who work in a group.

In a survey of 285 university students, Maurice and Day [4] found that some students were concerned that the online testing technology (conducted in a NSTE) gave students who worked in a group an advantage over those who worked individually. Given students' usual propensity to work in groups to increase their grade, our method tries to encourage all students to work and understand a problem or question before answering it. This is done by generating *unique* quizzes for all students. For example, as will be explained below, in a class of 600 students, our system can easily generate 600 unique quizzes with different questions and choices from the *same* set of questions without changing the degree of difficulty across quizzes. This means that one does not need a large database of questions, even with a small set of questions, our system can generate 600 different versions of the same question—how this is done is explained below.

The goal of our method is to encourage learning, leading to an accurate evaluation of student understanding of course material, not to prevent students from working in a group. Having unique quizzes for each student, and different, shuffled, choices (answers) to questions, encourages students to re-evaluate every question, and choice set, in every quiz. Thus preventing the sharing of answers between QTs, which would not be the case if all students received the same quiz. In NSTEs, our spreadsheet method is, we believe, an improvement over giving all students the same quiz or test.

The literature has yet to address issues surrounding NSTEs, such as the likelihood of inaccurately evaluating student understanding of course material. The need to evaluate students accurately is one reason why we invigilate exams. NSTEs present several advantages that are slowly becoming apparent to our instructors, as they see class sizes increase along with students' desire to be evaluated on their understanding of the course material. However, the proliferation of cheating behavior in NSTEs has been a concern in our faculty that needed to be addressed.

As many instructors take advantage of NSTEs, our spreadsheet method addresses the disadvantages. While we cannot prevent students from cheating, we can encourage them to learn by devising tests and quizzes that encourage students to *work* and understand a problem. Students should be evaluated accurately on their understanding of course material [5]. To take an excerpt from a national parenting magazine [3, pp. 1–3] (with the underlying assumption that students are being evaluated accurately):

> Just as a doctor uses a test to diagnose a medical condition, schools can use tests to pinpoint problems. And when a child's strengths and weaknesses are revealed in a test, teachers and schools can make the necessary changes to be more effective. Tests can be a critical tool to track the progress of student, curriculum, and school.

THE SPREADSHEET METHOD AND IMPLEMENTATION

Instructors enter their quiz questions and choices in a spreadsheet template—call these people the quiz makers (QM). The test or quiz is then delivered to, and taken by, students in their course—call them the quiz takers (QT). (It is beyond the scope of this paper to go into the details of the VBC embedded in the template. But since this template represents the core of our method, interested parties can contact the authors of this paper for further details on its exact operations and functions. The authors would be happy to help those thinking of implementing our method at their institution.)

One can define the spreadsheet method as follows. Each spreadsheet contains a set of *parametric* questions, **Q**. Each question in **Q** has parameters **x**, and associated choices **c**, where:

$\mathbf{Q} = \{(\{q^1\}, \ldots, \{q^n\}) \mid \{q^i\}$ is quiz question i made up of text and parameters $\mathbf{x}^i$, $i = 1 \ldots n\}$ (A)

$\mathbf{x}^i = \{(x_1, \ldots, x_n) \mid x_j + s_j \in [\mathrm{a}_j, \mathrm{b}_j]$, s_j is a step variable where: $s_j \leq \mathrm{b}_j - \mathrm{a}_j$, or $x_j \in \mathbf{C}^j$, or $x_j \in$ {JPG, GIF}, $j = 1 \ldots n$, $\mathbf{C}^j$ is some set j containing a singleton, $\mathrm{a}_j \leq \mathrm{b}_j$, JPG and GIF are popular Internet image formats, and $\mathbf{C}^j \subseteq \mathbf{R}$, a, b $\in \mathbf{R}$, $\mathbf{R}$ is the set of real numbers$\}$ (B)

$\mathbf{c}^i = \{(\{c_1\}, \ldots, \{c_n\}) \mid \{c_j\} \subseteq \mathbf{F}$, or $\{c_j\} \subseteq \mathbf{C}^j$, or

$\{c_j\} \subseteq$ {JPG, GIF}, or $\{c_j\}$ is text, $j = 1 \ldots n$, $\mathbf{F}$ is a set of all formulas in the Excel application or a formula defined by the QM, $\mathbf{C}^j$ is some set j containing a singleton, JPG and GIF are popular Internet image formats, and $\mathbf{C}^j \subseteq \mathbf{R}\}$ (C)

the index i on $\mathbf{x}^i$ and $\mathbf{c}^i$ indicates that $\mathbf{x}$ and $\mathbf{c}$ belong to quiz question i; furthermore: $\mathbf{x}^{ia} \subseteq \{c_j\} \in \mathbf{c}^i$ or $\mathbf{x}^{ia} \not\subset \{c_j\} \in \mathbf{c}^i$ or $\mathbf{x}^{ia} \subseteq \{c_j\} \subset \mathbf{c}^k$, $i \neq k$, $k = 1, \ldots, n$, and $\mathbf{x}^i \subseteq \{q^i\} \in \mathbf{Q}$, or $\mathbf{x}^{ia} \subseteq \{q^i\} \in \mathbf{Q}$ (D)

We can represent all chosen, or actual, values of $x_j \in \mathbf{x}^i$ as x_j^a, such that $x_j^a \in \mathbf{x}^{ia}$. The set definition in (A) says that $\mathbf{Q}$ is a set of questions in the quiz, with each question written in text and containing a $\mathbf{x}$. The set definition in (B) says that elements in $\mathbf{x}^i$ must satisfy $x_j + s_j$ in the closed interval $[a_j, b_j]$, or x_j is a constant number, or x_j is a graphic image. Similarly for (C), $\mathbf{c}^i$ requires that $\{c_j\}$ be any formula in the Excel application or a formula defined by the QM, or $\{c_j\}$ is a constant, or $\{c_j\}$is a graphic image, or $\{c_j\}$ is text. Lastly, (D) indicates that for a given $\{q^i\}$, all elements in $\mathbf{x}^{ia}$ may be referenced in each choice $\{c_j\} \in \mathbf{c}^i$, or *not*, and $\mathbf{c}^i$ can contain other $\mathbf{x}^a$; all parameters in $\mathbf{x}^i$ must appear in the question, $\{q^i\}$, and all parameters in $\{q^i\}$ must be replaced by the actual values of $\mathbf{x}^i$, represented by $\mathbf{x}^{ia}$. For $\mathbf{Q}$ to be a parametric question, it must be that for any$\{q^i\} \in \mathbf{Q}$, $\{q^i\}$ contains all $x_j \in \mathbf{x}^i$, which range in $[a_j, b_j]$. Choices can be any formula in the spreadsheet and can refer to any cell, such as the cells containing $\mathbf{x}^{ia}$. Since each choice set, $\mathbf{c}^i$, contains the correct answer to a $\{q^i\} \in \mathbf{Q}$, the spreadsheet application will randomly shuffle the choice set $\mathbf{c}^i$, hence, *potentially*, re-locating the correct answer in $\mathbf{c}^i$. (Given random variables, there is the potential for questions in different quizzes being the same. The probability of this happening is negatively related to the size of the closed interval [a,b]: the larger the interval, the less likely we are to have the same question in different quizzes. Therefore, the larger the interval [a,b] the higher the degree of uniqueness between quizzes.) It should be noted that the QM determines $\mathbf{Q}$, [a, b], $\mathbf{x}$, and $\mathbf{c}$. The VBC determines $\mathbf{x}^a$ and shuffles $\mathbf{c}$. Therefore, our spreadsheet program allows:

$$\text{Choices} = f(\text{Questions}) \tag{1}$$

Equation 1 says that choices are a function of the questions. Since $\{q^i\}$ contains random variables, it is itself a random variable, as are the choices—if they reference cells containing $\mathbf{x}^{ia}$. By having choices, hence the correct answer, as a function of the question, the VBC in the spreadsheet *is* creating different questions and choices for different QTs.

To summarize, questions are parametric, and choices can utilize any Excel formula or be constant numbers or text or a graphic image. Choices can reference *any* cell in the spreadsheet. The parameter(s) in the questions are allowed to range in the closed interval [a, b]. The VBC randomly chooses a number in [a, b], inclusive, causing changes in the questions and choices. The VBC further shuffles the choices with the objective of changing the location of the correct answer; the correct answer is chosen by the QM. (All choices are chosen by the QM. Specifically, the QM specifies the wrong choices and the correct choice.) The use of any type of Excel formula, graphic image, constant number, or text in the spreadsheet, makes our method both flexible and effective. How this whole process is operationalized in a dynamic and random manner is explained next.

It is the role of the VBC, embedded in the spreadsheet, to automatically carry out the following critical functions. These functions encapsulate the discussion above, for each $\{q^i\} \in \mathbf{Q}$:

a) VBC randomly chooses all x_j ranging over $[a_j, b_j]$ in $\mathbf{x}^i$—resulting in $\mathbf{x}^{ia}$
b) replaces $\mathbf{x}^i$ in $\{q^i\}$ with $\mathbf{x}^{ia}$ in (a)
c) recomputes all formulas in $\mathbf{c}^i$
d) shuffles the set $\mathbf{c}^i$
e) writes out $\{q^i\}$ and $\mathbf{c}^i$ to a quiz file.

Different quizzes can be created by iterating through a) to e) for each question in the quiz, and for each QT. The algorithm used by the VBC, shown in a) to e), can be demonstrated in the pseudo-code shown in Fig. 1.

After the run in Fig. 1, the QM distributes the quizzes to his/her students. The goals for our method are the following. It should:

1. Discourage students from memorizing answers to questions—this can happen if another QT has taken the same test and shares his/her answers with the QT yet to take the test. **Remedy:** the shuffling of the choice set $\mathbf{c}$ helps to minimize this behavior, because the answer location in $\mathbf{c}$ is likely to be different –for the same question.
2. Encourage students to understand, hence learn, the material before answering the question, even if being helped by a colleague. **Remedy:** since the questions, and choices, in each quiz, are likely to be different, a QT may not be able to use the same answer value that his colleague used. If the question requires computations, the QT must still perform these computations to obtain the correct answer. (However, if the QT decides to guess at the answer, our spreadsheet method cannot prevent this from happening,

```
For each QT
    For each {q} in Q
        For each x in {q}
            DO a)– d)
        End For
    DO e)
    End For
End For
```

Fig. 1. Pseudo-code for generating quizzes for QTs.

	A	B	C	D	E	F	G	H
34	**Begin Question**	**Input Variable**	**Min**	**Max**	**Step**	**Actual**	**Choices**	**Correct Answer**
35	The probability of developing a knee injury is dependent on a person's involvement in high	h	0.45	0.75	0.05	0.5	0.5875	
36	impact sports, with persons involved in these sports having a [h] chance of developing a	n	0.05	0.35	0.05	0.25	0.0875	
37	knee injury. Otherwise, the probability drops down to [n]. If a survey of the province indicates	p	0.25	0.75	0.05	0.65	0.7500	
38	that [p] of the population participates in high impact sports, what is the probability that a						0.2625	*
39	randomly selected individual will not have any knee injury and not participate in high impact							
40	sports?							

Fig. 2. Example of a parametric question, in raw form, in a spreadsheet.

nor can it prevent a student from using someone else to take the test for him/her. Given the uniqueness of each test, using someone else to take the test on the student's behalf has been made more difficult and time consuming by our method and acts as a discouraging factor.) While students can still count on luck to choose the right answer—this can be counterbalanced by presenting several questions to the QT—the probability of getting a good grade based purely on luck gets lower as more questions (and more choices) are presented to the QT.

Note that the degree of difficulty of questions is not affected by our method because we are not changing the intent or wording of a question, only the numerical values contained in the questions. Figure 2 shows an actual parametric question in a spreadsheet.

The parameters, **x**, in the question, shown in the **Begin Question** column (cells A35-A40), are indicated by [h], [n], [p], and listed in the **Input Variable** column, along with the **Min** and **Max**. The **Step** variable tells the VBC to randomly increase [h], [n], and [p] by their step value. The choices are listed in the **Choices** column and can use any Excel formula and reference any cell—preferably cells in the **Actual** column. The **Actual** column is where the VBC will place the randomly chosen x_j—this is the x_j^a, or $\mathbf{x}^a$. The QM indicates the correct answer to a question by an asterisk, shown in the **Correct Answer** column, i.e. H38. To state this more clearly—using the cell references in Fig. 2. Each parameter, [h], [n], and [p], will be replaced by the actual values of h, n, and p: 0.5 (cell F35), 0.25 (cell F36), and 0.65 (cell F37), respectively. The VBC will also shuffle all choices in the **Choices** column. When it is time for the QM to distribute this quiz, the QT will see only the question with the parameters [h], [n], and [p], replaced by their actual values, and the shuffled choices. Fig. 3 shows the resulting question in Fig. 2, after the code in Fig. 1 has run.

The probability of developing a knee injury is dependent on a person's involvement in high impact sports, with persons involved in these sports having a 0.5 chance of developing a knee injury. Otherwise, the probability drops down to 0.25. If a survey of the province indicates that 0.65 of the population participates in high impact sports, what is the probability that a randomly selected individual will not have any knee injury and not participate in high impact sports?

a) 0.0875
b) 0.5875
c) 0.2625
d) 0.7500

Fig. 3. Resulting question, shown in raw form in Fig. 2, after the run in Fig. 1.

The VBC handles everything in the 'background' transparent to the QTs and the QM. If choices reference a randomly chosen variable, the choices will also be random. The VBC does not automatically choose the wrong choices because the QMs wanted to *strategically* choose wrong choices. For example, the choices in the spreadsheet were entered into the spreadsheet by the QM as follows:

= (1-F35)*F37+(1-F36)*(1-F37)
(when executed outputs 0.5875)

= F36*(1-F37)
(when executed outputs 0.0875)

= 1-F36
(when executed outputs 0.7500)

= (1-F36)*(1-F37)
(when executed outputs 0.2625)

The VBC automatically instructs Excel to execute the above choices (or functions). Figure 4, below, shows the schematic of the spreadsheet process. QMs follow steps **i-v** when generating quizzes for their class.

Figure 4 can be explained as follows:

- **Step i:** Instructors log into the test generating website, and download the spreadsheet test template, on their computer, and enter their quiz questions and choices, indicating the correct choice for each question.
- **Step ii:** Once they finish entering their quiz, they upload their quiz back to the test generating website.
- **Step iii:** They press a button on the website to generate quizzes or they exit the website.
- **Step iv:** If they do not want to generate quizzes, they exit the website.
- **Step v:** If user wants to generate quizzes, the request is sent to the web server which will then execute the VBC (shown in Fig. 1) in the spreadsheet test template and output the quizzes.

The generated quizzes are then registered on the Generated Quizzes Website for instructors to

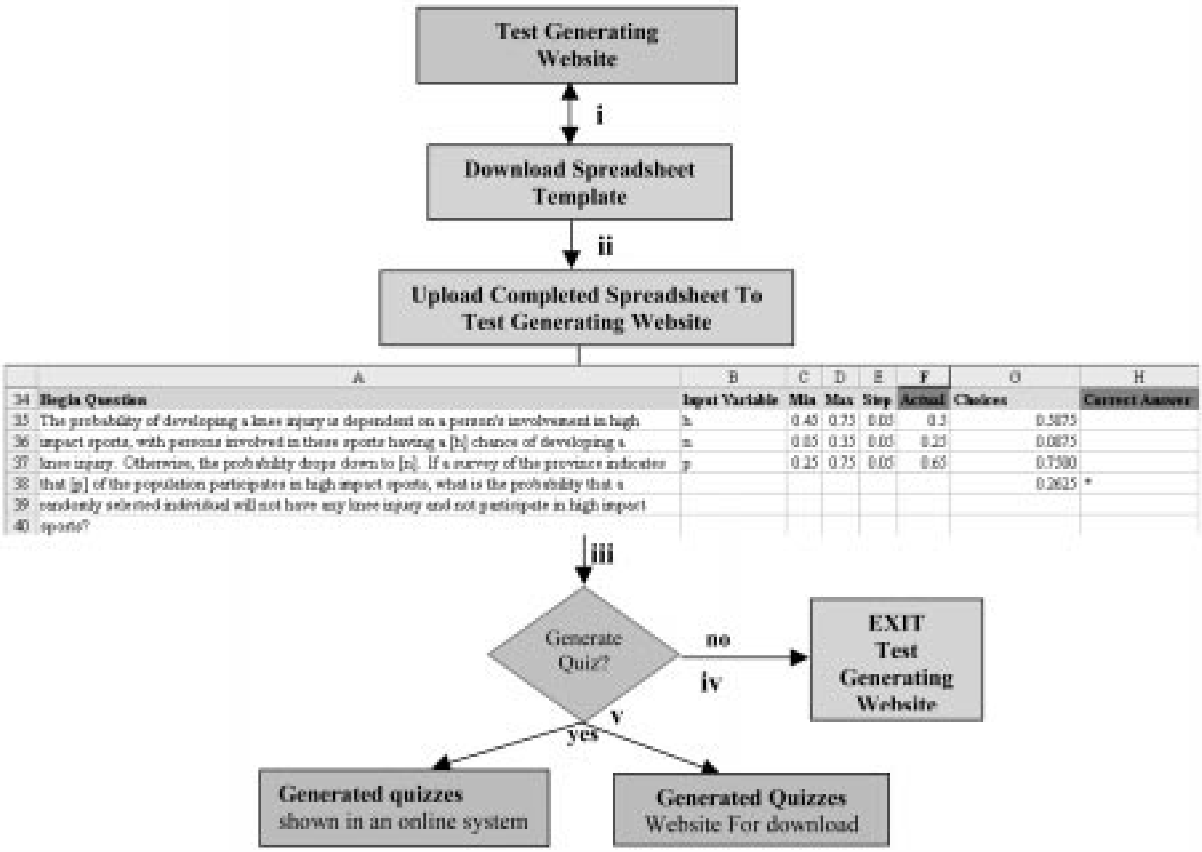

Fig. 4. Spreadsheet process and implementation design.

download or print. The quizzes are then distributed to their students. Or, the spreadsheet method can be implemented in an online system allowing students to do their quizzes online removing the need to distribute quizzes to each and every student manually. (Readers are encouraged to contact the authors on details on how the spreadsheet method can be implemented in a user-friendly online system.)

Associated with each generated quiz is a personal identification number (PIN) that uniquely identifies each quiz and associated answer key. (The answer key is generated for each quiz. The PIN of the quiz is equal to the PIN of the answer key, facilitating the marking of quizzes by matching the answers in the answer key to those provided by the student for a particular question in a particular quiz. While some data entry is required, i.e. entering students' answers, the time and effort it takes to mark a quiz can be greatly minimized by automation.)

Since the VBC will randomly generate quizzes and choices, and shuffle the choice set, along with the correct answer, it becomes important that correcting quizzes does not require a massive administrative effort, especially for large classes. By identifying each quiz by its PIN, and associated answer key, our spreadsheet method easily lends itself to automatic marking.

Figure 4 further illustrates that our spreadsheet method can be implemented on a website and all its major functions automated requiring very little effort to use. With the addition of automatic marking of quizzes, the benefits from using our method/process far outweigh the costs.

FIELD STUDY

Our spreadsheet method was implemented in a second-year undergraduate course in the University of Calgary. Total enrollment in the course was 609 students. The course was called ENGG 205: Mechanics I. The instructors in ENGG 205 presented all students two types of quizzes or assignments. One was our online spreadsheet method, and the other was a normal paper assignment. Students were given each type of assignment on a two-week rotating basis. So first two-weeks they were asked to complete the online spreadsheet assignment, and the next two-weeks they were given the paper assignment. Each type had to be completed in a week.

We implemented the spreadsheet method in an online web platform described in [4], therefore the 'Generated Quizzes' in Fig. 4 was sent directly to a website where a student could complete this quiz or assignment online. The students were given details on how to do this by their instructors with the appropriate supporting material. We surveyed 609 students in ENGG 205 using the questionnaire shown in the Appendix. Total

Table 1. Average survey results from 157 students

Question	Variable	Average Response	Confidence Interval For Mean (95%)	Variance
Ease of use of the quiz system	EU_1	4.05	(3.89, 4.21)	0.94
The on-line quiz helped me learn the course material	LEARN	3.57	(3.39, 3.75)	1.19
The on-line quiz helped me prepare for the type of examination questions on the mid-term and final examination.	PREPARE	2.93	(2.74, 3.12)	1.42
How does the online Quiz system compare to in-class paper quizzes? Keeping in mind issues such as flexibility of writing the Quiz at your leisure, getting your Quiz grade instantly, receiving an instant email detailing the questions you got wrong/correct, etc.	COMPARE	3.74	(3.55, 3.93)	1.40

survey responses received was 157. Table 1 summarizes the results gathered. (The COMPARE variable picks up the students' like or dislike towards the technology. If students choose 1, then the system compares poorly to paper quizzes, if they choose 5, it is more superior to in-class tests.)

As can be seen in Table 1, many students found that the online system, using the spreadsheet method, did help them learn the course material with a result of 3.57 out of 5. Also, majority of students felt that the online system helped them to prepare for the exams with a result of 2.93 out of 5. Many felt that the ease of use of the system was good with a result of 4.05 out of 5. And many felt the online quiz system was better than in-class paper quizzes with a result of 3.74 out of 5.

There are several things to take away from the results in Table 1. First, the above results show that our spreadsheet method can be implemented in an online system with a user interface. If done correctly, the spreadsheet method can be effective in testing students online in a user-friendly way. Second, our method did help students learn and prepare for the exams. Helping students learn and prepare for exams is one of the main objectives of the spreadsheet method and it was reassuring to see students felt that it was doing that. Thirdly, students preferred the online method to the paper method. While we don't want to generalize beyond our sample, the students in ENGG 205 can be viewed as representative of many other second-year students in other universities; there was nothing special about our sample. However, more research needs to be done to see if our spreadsheet method encourages learning and helps students prepare for exams in a broader context.

COMPARISON OF TESTING METHODOLOGIES

This section examines how our method compares to other testing methodologies. Several popular testing applications offered by companies such as Blackboard Inc. and Education Testing Service Inc. use a 'silo' approach common in Computer Adaptive Testing (CAT). (Blackboard (www.blackboard.com) offers a web-based portal with a testing component. It is implemented in the University of Calgary. ETS (www.ets.org) is a large testing organization that adjudicates several major tests such as the GRE, GMAT, SAT, TOEFL, etc.)

The silo approach [6] is a database that contains a library of questions that can be chosen at random by a computer program. This program can choose, at random, a subset of questions from this database. The questions are processed using a Kuder-Richardson formula (KR20) or coefficient alpha [6]. (This test reliability statistic measures inter-item consistency. A high KR20 value indicates a strong relationship between items on the test. A low value indicates a weak relationship between items on the test.)

A SIBTEST methodology [8] is bundles test items into meaningful and statistically dimensionally distinct categories. The purpose of this categorization is to show whether ethnic or gender differences exist in test performance [8]. Therefore, the SIBTEST bundle method can be applied to test equity evaluation and future test development in content domains other than mathematics [8].

Item modeling is a term used to refer to generating tests on-the-fly [7]. Specifically, Bejar *et al.* state that the goals of item modeling is to reduce the exposure, or reduce the frequency, of certain items (questions) so that its security is not compromised. In other words, an item model is simply a procedure for creating or generating tests of similar or comparable content that are exchangeable psychometrically [7]. Item modeling is a construct driven approach that understands the goals of the assessment. Tests are generated on-the-fly from an item pool that contains the item models and the items; instances of the model are presented to the test taker at delivery.

The above methods are not dependent on a spreadsheet and do not parameterize the questions or choices while our approach does. Our method is simply a different approach to testing. The above methods are useful mainly when test questions are text based. Our method is most useful when questions are mathematical in nature. Specifically, because we employ a spreadsheet, our method allows instructors to leverage the mathematical (and statistical) power of the spreadsheet application. (The added advantage of this is that we don't need to write complex mathematical parsing code—the spreadsheet does the parsing and computations for us.) Moreover, our method can generate different questions (and answer choices) by simply changing the numbers in the formula or cells—embedded in the questions or choices—from any interval specified by the test maker.

The power of our approach is in part due to the coupling of our method with the spreadsheet. The spreadsheet gives our method the flexibility of choosing numbers in any cells as input into the questions and answer choices. The parametric nature of the questions and the answer choices reduces the need to have a large database of questions while maintaining an equal degree of difficulty between generated questions and answer choices. While the above methods have value in their own respect, they fail to address the issues surrounding questions (and choices) that are more mathematical and statistical (and more complex) in nature while maintaining a high degree of uniqueness and equity in difficulty between tests. We feel that parameterizing questions and choices minimizes this difficulty because it does on modify the formulas only the inputs. (These inputs can be chosen randomly by our method: it can be a constant value or a value in an interval.) Parameterization, as implemented in our approach, is an easy and cost effective way to generate unique tests, that may be mathematically or statistically complex, on-the-fly. As we have argued in this paper, the uniqueness between tests in a NSTE is important to encourage students to be evaluated effectively and fairly while maintaining the integrity of a testing environment.

DISCUSSION AND CONCLUSION

Very little research has been done on the affects of NSTEs on the learning process of students. This paper has highlighted some of the issues surrounding NSTEs and proposes a method to address these issues. Our spreadsheet method has the potential to enhance the integrity of a NSTE by encouraging students to learn and understand the course material by creating *unique* quizzes for each student in an effort to deter cheating behavior. It is a method that promotes learning but cannot completely eliminate the possibility of evaluating students incorrectly especially when students guess at answers, and/or use someone else to write a test on their behalf. However, we feel our method has an advantage, and is an improvement over the alternative of providing the same test or quizzes to all students in NSTEs. There is an imputed benefit to students and instructors from our method that cannot be had using the alternative. Our method is both flexible and effective requiring minimal effort to use once implemented. As shown in the field study, many students indicated that the spreadsheet method helped them learn the course material and helped them to prepare for the exams. Our method was also easy to use and preferred over paper quizzes.

Further research should focus on other methods used to encourage learning in NSTEs. Research should analyze the negative implications of inaccurate student' evaluations in NSTEs. More research should be done, generally, into the differences between STEs and NSTEs, such as effects of time constraints, and the pressure of being monitored in a STE, on student performance. Our method could be used as a basis for further research to see how grades between students in STE and NSTEs differ by way of an experiment.

Our method may help other institutions dealing with similar issues, especially in these times when institutions need to balance resources, with larger class sizes, busier students, faculty, and staff, with students' desire to be evaluated on their understanding of course material. Accurate evaluation of students using tests is critical to track the progress of student, curriculum, and schools. We hope this paper will motivate more research into issues surrounding NSTEs.

Acknowledgments—Authors wish to thank the two anonymous referees for their thoughtful comments.

REFERENCES

1. R. Beare, A Spreadsheet system for educational use based on Microsoft Excel, *Australian Educational Computing*, July 1993.
2. P. Blayney, Use of a spreadsheet-based marking system for assisted learning and assessment, *Proc. 12th Annual Conference of ASCILITE*, December 1995, pp. 34–40.
3. C. Adams, Support for testing remains strong amidst debate, *Parent Power*, **3**(3), April 2001.
4. S. A. Maurice and R. L. Day Online Testing Technology: Important Lessons Learned, *Int. J. Eng. Educ.*, **20**(1), 2004.
5. N. Kober, Teaching to the test: the good, the bad, and who's responsible, *TestTalk for Leaders*, Center on Education Policy, Issue 1 (June 2001), pp. 1–12.
6. C. Norris and R. Lajoie, *Meeting the Challenges of Online Testing; Reliability, Scalability and Security*, SSGRR, 2003.

7. I. I. Bejar, R. R. Lawless, M. E. Morley, M. E. Wagner, R. E. Bennett and J. Revuelta, *A Feasibility Study of On-the-Fly Item Generation in Adaptive Testing*, GRE Board Professional Report No. 98-12P, ETS Research Report 02-23 (2002).
8. W. Stout, D. Bolt, A. G. Froelich, B. Habing, S. Hartz and L. Roussos (2003), *Development of a SIBTEST Bundle Methodology for Improving Test Equity With Applications for GRE Test Development*, GRE Board Professional Report No. 98-15P, ETS Research Report 03-06.

APPENDIX

Survey questionnaire presented to students in electronic form

On a scale from 1–5, 5 being the best, please comment about the following:

1. Ease of accessing the online system

1 2 3 4 5

2. Ease of use of the online system

1 2 3 4 5

3. The online assignments helped me learn the course material

1 2 3 4 5

4. The online assignments helped me prepare for the type of examination questions on the mid-term.

1 2 3 4 5

5. Please provide comments below specifically about how you think access and ease of use could be improved:

Sebastian Maurice is a research member of the laboratory for Software Engineering Decision Support at the University of Calgary. His general area of research is value-based software engineering. He has several academic publications in International Engineering journals and active in the area of online testing technology. He has also received several fellowships for his research initiatives. He is the founder and chair of the Software Engineering Consulting Consortium at the University of Calgary and has over 10 years of combined professional experience as a researcher, software developer and project manager. He is also a member of the Project Management Institute. Sebastian has a bachelor of social science degree in Economics from the University of Ottawa (1993), a bachelor of science degree in pure mathematics from the University of Calgary (1997), and a Master of Science degree in Agricultural Economics from the University of Alberta (1997), and a Master of Science in Software Engineering. Sebastian is currently an IS Project Manager at EPCOR Utilities Inc. in Calgary, Canada.

Robert Day is Associate Dean, Planning, Faculty of Engineering and Professor of Civil Engineering. In the Faculty of approximately 2500 undergraduate students offering ninr engineering degree programs, he was Associate Dean, Academic for 10 years and Associate Dean, Student Affairs for four years prior to assuming his current position. He has led several major curriculum and accreditation initiatives, including the major redesign of the first two years of all programs and the recent establishment of a full year hands-on design experience for all 600+ first-year students.