A Test Module for MOARS (the Mobile Audience Response System)

William R. Pellowe
ペロウ　W.

This was published in 2012 in *Reports of Faculty of Humanity-Oriented Science and Engineering, Kinki University*, Issue 17, pp. 54-58.

Abstract: This paper describes the in-progress development of a Testing Module for an open-source audience response system called MOARS. This module allows teachers to collect multiple choice test responses from students using mobile devices such as iPod Touch. The module includes test item analysis features for classroom teachers.

1. Introduction

The past few decades have seen a great increase in the classroom use of audience response systems in which the students send answers electronically to the teacher's computer through handheld devices (see Banks, 2006 for a comprehensive overview). To provide classroom teachers with an affordable alternative, I developed MOARS (Mobile Audience Response System), a web-based system to use with iPod Touch or other mobile devices (Pellowe, 2010; Pellowe et al, 2010).

MOARS was designed to allow add-on modules. Current projects include the Peer Assessment module (Holster & Pellowe, 2011) and the Extensive Reading Survey module (Holster & Pellowe, 2012). In this article, I introduce a new project, the Testing module.

2. Overview

The test is on paper. Students use iPod Touch devices to submit their answers. The screen shows one question at a time, with multiple choice answers. The results are available to the teacher immediately.

3. Rationale

A good analysis of test results can help the teacher address any problems in the test. However, this type of analysis can be very time consuming, involving encoding the data into spreadsheets, and either exporting it to statistics software, or running various equations within the spreadsheet software. For this reason, the Testing Module includes some basic item analysis and statistics to help the classroom teacher. This provides the teacher with instant feedback on how well the test is performing.

Furthermore, teachers who use MOARS in their regular classes as a quiz and/or survey tool are already very familiar with the procedures, and their students are very familiar with using a mobile device (such as an iPod Touch or other type of tablet device) in class. With such a system already in place, it is an obvious next step to use it to make the scoring of tests easier on the teacher. Using an automatic system to score multiple-choice tests frees up some of the teacher's time for more qualitative grading of other types of assessment (such as writing or speaking).

4. Process: Students

Students access the MOARS student page through their iPod Touch web browser. The page is a simple log in form. The students log in. Students access the MOARS activities with numerical shortcut codes. The teacher tells the students the shortcut code for the test, and students submit that number to begin their test.

Each question has its own page. The student selects an answer (up to five options) and clicks the "submit" button. This brings the student to a page with the next question. This continues until the test is finished.

5. Process: Teachers

The Testing module page for teachers has three options:

- Make a test
- List all tests
- See test results

When teachers select "Make a test", they use a pull-down menu to input the number of questions on the test (1 to 150). The "Make a test" page has input fields for the test title and teacher notes. The question and answer option fields are pre-populated with question text (consisting only of a question number) and option text (consisting of the letters A, B, C, D and E). Any of these fields can be changed by the teacher. Teachers must select the option that is the correct answer. Any option left blank will not appear on the student screen.

"List all tests" allows the teacher to see all of the tests that are currently in the system. The teacher can see the tests, and make editing changes if needed.

"See test results" gives teachers a list of all of the tests that students have taken. The teacher selects which class results to see. Teachers can pool students from different classes together if desired.

6. Test Results Page

On the results page, the first table contains the individual students' results. The test results are ordered from highest-scoring student to lowest-scoring. For each question, a "1" means "correctly answered". A "0" means "incorrectly answered". To see what time a student answered a particular question, the teacher can hover the mouse over the question result (the 0 or the 1) to see a timestamp (in the format year-month-day hour-minute-second). At the bottom of the table are values for *difficulty* and *discrimination*.

The difficulty index (p) is a simple ratio of the number of correct answers over the total number of answers.

$$p = \frac{N_p}{N}$$

Where N_p is the number of correct answers, and N is the total number of answers. The lower the value, the more difficult the question. Values range from 0.0 (very difficult) to 1.0 (very easy).

The discrimination index (D) tells us how well a test question differentiates between the top-performing students and the bottom-performing students.

$$D = \frac{U_p - L_p}{U}$$

U_p is the number of test takers in the upper group who got the test item correct, and L_p is the number of test takers in the lower group who got the test item correct, and U is the total number of test takers in the upper group (which is of course equal to the total number of test takers in the lower group). The recommended size of these groups varies, depending on the context, from 25% of the total to as high as 50%; for the Testing Module, I chose to take the top third and bottom third of the results. Values for D can theoretically go as low as -1 (which means that all of the upper group students got the answer wrong, while all of the lower group students got the answer correct), but any value below 0.0 is problematic, because it means that fewer of the top scorers and more low scorers are getting that question correct. The highest value (1.0) means that the top group all got the right answer on that question, and all of the low group got the wrong answer. A value of 0.0 means that the same number of students in both groups got the question right. If the question is just as easy or just as difficult for both the top and bottom groups, then the question does not help us to differentiate between the two groups at all, because it's either too easy for everyone (1.0) or too difficult for everyone (0.0).

The difficulty index and discrimination index values are also included in the "Distribution of Answer Choices" table in the teacher's test results page. In this table, the teacher can see how many students chose each correct answer as well as each distractor. If a distractor ("wrong" choice) attracted 0 answers, then perhaps it is too obviously wrong. However, if a distractor attracted an extremely large number of answers, then perhaps it is too confusingly similar to the correct answer.

Table 1 contains data from a pilot test of five classes taught by three teachers with a common exam; I have selected only a few of the 100 questions to display here in order to illustrate how the web page functions.

Question	Difficulty (p)	Discrimin-ation (D)	Status	Answer	A	B	C	D	E
9	.84	0.1	OK	B	17	122	1	6	n/a
21	.99	0	review	C	1	0	142	0	n/a
31	.55	0.4	OK	B	18	80	22	26	n/a
32	.45	0.2	review	C	58	9	65	14	n/a

Table 1: Selected rows from the "Distribution of Answer Choices" Table

The status is either "OK" or "review". A question with "review" is one that the teacher may need to take a look at. Obviously, question 21 is either too easy, or it is about something that all of the students were able to learn perfectly, hence the "review" label. The criteria used to determine these statuses is given in Table 2.

p=Difficulty D=Discrimination	Difficult $p \leq .3$	Medium $p > .3$ and $p < .8$	Easy $p \geq .8$
$D \leq 0.0$	review	review	review
$D > 0.0$ and $D < 0.3$	OK	review	OK
$D \geq 0.3$	OK	OK	OK

Table 2: Criteria to determine whether to flag a test item for reivew

These criteria are of course very context-specific, but it seems fine for classroom tests. The only one that may need to be explained further is the middle block (where p is between .3 and .8, and D is between 0.0 and 0.3). See question 32 as an example. This question of medium difficulty (p=.45), but there is not much difference between the high group and the low group (D = 0.2). The teacher should review this question, because something about the question is nearly equally confusing to both the high group and the low group. The distractor data gives us more information; we can see that nearly as many students chose "A" (wrong) as chose "C" (correct). Both "A" and "C" may be acceptable answers, or they may be close enough that students' choices are random.

The next table on the teacher's page contains a summary of the test results. The data below (Table 3) is from the above-mentioned test.

Description	Value
Total tests:	146
Mode:	79% (10 students)
Median:	76.0%

Description	Value
Mean:	74.260%
Standard Deviation:	10.347

Table 3: Summary table from Test Results page

The "Total tests" is of course the number of students took the test. The "mode" could be explained as "the most popular" score. In this case, 10 students achieved 79%. The "median" is the middle score, the one that separates the higher half of scores from the lower half of scores. The "mean" is better known as the average of all of the scores.

The formula I used for Standard Deviation is as follows:

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2},$$

Where N is the number of students, x is a score, x bar is the mean. (This formula image is from Wikipedia, 2012.) I "translated" this into PHP thus (a double slash indicates a comment):

```
// Each student's raw score was stored in an array called "studentRawTotal",
// with the array key corresponding to their student ID.
// Walk the array:
foreach($studentRawTotal as $key=>$value) {
    ...
    // Store the score as a percent in the array "calc_standard_deviation"
    $calc_standard_deviation[] = ($value / $testNumberQuestions)*100;
    ...
}
// Initialize variable
$total_square_result_minus_mean = 0;
// walk the "calc_standard_deviation array:
foreach($calc_standard_deviation as $key =>$value) {
    // Do the score minus mean squared summation:
    $total_square_result_minus_mean = $total_square_result_minus_mean
          + (($value - $mean_value)*($value - $mean_value));
}
// Divide by the population minus 1 and take the square root:
$standard_deviation = sqrt($total_square_result_minus_mean
          / ($totalStudents - 1));
```

The next table on the teacher page contains a simple summary of the raw score range. Each range group is one tenth of the possible raw score. The "%" column contains that range's lowest possible percentage. Each "*" represents one student, so that teachers can visualize the data distribution. In the following table, the Raw Score and the percent are equal because this particular test had 100 questions.

Raw Score	%	Occur-ances	* = 1 student
90+	90%+	5	*****
80+	80%+	43	***
70+	70%+	55	***
60+	60%+	33	*********************************
50+	50%+	6	******

Raw Score	%	Occur-ances	* = 1 student
40+	40%+	3	***
30+	30%+	1	*

7. Future Developments

Teachers using this for tests need to be aware that many of the newer devices are capable of multi-tasking, so students may have access to additional sources of information (such as dictionaries, websites, personal communication) through the devices. However, devices such as iPad 2 have a "kiosk" feature that allows teachers to lock the students' devices onto a particular application; for seriously high-stakes tests, this type of security would be a minimum requirement. For classroom tests, though, this is not too problematic, as long as the teacher remains just a bit more vigilant than during a regular test.

One feature that is missing from the student side is an easy way to return to previous questions in order to change answers. On iPod Touch 3, such as we use at Kinki University, the student may use the "back" button on the browser to return to a previous page. However, attempting this on older iPod Touch (such as the iPod Touch 1, which students at another university used during the pilot test described herein) will actually return the student to the log-in page, forcing them to re-enter their test answers.

On the teacher side, the "Distribution of Answer Choices" would be helpful to teachers if it have more detail. For example, in question 32, we saw that 58 out of 146 students chose "A", the wrong answer; it would be helpful to know how many of those 58 were in the top third of the test scores, and how many were in the bottom third. Such information would alert us to potential problems -- if a disproportionate number of top students were attracted to a particular distractor, then something about that distractor is actually causing the better students to receive lower scores than they would have otherwise. Such information would alert the teacher that the item required some investigation.

Also, teachers wanting to do serious test analysis would appreciate direct access to all of the data, in order to use it with established statistical analysis software. For those teachers and researchers, an "export data" function is required. Such export functions have been included in other add-ons to MOARS; adding it here would be just as straightforward.

After these improvements have been made, and after some more beta testing, the Testing Module will be released on moars.com.

References

Banks, D.A. (2006) *Audience Response Systems in Higher Education*. Information Science Publishing.

Holster, T., W. Pellowe (2012). "Using a Mobile Audience Response System to Monitor Extensive Reading." *The 5th Annual ER Seminar: Extensive Reading: Research and Practice*. Sugiyama Jogakuen University, July 1, 2012.

Holster, T., W. Pellowe (2011). "Using a Mobile Audience Response System for Classroom Peer Assessment." *JALTCALL 2011* (Japan Association for Language Teaching's Computer-Assisted Special Interest Group 2010 Conference). Kurume University, Kurume, June 4-5, 2011.

Pellowe, W. (2010) MOARS: Mobile Audience Response System http://moars.com

Pellowe, W., T. Holster & K. Ryan (2010). "Student Response System for Mobile Devices." *JALT 2010* (Japan Association for Language Teaching 36th Annual International Conference on Language Teaching and Learning). Aichi Industry and Labor Center, in Nagoya City, Aichi, November 19-23, 2010.
Pellowe, B. (2011). "Student Response Systems." *The Language Teacher*, 35.6 (November / December 2011), pp. 53-55.
Wikipedia (2012), Sample standard deviation formula, retrieved from