A PROGRAM THAT ANALYZES MULTIPLE-CHOICE TEST RESULTS USING OUTPUT FROM SAS'S ITEM MACRO

Marvin S. Margolis
Millersville University

Abstract
In its macro library, SAS provides an item macro that computes statistics for the analysis of multiple-choice test data. In other words, the macro can help grade a test. In the paper, I present a SAS program that demonstrates how I apply the macro's output as input data to automate test grading tasks. The program provides statistical analysis beyond that of the item macro. It simplifies test grading work while expanding the information a test provides to students and professor. It provides the test takers with individual printed reports, that is, their score, the test average grade, and their individual test item results. After examining the SAS program and item macro statistics, the test giver can adjust (curve) test scores accordingly.

Introduction
University students took a multiple-choice test in class using number two pencils and a standard answer sheet suitable for scanning. I (the test giver) completed the official answer sheet which I placed on top of the student answer sheets. I took the package to our university computer center. The center staff optically scanned the answer sheets and put the raw data file in my IBM 4300 account. Using the CMS operating system, I directed the 4300 computer to grade the test using a previously written SAS program. The program reads in the raw data file, calls upon the test grading capabilities of SAS's item macro, uses the item macro output as input for further analysis, and prints out the results. The following sections explain some of the pages of the program printout.

The Raw Input Data File
Figure 1 shows a printout of a sample of the variables in the raw data file as they will be input into the item macro. The scanning program converts a, b, c, d, and e choices into the numbers 1-5. There are thirty-three observations (not all shown) which means 32 students took the test and fifty-three variables (not all shown) which means there were 50 questions on the test. The first observation for each Qi variable is the official answer for the 50 multiple-choice questions. The student answers are real; I have changed the student names.

Variable SSN (not shown) is the social security number, the name variable is the last name, and the course variable (not shown) is section 6 of economics 102. In examining the raw data, I can check for (1) missing student answers and (2) if all numbers down a column are the same, then no students missed the question.

The Test Score Descriptive Statistics And Frequency Distribution
Figure 2, panels a-c, begin the SAS item macro output. Panel a presents the grade report showing the student names, the number of questions correct in column 2, the number of questions not filled in in column 3, the number of questions with more than one student choice and therefore invalid in column 4, and the percentage of correct questions in column 5.

Panel b presents the summary test statistics. The Kuder-Richardson 20 and 21 statistics measure the internal consistency of the test questions, that is, the question reliability and homogeneity (Anastasi, 1988). The KR 21 approximates KR 20 and usually underestimates it. A KR 20 or 21 value of 1 indicates perfect reliability while 0 indicates perfect unreliability. The total number of subjects equals the number of subjects times the number of items. The Spearman-Brown prophecy indicates how many multiple-choice questions I would need to include on the test to raise the KR 21 value to 0.9 reliability. The prophecy number of 254 items indicates a large scope to improve the reliability of the test questions.

Panel c presents the test frequency distribution. The item macro computes the standard scores in column 2 as follows: assign the raw score average the value of 500 and the sample
standard deviation a value of 100. Therefore the lowest raw score is within two standard deviations of the mean while the highest is slightly greater than two standard deviations above the mean.

The Item Macro Item Analysis
Figure 3 shows the item macro's printout of its item analysis for the first three multiple-choice questions while Figure 4 shows the item analysis for multiple-choice questions 43-45. I have omitted the item analysis for other questions on the test.

For the first question (item Q1), the first row lists the possible choices: 1 = a, 2 = b, 3 = c, 4 = d, and 5 = e. The asterisk indicates the correct choice. The second row is the percentage of students who selected a particular choice. For a five-option multiple-choice question, the average proportion correct should be 69% for greatest discrimination among the student abilities.

The third and fourth rows report the percentage distribution of answers for the extreme groups. The third row reports the percentage distribution for the 33% of the students who scored the highest on the test. In other words it is the percentage of the upper third percentile group responding to each alternative. The fourth row reports the percentage distribution of the 33% who scored the lowest on the test. The better the reliability of a question, the higher the percentage of students who get the question correct who are in the upper 33% rather than in the lower 33%.

The item-total rows are point biserial and biserial correlation coefficients (Anastasi, 1988) between the percentages in the upper and lower 33% values. A negative point biserial correlation coefficient indicates more students in the lowest 33% got a question correct than did those in the upper 33%. A negative value for the correct choice signals that there is a problem with the particular question.

For Q2 100% of the students chose the correct answer. I could discard the question from the test with no loss of test discrimination. Q1 and Q2 explain why the KR 20 reliability test statistic is no higher than it is.

Q3 yields almost ideal outcomes for a multiple-choice question. First we observe the 69% average proportion correct that is ideal for all questions. Second the 77.8% in the upper 33% who got the question correct is larger than the 50% in the lower 33%. The item-total point biserial correlation coefficient for the correct choice is positive at 0.424. Students chose incorrect answers with approximately equal but low probability. The item-total point biserial correlation coefficients for the incorrect choices are negative except for choice 4.

Figure 4 shows an example of problem questions. Unlike Q1 and Q2 in Figure 3 which do no harm even if they contribute little to the test's discrimination ability, Q43 and Q44 do cause problems because they are poorly designed. In Q43 only 22% of the students chose the correct choice while in Q44 only 19% chose the correct choice. In both questions over 50% of the students chose a choice that I call incorrect. In Q44, the worst question, only 11.1% of the upper 33% answered correctly while 30% of the lower 33% answered correctly.

The item-total point biserial correlation coefficient is negative for the correct choice and more highly positive for some of the incorrect choices. I argue later in the paper that the Q43 and Q44 item analysis results call for a grade adjustment, that is, curving the grades.

Figure 4 ends the SAS item macro output. The remaining figures come from additional analysis that I programmed SAS to do.

Scatter Plots Of Test Scores Versus Test Taker Names
Figure 5 presents a scatter plot of the raw test scores (multiplied by 2) versus the student names. I have drawn dashed horizontal lines at scores of 50-60-70-80-90. I have also sorted the student names in alphabetical order on the horizontal axis.

Curved Test Scores
Figure 6, panel A, presents the curved score results. The program does not curve the raw scores (multiplied by 2) if the raw score average is above 75. I set 75 as the desired average score for every test. However, if the raw test scores average above 75, I do not lower the average downward. For this test, the 69 raw
score average meant the program added six points to student scores.

First I sort the names in alphabetical order to simplify entering them into a physical gradebook. Figure 6, panel B, gives the frequency distribution of the curved grades. Usually a curved score of 90 or above is an A grade, between 89 and 80, a B grade, etc. Note how the curved grades are bell shaped.

Individual Printed Reports For Students

Figure 7 shows a sample part of an individual student report that I return to students after a test. The program produces one report page per student that I generate using SAS's by and pageby options. The report contains among other results the student test score, the curved test average, and individual test item results. I physically deliver the report to students to preserve their privacy.

The page heading gives the date students took the test (not shown), the class, section, and test number, the professor, and the course name. The student's score, the class minimum, average, and maximum follow next. If there is a curve, its amount shows under the variable name curve. Next 50 Q variables (not all shown) report the student's answer choices for each question. Next 50 A variables report the official (correct) answers for each question. The 50 BIN variables are binary variables with either zeroes or ones. If students answered a question correctly, the appropriate BIN value equals 1; if they answered incorrectly, it equals zero.

Students can check their curved score using the following formula:

\[ \text{Curved Score} = 100 - 2^a \times \text{number wrong} + \text{curve.} \]

For student Beach, the calculation is 78 = 100 - 2^4(14) + 6.

A Table of Questionable Questions

Figure 8 presents summary information for the questions containing unusual item statistics. The input for the table is the output from SAS's item macro. To qualify to be listed, a question must be "too" hard (the overall proportion of students who answer correctly must be less than 50%). Alternatively, more low than high scorers must answer correctly (the item-total point biserial correlation coefficient must be less than zero for the correct answer). The table locates problem questions and saves me the effort of reading through every question in the item analysis.

Conclusions

Future improvements include setting up an electronic gradebook for all tests students take in a semester. Also I would like a more efficient method of changing the titles, the number of questions, and the curving procedures for each test. In particular, I would like to relate the curving procedure directly to a test's problem questions.

Trademarks

SAS is a registered trademark of SAS Institute, Inc. in the USA and other countries.

References


Acknowledgments

I thank Barry Walton of Millersville University's academic computer assistance department for assistance over many years and particularly on this project. Many SAS-L listserv contributors inspired me to complete the project.

\[^1\text{Associate Professor of Economics, Economics Department, Millersville University, Box 1002, Millersville, PA 17551. Telephone 717-872-3561, E-Mail: Margolis@MU2.Millersv.Edu} \]
Figure 1
Sample Raw Data
(10 names, 20 questions)

| Name     | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q |
| Margolis | 4 | 5 | 3 | 5 | 1 | 4 | 3 | 1 | 4 | 3 | 4 | 3 | 2 | 3 | 4 | 3 | 1 | 1 | 1 | 1 | 1 | 3 |
| Porter   | 4 | 5 | 1 | 2 | 2 | 5 | 3 | 2 | 4 | 3 | 3 | 3 | 5 | 4 | 5 | 3 | 1 | 1 | 1 | 1 | 1 | 5 |
| Mueller  | 4 | 5 | 3 | 5 | 2 | 2 | 3 | 1 | 4 | 3 | 4 | 2 | 2 | 3 | 4 | 3 | 1 | 1 | 1 | 1 | 1 | 3 |
| Malvern  | 4 | 5 | 3 | 5 | 2 | 1 | 3 | 1 | 2 | 3 | 4 | 3 | 2 | 3 | 3 | 4 | 3 | 1 | 1 | 1 | 1 | 3 |
| Mendel   | 4 | 5 | 2 | 1 | 2 | 4 | 5 | 2 | 1 | 3 | 3 | 3 | 2 | 3 | 4 | 3 | 1 | 1 | 1 | 1 | 1 | 3 |
| Marlow   | 4 | 5 | 3 | 5 | 3 | 3 | 3 | 4 | 3 | 4 | 3 | 2 | 3 | 4 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 3 |
| Dichter  | 4 | 5 | 1 | 4 | 2 | 3 | 5 | 1 | 1 | 3 | 4 | 1 | 2 | 4 | 2 | 3 | 1 | 1 | 1 | 1 | 1 | 3 |
| Korman   | 5 | 5 | 4 | 5 | 2 | 4 | 3 | 1 | 2 | 3 | 4 | 5 | 3 | 3 | 4 | 3 | 1 | 1 | 1 | 1 | 1 | 3 |
| McIntyre | 4 | 5 | 1 | 4 | 3 | 3 | 5 | 1 | 4 | 3 | 4 | 3 | 2 | 3 | 4 | 3 | 1 | 1 | 1 | 1 | 1 | 3 |

Figure 2
Panel A
Sample Grade Report

<table>
<thead>
<tr>
<th>Name</th>
<th>Score</th>
<th>#Missing</th>
<th>#Invalid</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porter</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>Mueller</td>
<td>46</td>
<td>0</td>
<td>0</td>
<td>92</td>
</tr>
<tr>
<td>Malvern</td>
<td>31</td>
<td>0</td>
<td>0</td>
<td>62</td>
</tr>
<tr>
<td>Mendel</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>70</td>
</tr>
<tr>
<td>Marlow</td>
<td>37</td>
<td>0</td>
<td>0</td>
<td>74</td>
</tr>
<tr>
<td>Dichter</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>54</td>
</tr>
<tr>
<td>Korman</td>
<td>34</td>
<td>0</td>
<td>0</td>
<td>68</td>
</tr>
<tr>
<td>McIntyre</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>56</td>
</tr>
</tbody>
</table>
Figure 2
Panel B
Summary of Test Statistics

<table>
<thead>
<tr>
<th>Number of Items</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Subjects</td>
<td>32</td>
</tr>
<tr>
<td>Mean Score</td>
<td>34.563</td>
</tr>
<tr>
<td>Variance of Scores</td>
<td>29.544</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>5.435</td>
</tr>
<tr>
<td>Kuder-Richardson 20</td>
<td>0.730</td>
</tr>
<tr>
<td>Kuder-Richardson 21</td>
<td>0.640</td>
</tr>
<tr>
<td>Standard Error (from KR20)</td>
<td>2.822</td>
</tr>
<tr>
<td>Standard Error (from KR21)</td>
<td>3.261</td>
</tr>
<tr>
<td>Minimum Score</td>
<td>24</td>
</tr>
<tr>
<td>Maximum Score</td>
<td>46</td>
</tr>
<tr>
<td>Total Number of Answers</td>
<td>1600</td>
</tr>
<tr>
<td>Number missing</td>
<td>0</td>
</tr>
<tr>
<td>Number invalid</td>
<td>0</td>
</tr>
</tbody>
</table>

Spearman-Brown Prophecy (from KR21): To obtain a reliability of 0.90, the test should contain 254 items.

Figure 2
Panel C
Test Frequency Distribution

<table>
<thead>
<tr>
<th>Raw Score</th>
<th>Standard Score</th>
<th>Percentile</th>
<th>Percent</th>
<th>Frequency</th>
<th>Cum Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>710</td>
<td>98</td>
<td>3.1</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td>45</td>
<td>692</td>
<td>96</td>
<td>0.0</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>44</td>
<td>674</td>
<td>96</td>
<td>0.0</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>43</td>
<td>655</td>
<td>95</td>
<td>3.1</td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td>42</td>
<td>637</td>
<td>92</td>
<td>3.1</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>41</td>
<td>618</td>
<td>89</td>
<td>3.1</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>40</td>
<td>600</td>
<td>84</td>
<td>6.3</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>39</td>
<td>582</td>
<td>78</td>
<td>6.3</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>38</td>
<td>563</td>
<td>73</td>
<td>3.1</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>37</td>
<td>545</td>
<td>67</td>
<td>9.4</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>36</td>
<td>526</td>
<td>57</td>
<td>9.4</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>35</td>
<td>508</td>
<td>48</td>
<td>9.4</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>34</td>
<td>490</td>
<td>42</td>
<td>3.1</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>33</td>
<td>471</td>
<td>37</td>
<td>6.3</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>32</td>
<td>453</td>
<td>32</td>
<td>3.1</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>31</td>
<td>434</td>
<td>26</td>
<td>9.4</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>30</td>
<td>416</td>
<td>20</td>
<td>3.1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>29</td>
<td>398</td>
<td>17</td>
<td>3.1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>28</td>
<td>379</td>
<td>14</td>
<td>3.1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>27</td>
<td>361</td>
<td>9</td>
<td>6.3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>26</td>
<td>342</td>
<td>6</td>
<td>0.0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>324</td>
<td>6</td>
<td>0.0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>306</td>
<td>3</td>
<td>6.3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
### Item Macro Item Analysis

#### Item Q1: is 1 2 3 4 5

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4*</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Responses</strong></td>
<td>3.1%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>93.8%</td>
<td>3.1%</td>
</tr>
<tr>
<td>Upper 33%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Lower 33%</td>
<td>10.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>90.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

**Item-Total:**
- **Point Bis:** -0.187 0.000 0.000 0.148 -0.019
- **Biserial:** -0.462 0.000 0.000 0.291 -0.047

#### Item Q2: is 1 2 3 4 5

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Responses</strong></td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Upper 33%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Lower 33%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

**Item-Total:**
- **Point Bis:** 0.000 0.000 0.000 0.000 0.000
- **Biserial:** 0.000 0.000 0.000 0.000 0.000

#### Item Q3: is 1 2 3 4 5

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3*</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Responses</strong></td>
<td>12.5%</td>
<td>9.4%</td>
<td>68.8%</td>
<td>6.3%</td>
<td>3.1%</td>
</tr>
<tr>
<td>Upper 33%</td>
<td>0.0%</td>
<td>11.1%</td>
<td>77.8%</td>
<td>11.1%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Lower 33%</td>
<td>30.0%</td>
<td>10.0%</td>
<td>50.0%</td>
<td>0.0%</td>
<td>10.0%</td>
</tr>
</tbody>
</table>

**Item-Total:**
- **Point Bis:** -0.481 -0.134 0.424 0.190 -0.254
- **Biserial:** -0.773 -0.233 0.555 0.374 -0.628
### Figure 4

#### Item Macro Item Analysis

<table>
<thead>
<tr>
<th>Item Q43: * is keyed</th>
<th>1</th>
<th>2*</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Responses</strong></td>
<td>6.3%</td>
<td>21.9%</td>
<td>15.6%</td>
<td>3.1%</td>
<td>53.1%</td>
</tr>
<tr>
<td><strong>Upper 33%</strong></td>
<td>0.0%</td>
<td>33.3%</td>
<td>22.2%</td>
<td>0.0%</td>
<td>44.4%</td>
</tr>
<tr>
<td><strong>Lower 33%</strong></td>
<td>20.0%</td>
<td>20.0%</td>
<td>20.0%</td>
<td>0.0%</td>
<td>40.0%</td>
</tr>
</tbody>
</table>

**Item-Total:**

- **Point Bis:** -0.268 0.128 -0.126 0.082 0.087
- **Biserial:** -0.528 0.179 -0.190 0.202 0.109

<table>
<thead>
<tr>
<th>Item Q44: * is keyed</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4*</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Responses</strong></td>
<td>12.5%</td>
<td>56.3%</td>
<td>6.3%</td>
<td>18.8%</td>
<td>6.3%</td>
</tr>
<tr>
<td><strong>Upper 33%</strong></td>
<td>22.2%</td>
<td>44.4%</td>
<td>0.0%</td>
<td>11.1%</td>
<td>22.2%</td>
</tr>
<tr>
<td><strong>Lower 33%</strong></td>
<td>0.0%</td>
<td>60.0%</td>
<td>10.0%</td>
<td>30.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

**Item-Total:**

- **Point Bis:** 0.225 -0.143 -0.124 -0.140 0.335
- **Biserial:** 0.362 -0.180 -0.243 -0.203 0.659

<table>
<thead>
<tr>
<th>Item Q45: * is keyed</th>
<th>1</th>
<th>2</th>
<th>3*</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Responses</strong></td>
<td>3.1%</td>
<td>0.0%</td>
<td>96.9%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>Upper 33%</strong></td>
<td>0.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>Lower 33%</strong></td>
<td>10.0%</td>
<td>0.0%</td>
<td>90.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

**Item-Total:**

- **Point Bis:** -0.254 0.000 0.254 0.000 0.000
- **Biserial:** -0.628 0.000 0.628 0.000 0.000
Figure 5: Plot of Score Versus Name

Score

Beach 0 10 20 30 40 50 60 70 80 90 100
Dichter
Hartman
Hein
Korman
Lightman
Lowper
Malvern
Marlow
McIntyre
Mendel
Mueller
Peters
Porter
Shrader
Stockard
Webster

...
Figure 6
Panel A
Curved Score Results

<table>
<thead>
<tr>
<th>Name</th>
<th>Curved Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beach</td>
<td>78</td>
</tr>
<tr>
<td>Brach</td>
<td>92</td>
</tr>
<tr>
<td>Campisi</td>
<td>66</td>
</tr>
<tr>
<td>Christ</td>
<td>82</td>
</tr>
<tr>
<td>Crandall</td>
<td>90</td>
</tr>
<tr>
<td>Dichter</td>
<td>60</td>
</tr>
<tr>
<td>Frost</td>
<td>86</td>
</tr>
<tr>
<td>Hart</td>
<td>84</td>
</tr>
<tr>
<td>Hartman</td>
<td>70</td>
</tr>
<tr>
<td>Hein</td>
<td>72</td>
</tr>
<tr>
<td>Hurt</td>
<td>88</td>
</tr>
<tr>
<td>Korman</td>
<td>74</td>
</tr>
<tr>
<td>Kristol</td>
<td>60</td>
</tr>
<tr>
<td>Ledyard</td>
<td>80</td>
</tr>
<tr>
<td>Lightman</td>
<td>86</td>
</tr>
<tr>
<td>Lowper</td>
<td>72</td>
</tr>
<tr>
<td>Malvern</td>
<td>68</td>
</tr>
<tr>
<td>Marlow</td>
<td>80</td>
</tr>
<tr>
<td>McIntyre</td>
<td>62</td>
</tr>
<tr>
<td>Mendel</td>
<td>76</td>
</tr>
<tr>
<td>Mueller</td>
<td>98</td>
</tr>
<tr>
<td>Page</td>
<td>84</td>
</tr>
<tr>
<td>Peters</td>
<td>68</td>
</tr>
<tr>
<td>Porter</td>
<td>54</td>
</tr>
<tr>
<td>Pounds</td>
<td>78</td>
</tr>
<tr>
<td>Prince</td>
<td>80</td>
</tr>
<tr>
<td>Redstone</td>
<td>76</td>
</tr>
<tr>
<td>Roberts</td>
<td>76</td>
</tr>
<tr>
<td>Shrader</td>
<td>64</td>
</tr>
<tr>
<td>Stockard</td>
<td>78</td>
</tr>
<tr>
<td>Thompson</td>
<td>68</td>
</tr>
<tr>
<td>Webster</td>
<td>54</td>
</tr>
</tbody>
</table>

Figure 6
Panel B
Frequency Distribution of Curved Scores

<table>
<thead>
<tr>
<th>Ninety</th>
<th>Eighty</th>
<th>Seventy</th>
<th>Sixty</th>
<th>BelowSixty</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>9</td>
<td>10</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 7
Individual Student Report
EC 102-06, Fourth Test
M. Margolis, Fall, 1993, Principles of Economics II

Name = Beach

| C | S | C | C | C | Q | Q | Q | Q | Q | Q | A | A | A | A | A | A | A | A | B | B | B | B | B | B |
| o | c | i | l | l | u | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | i | i | i | i | i | i |
| u | o | a | a | a | r | | n | n | n | n | n | n | n | n | n | n | n | n | n | n | n | n | n | n | n | n | n |
| r | r | r | r | s | s | s | s | s | v | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| s | e | M | A | M | e | c | l | v | a | n | e | x |

| 1 | 7 | 5 | 7 | 9 | 6 | D | E | C | E | B | D | C | E | B | D | E | C | E | A | D | C | A | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| 0 | 8 | 4 | 5 | 8 | 2 | 0 | 6 |

Figure 8
SAS Data Set Containing Unusual Item Statistics
All Questions With Propor < .5 or IT_PB < 0

<table>
<thead>
<tr>
<th>Name</th>
<th>Respon</th>
<th>Keyed</th>
<th>Propor</th>
<th>IT_PB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q5</td>
<td>1</td>
<td>*</td>
<td>.28125</td>
<td>.09013</td>
</tr>
<tr>
<td>Q6</td>
<td>4</td>
<td>*</td>
<td>.43750</td>
<td>.15455</td>
</tr>
<tr>
<td>Q24</td>
<td>3</td>
<td>*</td>
<td>.43750</td>
<td>.40182</td>
</tr>
<tr>
<td>Q26</td>
<td>2</td>
<td>*</td>
<td>.37500</td>
<td>.48564</td>
</tr>
<tr>
<td>Q30</td>
<td>5</td>
<td>*</td>
<td>.46875</td>
<td>.41627</td>
</tr>
<tr>
<td>Q31</td>
<td>3</td>
<td>*</td>
<td>.28125</td>
<td>.41493</td>
</tr>
<tr>
<td>Q36</td>
<td>5</td>
<td>*</td>
<td>.34375</td>
<td>.52653</td>
</tr>
<tr>
<td>Q39</td>
<td>4</td>
<td>*</td>
<td>.37500</td>
<td>.28053</td>
</tr>
<tr>
<td>Q43</td>
<td>2</td>
<td>*</td>
<td>.21875</td>
<td>.12805</td>
</tr>
<tr>
<td>Q44</td>
<td>4</td>
<td>*</td>
<td>.18750</td>
<td>-.14030</td>
</tr>
<tr>
<td>Q49</td>
<td>4</td>
<td>*</td>
<td>.40625</td>
<td>.06764</td>
</tr>
</tbody>
</table>