Spring 1997

Responses to Teacher Feedback on Errors Differ by Age and Gender

Sandra P. Clarkson
Hunter College

William (Bill) H. Williams Ph.D
CUNY Hunter College

How does access to this work benefit you? Let us know!
Follow this and additional works at: http://academicworks.cuny.edu/hc_pubs

Part of the Educational Methods Commons, Higher Education Commons, and the Science and Mathematics Education Commons

Recommended Citation
Sandra P. Clarkson and W.H. Williams "Responses to Teacher Feedback on Errors Differ by Age and Gender;" The AMATYC Review 18, no. 2 (1997): 38 - 46.

This Article is brought to you for free and open access by the Hunter College at CUNY Academic Works. It has been accepted for inclusion in Publications and Research by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.
Responses to Teacher Feedback on Errors Differ by Age and Gender

by

Sandra P. Clarkson

and

Wm. H. Williams

Hunter College of the City University of New York (CUNY)
New York NY 10021

Sandra Clarkson and W.H. Williams are Professors of Mathematics and Statistics at Hunter College. Clarkson received her M.Ed. and Ed.D. from the University of Georgia. Williams received his M.S. and Ph.D. from Iowa State University. Their interests include the impact of reading proficiency on learning mathematics and statistics as well as the use of error analysis and cooperative groups in teaching mathematics and statistics.

Abstract

Many students enter Hunter College’s developmental mathematics program committing errors (mis)learned years earlier. These errors typically persist into the adult years and it is important to correct them specifically; simply reteaching concepts is not sufficient. Furthermore, there is a strong correlation between completion rate and student perception of the instructor’s concern. To address both factors, we developed and tested an instructional technique to see whether giving detailed feedback to students about their errors would facilitate progress through the course. We found that the use of the feedback method had a clear positive effect on women; while for men, complex age by sex interactions and a smaller male sample size made the results less clear.

Why Study Feedback

A first year Hunter College student faces many of the same hurdles faced by students in the first year of numerous community colleges. Approximately 67% are required to take developmental reading, writing or mathematics (Kourelakos, 1979; Weisgal, 1996). The completion rate of the students enrolled in developmental mathematics courses is very important to the CUNY mission of providing a college education for all city residents who seek it. Many students enter Hunter College’s developmental mathematics program committing errors (mis)learned many years earlier. Since these errors typically persist into the adult years, it is important to identify and correct them specifically (Ashlock, 1976); simply reteaching a concept is not sufficient. Furthermore, teacher evaluation forms used each semester in this developmental course indicate a strong correlation.
between the finish rate in a class and student perceptions of the instructor’s care and concern. Motivated by these factors, we were interested in developing, and justifying, an instructional technique which

a) indicated to the students that their instructor was concerned about their individual progress in the course and was helping them individually, not simply dealing with them as a “class” of students;

b) gave students help specific to their own particular misconceptions and incorrect use of algorithms;

c) facilitated their movement through the course.

After consideration, we hypothesized that if teachers gave detailed feedback to students about the specific errors they had made, students would relearn the material correctly. Consequently, this study was planned to test whether giving detailed feedback to students about their errors would facilitate their progress through the course. The feedback method was effective for women; for men, complex age by sex interactions and a smaller sample size made the results less clear.

Review of the Literature

Many features of the developmental arithmetic course at Hunter were instituted because research indicated they would produce positive results. For example, the developmental arithmetic course has specified learning goals and strict deadlines for taking exams. Greenwood (1977) examined the effect of the imposition of external pacing on an individualized instructional program. When time goals and deadlines were provided, individualized instruction was more successful than traditional instruction in 25 of 27 studies; traditional was superior in two studies. When no pacing was imposed, traditional instruction was superior in one of four studies, individualized in none, and there was no difference in three studies.

Greenwood (1977) also found that the superiority of individualized over traditional instruction rests with the involvement of the instructor. If the teacher merely supervises the students’ work, little is gained. Only when the teacher attends to individual problems and deficiencies does this system perform better than the traditional approach. Even without Greenwood’s data, a teacher’s intuition (skill from experience) would predict that this supportive role for instructors would produce superior results.

Research on error analysis is rather extensive. In particular, however, the types of errors that are common in adults are (a) wrong operation, (b) computational error, (c) defective algorithm, and (d) random response [see discussion in Clarkson (1981)]. Of these four types of errors, the latter is the only one that is not persistent. Someone answering randomly has, essentially, no clue about the problem and should be re-taught the material. In all the other error types, something about the problem has evoked an incorrect response; the error should be pointed out and, in most cases, must be specifically corrected to prevent its recurrence. An instructor who would take the time to write comments about individual errors would surely relay to the students that (s)he cared about that student’s progress. And the value of such written feedback — feedback specific to student errors — has been demonstrated in several studies in traditional classrooms (Schoen and Kreye, 1974). And, since “teacher-learner-subject matter
interactions...must be at the very core of every educational endeavor;” (Wheeler, 1989), the teacher comments about student errors made up an important part of this, otherwise, individualized study program.

The Environment of the Study

Hunter College was founded in 1870 to educate young women. It became coeducational in 1964 and currently about seventy percent of the 19,500 students are women. The Dolciani Mathematics Learning Center (DMLC) is organized within the Department of Mathematics and Statistics. The DMLC has its own library, a laboratory, and a computer center and employs nearly 70 people, mostly as course instructors and tutors. The Learning Center offers several introductory mathematics courses. At the time of the study, Mathematics 001 (MATH 001), an arithmetic/beginning algebra proficiency course, was the largest of these courses, with class sizes ranging from 75 to 90.

Upon entrance to Hunter, all undergraduates and nonmatriculated students must take two mathematics proficiency examinations, one mandated by CUNY and one by Hunter College. Depending on the results of the former exam, students may be required to take MATH 001 and, as a result, the students in MATH 001 are not there by choice. Successful completion of the placement examinations, or of MATH 001, is a prerequisite for any other math or science course at Hunter. MATH 001 is also the one course that can keep students from graduating. This, combined with the fact that most of the students have not had extensive training in mathematics, gives many of the MATH 001 students serious apprehensions about the course.

MATH 001 is a self-paced course (structured for completion within one semester, but with flexibility during the semester). Course attendance is mandatory but the student need not attend the instructor’s lectures, held in a separate classroom. Instead, during class time students have the option of working in the DMLC lab, which is staffed with several tutors. When additional tutoring is needed, students may use the DMLC laboratory, library, or computer center during any open hours.

MATH 001 contains learning goals, called objectives—specific statements of what students are to learn. Students are required to pass individual mastery-based unit tests on eight areas of mathematics and a final exam, which covers topics on the placement test, as well as additional material needed for subsequent college level courses. The eight units covered in the course are arithmetic of whole numbers; fractions; decimals; ratio, proportion, and percent; signed numbers; and three units of beginning algebra. Both computational skills and applications are taught in this course. Passing the final exam means making a score that, had these students been entering students, would have exempted them from remediation. The final exam is a different version of the same exam they failed upon entry; passing this exam is required for exit from the course.

In the semester of the experiment, the eight unit tests contained a total of 60 objectives. Students were retested on each unit until they correctly answered two out of three questions per objective on all but one objective on the unit tests. Unit tests for all students were graded by the instructor on a pass/fail basis. No partial
credit was given on these tests. Normally a student must pass all eight exams to be eligible to take the final examination. In the semester of the experiment, however, all students were allowed to take the final if they attended the course up until the end of the semester. Only students who had completed the required work and passed the exam were given a grade of Credit and allowed to progress to the next course.

Clearly, the completion rate of students enrolled in MATH 001 is very important to the overall success of the mission of the Dolciani Center. Not only are there obvious pedagogical reasons, but any reduction in the number of repeating students is very important financially.

**Design of the Study**

During the semester of the study, there were eleven sections of MATH 001 with about 75 students each. In each section, we focused on first semester freshmen who had never taken the course before. The freshmen were randomly placed in a control group and an experimental group, which we examined to ensure that they were balanced. These groups were balanced on all incoming measurements and test scores; statistically the two groups were as identical as possible. In all, 220 freshmen entered the study and 193 of them completed the final exam.

Students in the experiment were not singled out for any unique attention by the instructors. In fact, instructors did not know which students were in the experimental group and which were in the control group, although they did know which students were taking part in the study. Teacher training was routinely given to instructors to insure that the level of instruction they gave was consistent from section to section. In addition, a detailed course outline specified text sections, page numbers and problems that all students were expected to cover. The tests, taken by all students, were carefully constructed to be comparable from section to section. All students, including those taking part in the experiment, were given the opportunity to schedule their own exams, as long as they met deadlines. All these procedures were a part of the routine structure of the course and were not changed because of the experiment.

All students in the study — both control groups and experimental groups — had their exams graded by one person, the researcher. The only difference between the treatment of the two groups was the fact that the experimental group received error feedback on their exams.

Encouraging remarks were given to all students. For example, when a student passed all objectives on the exam, (s)he was given a statement like “Good work!” or “Keep up the good work!” or “Way to go!” or “Bravo!” or some such comment. When students were careless, they were cautioned “Be careful!” or “Check your work” or “Read the problem carefully.” In addition, however, students in the experimental group (as opposed to those in the control group) were given specific information on their errors.

For example, a student whose errors were caused by incorrect multiplication by 6 was told what the error was; the suggestion was then made to review the multiplication facts for six. Students solving equations incorrectly were given details about their unique errors and a complete, correct solution was also given to
them. Students who copied their answers incorrectly from their scrap paper to the
test sheet were told to number the problems on their scrap paper and to carefully
check that they had copied their answers correctly from their work to the test.
Remarks to the control group included positive and cautionary remarks comparable
to the experimental group. However, their errors were not specifically discussed.
Periodically, an incorrect digit was circled, but no written explanation was given.

The Data

Many measurements were obtained on each student, including information
relating to before, after and during the student’s attendance in MATH 001. The
variables relevant to this analysis are described below.

1) Sex

2) Age — Ages of the students’ ranged from 18 to over 50. However, there
were not many men above 25 nor women above 33.

3) CUNY—This is the student’s score on the entering CUNY mathematics
exam. This score is the basis for placing a student in MATH 001. Scores are
the number correct out of 40; a student who scores 25 or over exempts the
course.

4) Final Exam — The final exam is simply a different version of the CUNY
mathematics exam that the students took upon entrance to the college. As a
result, the CUNY score and the final exam score are directly comparable.

5) Change Ratio (CR) — The ratio of the CUNY exit exam to the CUNY
entrance exam. A higher ratio indicates greater improvement than a lower
one and can be less than one.

6) Percent Possible Gain (PPG) — The actual gain, or loss, a student makes in
the course as a percentage of the possible gain that the student can make,

\[
\text{Final Exam} - \text{CUNY} \over 40 - \text{CUNY} \times 100
\]

The Initial Results

We first analyzed the change ratio (CR) with the expectation that this would be
higher for the experimental group than for the control group. However, this was not
the case. The CR ratio for the control group was 1.597 and the experimental group
1.589, indicating that the control group showed slightly more improvement in the
course than the experimental group (see Table I). This difference between the
experimental and control groups is statistically insignificant with a small \( t \) value of
0.134.

<table>
<thead>
<tr>
<th></th>
<th>Count</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>St. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>99</td>
<td>1.597</td>
<td>0.487</td>
<td>0.049</td>
</tr>
<tr>
<td>Experimental</td>
<td>94</td>
<td>1.589</td>
<td>0.361</td>
<td>0.037</td>
</tr>
</tbody>
</table>

Table 1. Analysis of Change Ratios – Final Exam/CUNY
Based on these results, one might be tempted to conclude that providing feedback has no effect on Math 001 students, but as we shall show, this would be an incorrect conclusion. Table 2 extends Table 1 to include sex. The important feature of Table 2 is that women in the experimental group improved while the change ratio of the men dropped. While the main effect difference between the control group and the experimental group, and the main effect difference between males and females are not statistically significant without adjustment for the interaction, the $p$ value of the interaction between treatment group and sex is 0.1351. While this is slightly above standard significance levels, why such an interaction would appear at all is unclear. Did the women in the experimental group prefer the extra help, while the men in the experimental group detested it?

<table>
<thead>
<tr>
<th></th>
<th>Females</th>
<th>Males</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>1.586</td>
<td>1.694</td>
<td>1.597</td>
</tr>
<tr>
<td>Experimental</td>
<td>1.614</td>
<td>1.432</td>
<td>1.589</td>
</tr>
<tr>
<td>Combined</td>
<td>1.599</td>
<td>1.546</td>
<td>1.593</td>
</tr>
</tbody>
</table>

Table 2. Change Ratios by Group and Sex

At essentially the same time that the treatment group/sex interaction was observed, a second interaction was also observed. When the data were rotated in three dimensions, CR vs. age vs. incoming CUNY score, and the CR vs. age dimensions came into view, we could clearly see a dramatic difference in the behavior patterns of the men and women. In particular, the regression of CR on age for women drops slightly, but significantly, and the regression for men rises rapidly. However while the male regression is also statistically significant, the oldest male was 32 and the total number of males was under 30.

These clear indications of different behavior for men and women demanded that the data be studied more closely.

**The Extended Analysis**

The relationship between CR and the incoming CUNY scores exhibits a clear, statistically significant, downward slope with a correlation of -0.663. This implies that CR is a biased measurement of change because students with lower incoming CUNY scores have a greater potential for improvement. To correct this bias, we instead calculated each student’s gain as percentage of that student’s maximum possible gain, percent possible gain, PPG. The relationship between PPG and CUNY shows no evidence of a biasing relationship and indeed, the correlation between the new measure, PPG, and CUNY is negligible, (0.041). In addition to a negligible correlation with CUNY, the correlations of PPG with all of the other incoming student variables are also very small. This clear lack of systematic bias in PPG makes it a far superior measure of student gain. This uneven potential for gain has been observed before; for example, in 1994 Mary Margaret Shoaf-Grubbs used...
a graphical adjustment technique to correct the bias. We believe the use of PPG to be new.

Tables 3 and 4 exhibit analyses of PPG. Table 3 shows that the experimental treatment helped women but the men did noticeably worse. The interaction between treatment group and sex is very clear. The women improved from 45.89 to 49.21 while the men’s scores declined significantly from 60.43 to 36.81. These results are not surprising. The interactions, which we have already noted, mask the differences between the sexes within the experimental groups. It is worth noting again that the men and women coming into the program were equally able, on average, in that the incoming CUNY scores for men and women in the two experimental groups showed no significant differences. In Table 4, we can see that the treatment group/sex interaction is significant with a p value less than 0.005 and that the feedback treatment is also significant.

<table>
<thead>
<tr>
<th></th>
<th>Females</th>
<th>Males</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>45.89</td>
<td>60.43</td>
<td>53.16</td>
</tr>
<tr>
<td>Experimental</td>
<td>49.21</td>
<td>36.81</td>
<td>43.01</td>
</tr>
<tr>
<td>Combined</td>
<td>48.62</td>
<td>47.55</td>
<td>47.42</td>
</tr>
</tbody>
</table>

**Table 3. Percentage Possible Gain (PPG)**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sums of Squares</th>
<th>Mean Square</th>
<th>F-ratio</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (Tr)</td>
<td>1</td>
<td>2056.85</td>
<td>2056.85</td>
<td>4.7078</td>
<td>0.0313</td>
</tr>
<tr>
<td>Sex</td>
<td>1</td>
<td>22.9026</td>
<td>22.9026</td>
<td>0.05242</td>
<td>0.8192</td>
</tr>
<tr>
<td>Tr*Sex</td>
<td>1</td>
<td>3622.85</td>
<td>3622.85</td>
<td>8.2922</td>
<td>0.0044</td>
</tr>
<tr>
<td>Error</td>
<td>189</td>
<td>82573.6</td>
<td>436.898</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>192</td>
<td>88276.20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4. ANOVA of PPG**

Finally, recall that the three dimensional graphs we described earlier suggested an interaction of improvement with the age of the student. The regression of PPG on age for women has an estimated slope of -0.34 and is significant (p = 0.021) and for men the slope is 4.04 with a p value of 0.0069. But we are currently studying age further because local smoothing of the data hints that, for young men, the regression first slopes downward and after that turns quickly upward.

Altogether, the results of this analysis permits the following summary,

- error feedback methodology does have a significant effect,
women react positively to this particular procedure, men tend to react negatively,

feedback appears to have a lessening effect on older women and an improving effect on older men.

Discussion

The error feedback program had strong effects on the students, although the interactions were complex. A few explanations for the differential effects on males and females are plausible. Mathematics 001 is required for students who have not achieved an acceptable level of arithmetic and algebra skills. It is possible that the men, who typically expect themselves to be competent in mathematics, may have had a negative reaction to having their errors pointed out. Unlike their female counterparts, they did not appear to learn from their mistakes. It is likely, when told what they did wrong, the men "stored" that information, but did not act on it. Without practice, students do not retain the information from class to class.

Women, on the other hand, seemed to use the feedback information to help them learn the mathematics. There is a considerable body of research of elementary and middle school-age students that indicates that on algorithmic (step-by-step) computational problems, females outperform males (Zambo and Follman, 1994). Here, even in the "relearning" of such material, females do seem to excel.

Age had a significant effect. Although, women reacted positively to the detailed feedback and men negatively, there was some indication that this might reverse with age. Older men seem to react better than younger men; perhaps they understand better that they need help with the material and so may be less subject to peer pressure. In contrast, older women tend to do worse than younger women.

References


...by a phenomenon that everybody who teaches mathematics has observed: the students always have to be taught what they should have learned in the preceding course. (We, the teachers, were of course exceptions; it is consequently hard for us to understand the deficiencies of our students.) The average student does not really learn to add fractions in an arithmetic class; but by the time he has survived a course in algebra he can add numerical fractions. He does not learn algebra in the algebra course; he learns it in calculus, when he is forced to use it. He does not learn calculus in a calculus class either; but if he goes on to differential equations he may have a pretty good grasp of elementary calculus when he gets through. And so on throughout the hierarchy of courses; the most advanced course, naturally, is learned only by teaching it. This is not just because each previous teacher did such a rotten job. It is because there is not time for enough practice on each new topic; and even if there were, it would be insufferably dull.

R.P. Boas

+++ The mind is not a vessel to be filled, it is a fire to be kindled. Plutarch

+++ The goal of teaching is learning, not teaching. Hugo Rossi

46