2010

Grading Practice Influence on the Value of an Assigned Grade

Carrie Bala
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/gradreports

Part of the Applied Mathematics Commons

Recommended Citation
Bala, Carrie, "Grading Practice Influence on the Value of an Assigned Grade" (2010). All Graduate Plan B and other Reports. 1254.
https://digitalcommons.usu.edu/gradreports/1254

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact dylan.burns@usu.edu.
GRADING PRACTICE INFLUENCE
ON THE VALUE OF AN ASSIGNED GRADE

By

Carrie Bala

A report submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Mathematics

UTAH STATE UNIVERSITY
Logan, Utah

2010
Grading Practice Influence on the Value of an Assigned Grade

Carrie Bala
Brynja Kohler
John Stevens
Utah State University
Abstract

This article presents results of a study of grading practice influence on the value of an assigned grade. The value of an assigned grade, as an indication of student achievement of learning goals, was measured using the Utah Criterion-Referenced Test (CRT) in the subjects of Geometry and Algebra 2. The grading policies of six mathematics teachers at the same high school were categorized according to grading practice and their combined 587 students’ scores on the Utah CRT were collected and analyzed. The findings suggest that certain grading practices, such as the use of a 1-4 scale and a criterion-referenced grading system, have a significant effect on the usefulness of a grade, as measured by its prediction value of an achievement test score.
Grading Practice Influence on the Value of an Assigned Grade

Perhaps the most important purpose of a grade is to communicate an overall achievement level of a student in a particular subject of study. When a student receives a passing grade, one would expect that student to have reached certain levels of achievement of specified learning goals. Similarly, when a student receives a failing grade, one would assume that student failed to attain certain achievement levels of the specified learning goals. These learning goals are content specific and also vary in learning level.

States' Offices of Education select learning objectives and design criterion-referenced tests to measure achievement of these objectives at learning levels of simple knowledge and algorithmic skill. These tests are administered at the end of a course and are often interpreted as summative evaluations of a student's work throughout the school year. However, the main purpose of these tests is to provide feedback on specific objectives, while a teacher may prepare a particular course with many more learning objectives. In general, an assigned grade reflects achievement levels of selected learning goals which may extend beyond what a State Office of Education has selected to measure on a criterion-referenced test.

However, with variation in teacher grading practices, a student's grade is not necessarily an accurate gauge of achievement levels, even with standards collaboratively selected. While it is noted that grades can reflect much more, one can conceivably use a State Criterion-Referenced Test to minimally measure achievement levels, and therefore study grading practices. According to Willingham, Pollack, and Lewis, "...grading variations are an important source of discrepancies between grade averages and test scores" (2002, p.8). This research suggests that
when grading systems are not comparable, the correlation between grades and test scores lowers. However, there have been studies to suggest that improving the uniformity of grading practices increases the reliability of the grade as a predictor of success on high-stakes tests (Elliot and Strenta, 1988). It is important to note that only high-stakes tests that measure achievement of specified standards are appropriate. Therefore, if schools wish to increase the accuracy of grades in communicating achievement of learning goals, a more uniform grading system should be established.

If a common grading system can indeed be established, then best practices must be considered and implemented. According to Marzano’s research in 2000, successful grading procedures should give timely information focused on academic achievement; teachers should give different scores reflecting the different standards assessed; teachers should use a common system of weighting or combining these scores to determine a grade, as well as a relevant scale to report this information. These best-practices should result in assigned grades that are better predictors of success on high-stakes tests and next-level courses.

Unfortunately, these practices are researched but not often implemented. “Practices vary greatly among teachers in the same school—and even worse, the practices best supported by research are rarely in evidence” (Reeves, 2008, p. 85). For example, within a school, two different teachers of the same course could use a norm-referenced grading and a standards-based grading system, respectively.

Norm-referenced grading is a common practice among teachers. With this system, student achievement is measured in reference to other students in the class. There exist predetermined numbers of high-level grades, mid-level grades, and low-level grades. Therefore, students are not expected to achieve high-levels of understanding, but rather to compete for the
high-level grades. With norm-referenced grading practices, "[h]igh grades are attained not through excellence in performance but simply by doing better than one's classmates" (Guskey, Bailey, 2001, p.36). Systematically assigning low-level grades to students, regardless of their achievement levels, is not a practice consistent with legal mandate that no child be left behind.

Another common practice is to assign grades based on reference to knowledge gained. With this practice, student effort is measured and grades assigned in reference to the entrance-level of each student to the course. "High grades are given to students who exhibit exceptional effort and improvement beyond what is expected of them" (Marzano, 2000, p.23). Again, however, equitable and high expectations for all students must be in place according to today's educational mandates.

Criterion-referenced grading is the system that research has found to be most effective for improving student achievement (Marzano, 2000). As the name implies, criterion-referenced (or standards-based) grading requires learning goals to be selected and presented to the students. Terwilliger suggests that grading be directly linked to an explicit set of instructional goals (1989). In addition, levels of understanding are presented in a rubric for each goal and students receive feedback in reference to this rubric throughout the unit. Hattie (1992) reports that there are marked student achievement gains when teachers provide students with specific feedback in reference to these objectives. Because this grading system can supply specific, timely achievement information for students, parents, and teachers, criterion-referenced grading is considered best practice.

Even with criterion-referenced grading in place, though, there needs to be a consistent practice as to what factors to admit into grading and to what extent these will be admitted. In his collection of multiple grading studies, Marzano states that "academic achievement is the primary
factor on which grades should be based” (2000, p.29). In a criterion-reference grading system, achievement is measured according to a rubric of understanding for each learning goal. The next question then is: To what extent, if any, should the following be included in the grade: Homework, Retakes, Effort, Attendance, Behavior? Again, best practice dictates that grades should be based solely on academic factors. Marzano states that it is appropriate to give feedback on non-academic achievement factors, but ideally should be kept separate from academic achievement reporting (2000). Cangelosi (1999) adds that awarding points for participation can suppress formative feedback and interfere with learning activities. Additionally, giving students formative feedback that does not affect their grade throughout a marking period allows students to make and learn from mistakes. As time progresses, teachers can then formally assess what students have learned in a particular marking period. “With standards-based grading, grades are based solely on summative assessments designed to measure content mastery” (Deddeh, Main, and Fulkerson, 2010, p.57). Criterion-referenced grading is a method for teachers to give effective and timely feedback to students while still allowing growth throughout the term.

Research also suggests teachers use a common, rational grading scale. One such example is a four-point scale. According to Reeves, “the four-point scale is a rational system, as the increment between each letter grade is proportionate to the increment between each numerical grade-one point”. In a four-point scale, a four is the highest score, representing a student’s ability to apply a skill to a new situation; a three represents a student’s ability to repeat procedures exactly as has been shown to them; a two score represents that a student can perform simple tasks individually and more complex tasks with help; a one score represents that the
student needs help with all tasks, simple and complex. There are many variations to this concept, but a common theme is an equal division of among grades.

At the high school studied, the Mathematics Department had six members, four of which were new to the district in 2004 or 2005. With such a new group, the high school administration held a summer Math Department meeting in 2005 to set goals and establish a policy of collaboration. The six math teachers had common prep time to continue discussions of student achievement and best teaching practices. The following school year, the school district adopted a weekly hour of collaborative meeting time for all teachers, specifically designed to focus on student achievement.

Among many topics of conversation, the mathematics department focused on a common grading policy. Because large variation in grade distribution for each course existed, dependent on the teacher, discussion and action began on creating uniformity in course content and grading policy. There also existed many problems with student achievement of high grade-point averages (GPAs) but inability to progress or qualify for advanced classes.

In 2007-2008, with district-level mandate, the Math Department created common assessments for Algebra 1, Geometry, and Algebra 2 (CDAS). Essential standards were chosen for these courses and a curriculum map was developed. During this school year, common student notebooks were also created and used in these courses. This common notebook was a daily template for note-taking, with accepted standards listed at the top of each lesson. Practice problems and assignments were listed for each lesson as well. Each class, regardless of teacher, used these notebooks to guide each lesson.

In addition, three of the six teachers separated the computerized grade book into agreed-upon standards for each quarter (following the curriculum map), replacing the typical system of
separating the grade book into tests, quizzes, homework, etc. All assessments, formal and informal, were inputted into a specific standard in the grade book.

An example is displayed for a geometry class in Figure 1. The standards for this quarter, listed more as topics in student-friendly language, were 1. Reasoning, 2. Points, Segments, Lines, 3. Coordinate Geometry, 4. Angles, and 5. Polygons and Polyhedrons. A category for the quarter exam was also inputted. Quizzes and CDAS (common assessments) were assigned to a particular category so that students, parents, and teachers could reference each assignment/assessment to a particular area of interest. The three teachers that used this grade book overwhelmingly endorsed this method of grade book set-up. While there were several comments from parents about the unfamiliar look of the grade book, the common thought among the three teachers was that this grade book set-up was the best method to monitor student progress on each standard.

Figure 1
In the following school year, 2008-2009, this goal was recorded from a math department meeting on grading policy:

**Goal:** *To ensure that a quarter grade reflects what the student knows.*
* (Not based on effort or participation)

Teachers discussed a different grading scale that was more proportionate than the traditional percentage scale. Three decided to use a 1-5 grading scale (similar to the 1-4 grading scale described) in addition to separating the grade book into quarter standards. Using this scale, any value from 1-5, including scores such as 3.3 or 3.5, was recorded. Homework was recorded but weighted zero. (One of these three teachers experimented with this weighting throughout the year). Retakes of assessments were allowed after working with the teacher, although this practice varied greatly with each teacher. Grades could be assigned based on a cumulative test at the end of each term, with the idea that student achievement progressed throughout the term.

During collaborative meetings, much discussion existed within the department on the benefits of this system, as well as the difficulties. Teachers that used the 1-5 scale and standards-based grading talked about how they liked the new system, while others questioned the practicality. Throughout all this discussion, common “cut-offs” or rubrics for assessments were developed, which made grading less subjective. Students quickly adapted and understood the 1-5 scale, which had replaced the 90,80,70,60 percentage scale they had experienced in the past. It did appear in general that there were less A’s on report cards, but significantly less F’s as well.

However, there was still a need to have more uniformity in the department’s policy. Students and parents still “teacher-chose”, based upon familiar grading procedures and perceived class difficulty. Common concern was that a student’s grade point average could be hurt by a specific teacher, specifically one who graded “harder”. Without a common grading practice in place, students would have a much different experience in these courses depending on their
teacher. Therefore, to promote fairness, additional changes were made in the 2009-2010 school year.

Two teachers were added to the Math Department to accommodate the ninth grade that was added to the high school. All 8 teachers for each course were required by the department chair to divide their grade books into the same quarter standards. (All teachers would be working under a standards-based system). In order to achieve this, teachers brought their computers to a meeting and each course was set up according to these standards. Maximum weights were established for homework in each course, although not enforced by any one in particular. Five teachers began using a 1-4 scale to grade all assessments, with common descriptors/cutoffs for common assessments. (The school Science Department was also working on grading policy and some had begun to use a four-point scale. Math teachers chose to be consistent with the Science Department and switched to a four-point scale). Three teachers used a traditional 90,80,70,60 percentage grading scale. While a criterion-referenced structure was in place, there was much variation in the grading procedures and scales used.

In this study, teachers collected high school students' math GPAs (grades in their math course only) for the 2009-2010 school year, along with each student's math State Criterion-Referenced Test (CRT) score for the same year. Students were grouped according to teacher, and therefore, according to a particular grading procedure. All teachers assigned grades using a criterion-referenced system. However, variations existed with regards to weighting and inclusion of certain factors such as effort, homework, and participation.

This study hopes to clarify whether or not a particular grading system produced an assigned grade that was a stronger indicator of achievement. Because the Math Department did not create a common summative assessment for the two courses of interest, the study uses the
Utah Criterion-Referenced Test results as an indicator of achievement. The CRT does not assess the same learning goals as the high school course. Hence, the use of the test limits the study to learning goals of basic skills and processes. This study focuses on the difference in teacher grading practices only and assumes that the six teachers are identically effective professionals.

The results of the study indicate that certain grading practices, such as standards-based grading and the four-point grading scale, improved the assigned grade as a measure of achievement.

Method

The authors collected grading policies for six math teachers at the same high school in study and assigned an A and B grouping to each grading policy. Each grading policy was assigned to two groups. Group A1 represented 4 teacher policies that used a 1-4 scale in grading. Group A2 represented 2 teacher policies that used a traditional 90-80-70-60-0 percentage scale in grading. Group B1 represented three teacher policies that used only standards-based factors in the grade assigned. For example, only scores from assessments on commonly-chosen standards were used to determine grades. Group B2 represented three teacher policies that used standards and non-standard factors in the grade assigned. For example, grades were determined from scores on assessments of commonly-chosen standards, as well as from participation, effort, and homework scores.

Individual student CRT scores for Geometry and Algebra 2 courses were then sorted by the corresponding teacher policy groups used to determine each student’s math GPA. Each student score had an A and B grouping. A total of 306 Geometry and 281 Algebra 2 scores and GPAs were studied.
Upon examination of the data, differences amongst teacher policies were immediately apparent. As evidenced in Figures 2 and 3, the spread of student math grades for each level of CRT score was different for the grading policies shown. For example, the math GPAs for students in Figure 2 who scored a 3 on the CRT ranged from 1.92 to 3.33; The math GPAs for students in Figure 3 who scored a 3 on the CRT ranged from 0.5 to 3.833. With these differences noted, a more sophisticated analysis was used to determine a relationship between grading policy used and a grade’s value as a measure of achievement.

Figure 2

Figure 3
Because of the polytomous nature of the data, a process of ordinal logistic regression was used to study the predictive value of a student’s math GPA on the state CRT score. This process also clarified the effect of a particular grading policy on the GPA’s predictive value. In this study, student math GPA along with the two grading policy categories, were the predictor variables (X) of a categorical response variable (Y), the state CRT score. Possible student math GPAs ranged from 0 to 4.0, with 4.0 representing an A grade. Grading policy categories were either A1 or A2, and B1 or B2, as described above. Possible student CRT scores (r) were 1, 2, 3, and 4, with 4 as the maximum.

The authors studied cumulative probabilities that the state CRT score was less than or equal to the predicted CRT score given certain covariates (Covariates were the combination of math GPA and grading group assigned). This probability was linked to the predictor variables with a logit link

\[ L_{r|i} = \log \frac{\pi_{r|i}^c}{1 - \pi_{r|i}^c} = \log \left( \frac{P(Y \leq r|covariates_i)}{P(Y > r|covariates_i)} \right) \]

with a logit link

With the use of the computer program SAS, the ordinal regression model for \( r = 1, 2, 3 \) (because of the cumulative probabilities) was

\[ L_{r|i} = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} \]

\( \beta \) coefficients were interpreted as the effect of each X variable on the log-odds of the cumulative response. \( X_1 \) represented the math GPA\(^2\) (the data was squared to help the fit of the model). \( X_2 \) represented the grading policy (a 1 for A1 or B1 and a 0 for A2 or B2). \( X_3 \) represented the combination of \( X_1 \cdot X_2 \). SAS gave estimates of each \( \beta \) as well as a p-value for each, corresponding to the null hypothesis for \( H_0: \beta = 0 \) (no effect).
The proportional odds assumption, in this model, stated that the $\beta$ values were the same regardless of $r$ score. For each course, Geometry and Algebra 2, the proportional odds assumption was tested for fit. The p-value for the test of the proportional odds assumption for the Geometry course data was a non-significant .0674 (The null hypothesis for this test was that the proportional odds assumption was correct and therefore, there were no problems using the model assumption). A significance threshold of 0.05 was used throughout the study.

This Proportional Odds Model was used to first study the Geometry student scores and their corresponding A category (A1: use of 1-4 grading scale, A2: use of traditional 90,80,70,60,0 percentage scale). The SAS report is shown on page 21. The math GPA(squared) and the grading system were significant predictors for the probability of given CRT scores, as their p-values were both less than .0001. However, the usefulness of the GPA(squared) as a predictor was largely dependent on the grading system. This dependence was shown with the significant p-value of the combined term less than .0001.

Since the GPA(squared) effect for both A groups was negative, an increasing GPA(squared) made lower CRT scores less likely. For the A1 group, every additional 1.0 on the GPA(squared) scale reduced the odds of a lower CRT score by 50 percent. For the A2 group, every additional 1.0 on the GPA(squared) scale reduced the odds of a lower CRT score by about 32 percent.

The Proportional Odds Model was then used to study the Geometry student scores and their corresponding B category (B1: grades based only on standards, B2: grades include standards and non-standard factors). The SAS report is shown on page 22. The proportional odds assumption was satisfied with a p-value of 0.2229. (The authors failed to reject the null hypothesis that the proportional odds assumption was correct). The math GPA(squared) and the
grading system were significant predictors for the probability of given CRT scores, as their p-values were .0001 and .0034, respectively. The usefulness of the GPA(squared) as a predictor, again, was dependent on the grading system. This dependence was shown with the p-value of the combined term at .0147.

Since the GPA(squared) effect for both B groups was negative, an increasing GPA(squared) made lower CRT scores less likely. For the B1 group, every additional 1.0 on the GPA(squared) scale reduced the odds of a lower CRT score by about 48 percent (1 - exp(-.6627)). For the B2 group, every additional 1.0 on the GPA(squared) scale reduced the odds of a lower CRT score by about 39 percent (1 - exp(-.4959)).

A visualization of the Proportional Odds Model results are found on page 23. For both the A and B comparisons, the slopes of each curve were steeper for the 1st conditions (A1 and B1). A steeper slope indicated a higher predictive value. These steeper slopes were supported statistically by the significant and negative interaction terms.

The Proportional Odds Model was then used to study the Algebra 2 student scores and their corresponding A and B categories. However, the proportional odds assumption was not met for A or B categories (test p-values of less than 0.0001 for A groups and .0035 for B groups). Therefore, the Non-Proportional Odds Model, or Generalized Logit Model, was considered for the Algebra 2 data.

\[ \tilde{L}_{r|i} = \beta_{0,r} + \beta_{1,r}X_{i,1} + \beta_{2,r}X_{i,2} + \beta_{3,r}X_{i,3} \]
In this model, all the $\beta$ terms depend on the student score, $r$. Also, the non-cumulative probabilities are related to the predictor variables using the logit link, with one level of $Y$ as a reference category ($Y = 4$).

$$
\pi_{r|i} = P(Y = r|\text{covariates}_i)
$$

$$
L_{r|i} = \log \frac{\pi_{r|i}}{\pi_{B|i}}
$$

Pages 24 and 25 contain the SAS output for this model study. For the study of the Algebra 2 scores, grades assigned from both a four-point scale and a traditional percentage scale were both good predictors of state CRT scores. While the overall predictive value of the GPA(squared) was more difficult to quantify with this model, the interaction or combination term in the model had a significant p-value of .0002 in the “Type 3 Analysis of Effects” section. Therefore, the predictive usefulness of the GPA(squared) depended upon the grading policy.

For the study of the Algebra 2, grades assigned from both a standards-based system and a mixed standards/non-standards-based system were good predictors of state CRT scores. Since the interaction term in the model had a p-value of .0023 in the “Type 3 Analysis of Effects” section, the predictive usefulness of the GPA(squared) depended upon the grading policy (less so than with the A groups).

Again, looking at a visualization of the Non-Proportional Odds Model results on page 26, the slopes of the curve were steeper for both the A1 and B1 groups (compared to A2 and B2, respectively). These steeper slopes were supported statistically by the significant overall interaction tests in the “Type 3 Analysis”.

Results

The GPA(squared) is a better predictor of CRT score for those in the A1 (four-point scale) group compared to those in the A2 (traditional percentage scale) group. In the geometry course study, because the p-value for the combined term in this model was less than .0001, the
prediction value of the GPA(squared) in the A1 group is significantly better than that of the A2 group. In the Algebra 2 course, the combined term in this model had a p-value of .0002 improving the predictive value of the GPA(squared) for those in the A1 group. From this data, one can conclude that the use of a four-point scale can improve the value of a math grade as a measure of achievement, as compared to the use of a traditional 90-80-70-60-0 percentage scale.

The GPA(squared) is a better predictor of CRT score for those in the B1 (standards-based grading) group compared to those in the B2 (mixed standards/non-standards grading) group, as well. In the geometry course study, the p-value for the combined term in this model was 0.0147. Therefore, the predictive value of the GPA(squared) in the B1 group is better than that of the B2 group (although, without such significant results as with the A groups). In the Algebra 2 course, the combined term in the model had a p-value of .0023, improving the predictive value of the GPA(squared) for those in group B1. One can therefore conclude that the inclusion of strictly standards-based factors in a grade can improve the value of a math grade as a measure of achievement, as compared to including non-standards factors in the grading system.

While both a 1-4 grading scale and the use of achievement in standards only in grade determination improved the predictive value of the math GPA for the state CRT score individually, the combination of the two techniques could make a much stronger math GPA predictive value. The linear model used in this study analyzes only A groups or B groups, not the influence of both, in the same model. This study does not extend to that combination, but would certainly be of interest to the teachers who participated in the study.

Discussion

It is important to note that the main purpose of the math grade assigned is to communicate achievement levels of specific learning goals. This study uses the Utah CRT to
measure accomplishment of the selected learning goals of Geometry and Algebra 2. However, the CRT is considered a minimal measure of the learning goals, more specifically measuring objectives of algorithmic skills, simple knowledge, and comprehension and communication. It would certainly be more appropriate to collaboratively design an assessment that more specifically identified and measured achievement of objectives with different learning levels and to use such an assessment to study assigned grades. In the absence of such an assessment, though, grading practices were studied with seemingly more positive results for those with a standards-based grade book and proportional grading scale.

Partly due to these results, the math department in the school of study continues to work on a more uniform grading policy. For the 2010-2011 school year, all math teachers in the department adopted the 1-4 scale as their scoring rubric and created a common standards-based breakdown for the grade book. In a group vote, the eight teachers agreed to minimally include non-standards factors in the grade book, ensuring that 95% of a student's grade would be standards-based. Additionally, a department disclosure was developed for all math teachers with this information. A huge effort was extended to increase parent participation in the Back-to-School Night event, in order to share and explain how student grading would look for the school year.

There continues to be difficulty in working within the school computerized grading system (SIS). Parent and student experience with the traditional percentage grading system often skews the perception of grades reported with a 1-4 scale. The computerized grading system must output a percentage and a letter grade for each student. However, using a 1-4 scale, the math department has adjusted the percentage scale to produce the appropriate letter grade. This
change in the relationship between percentages and letter grades is unfamiliar to most students and parents (and other teachers) and has been the topic of much conversation.

After working through this study, though, the authors recommend that school and district departments work towards a more uniform grading policy, specifically employing a standards-based grading system with a proportional grading scale. Adopting a standards-based grading system will improve the achievement levels of students on particular learning goals. With a more focused approach to assigning grades, teachers and students can track progress more accurately and remediate more effectively than in the past. And a final grade will be a better overall reflection of achievement levels of specific learning goals for the course.
References

Cangelosi, James. *Assessment Strategies for Monitoring Student Achievement.* (Pearson Education, 1999)


Geometry: A1 vs. A2

The LOGISTIC Procedure

Response Variable CRT_Score
Number of Response Levels 4
Model cumulative logit
Optimization Technique Fisher's scoring

Number of Observations Read 307
Number of Observations Used 306

Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>CRT_Score</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>39</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>127</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>117</td>
</tr>
</tbody>
</table>

Probabilities modeled are cumulated over the lower Ordered Values.

NOTE: 1 observation was deleted due to missing values for the response or explanatory variables.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Score Test for the Proportional Odds Assumption

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.7652</td>
<td>6</td>
<td>0.0674</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-1.7581</td>
<td>0.4559</td>
<td>14.8687</td>
<td>0.0001</td>
</tr>
<tr>
<td>Intercept</td>
<td>2</td>
<td>0.1877</td>
<td>0.3854</td>
<td>0.2372</td>
<td>0.6263</td>
</tr>
<tr>
<td>Intercept</td>
<td>3</td>
<td>3.9705</td>
<td>0.4966</td>
<td>63.9181</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
| GPA_
| Intercept                | 1  | 2.4168   | 0.5215         | 21.4807         | <.0001     |
| GPA_
| Intercept                | 1  | -0.3915  | 0.0694         | 18.8597         | <.0001     |
GRADING PRACTICE INFLUENCE ON THE VALUE OF AN ASSIGNED GRADE

Geometry: B1 vs. B2

The LOGISTIC Procedure

Response Variable CRT_Score
Number of Response Levels 4
Model cumulative logit
Optimization Technique Fisher's scoring

Number of Observations Read 307
Number of Observations Used 306

Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>CRT_Score</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>39</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>127</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>117</td>
</tr>
</tbody>
</table>

Probabilities modeled are cumulated over the lower Ordered Values.

NOTE: 1 observation was deleted due to missing values for the response or explanatory variables.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Score Test for the Proportional Odds Assumption

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.2136</td>
<td>6</td>
<td>0.2229</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-0.7262</td>
<td>0.3423</td>
<td>4.5012</td>
<td>0.0339</td>
</tr>
<tr>
<td>Intercept</td>
<td>2</td>
<td>1.0490</td>
<td>0.3142</td>
<td>11.1450</td>
<td>0.0008</td>
</tr>
<tr>
<td>Intercept</td>
<td>3</td>
<td>4.7464</td>
<td>0.4586</td>
<td>107.1359</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>GPA*Grading_Group_B</td>
<td>1</td>
<td>-0.4959</td>
<td>0.0643</td>
<td>83.4336</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>GPA*Grading_Group_B</td>
<td>1</td>
<td>1.4663</td>
<td>0.5011</td>
<td>8.5846</td>
<td>0.0034</td>
</tr>
<tr>
<td>GPA*Grading_Group_B</td>
<td>1</td>
<td>-0.1668</td>
<td>0.0584</td>
<td>5.9489</td>
<td>0.0147</td>
</tr>
</tbody>
</table>
GRADING PRACTICE INFLUENCE ON THE VALUE OF AN ASSIGNED GRADE

Geometry: A1 vs. A2
Grading_Group_A=1

Geometry: A1 vs. A2
Grading_Group_A=2

Geometry: B1 vs. B2
Grading_Group_B=1

Geometry: B1 vs. B2
Grading_Group_B=2

PLOT P(CRT <= 1) P(CRT <= 2) P(CRT <= 3)

PLOT P(CRT <= 1) P(CRT <= 2) P(CRT <= 3)
Algebra: A1 vs. A2

The LOGISTIC Procedure

Response Variable CRT_Score
Number of Response Levels 4
Model generalized logit
Optimization Technique Fisher's scoring

Number of Observations Read 281
Number of Observations Used 281

Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>CRT_Score</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>73</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>107</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>31</td>
</tr>
</tbody>
</table>

Logits modeled use CRT_Score='4' as the reference category.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Type 3 Analysis of Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA(\text{sq})</td>
<td>3</td>
<td>19.4359</td>
<td>0.0002</td>
</tr>
<tr>
<td>Grading_Group_A</td>
<td>3</td>
<td>11.3164</td>
<td>0.0101</td>
</tr>
<tr>
<td>GPA(\text{sq})\times\text{Grading_Group_A}</td>
<td>3</td>
<td>20.2318</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CRT_Score DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>12.6520</td>
<td>4.4243</td>
<td>8.1776</td>
<td>0.0042</td>
</tr>
<tr>
<td>Intercept</td>
<td>2</td>
<td>12.1153</td>
<td>4.4225</td>
<td>7.5048</td>
<td>0.0062</td>
</tr>
<tr>
<td>Grading_Group_A</td>
<td>3</td>
<td>10.4731</td>
<td>4.3969</td>
<td>5.6737</td>
<td>0.0172</td>
</tr>
<tr>
<td>GPA(\text{sq})</td>
<td>1</td>
<td>-0.9465</td>
<td>0.3088</td>
<td>9.3945</td>
<td>0.0022</td>
</tr>
<tr>
<td>GPA(\text{sq})</td>
<td>2</td>
<td>-0.8784</td>
<td>0.3076</td>
<td>8.1554</td>
<td>0.0043</td>
</tr>
<tr>
<td>GPA(\text{sq})</td>
<td>3</td>
<td>-0.6874</td>
<td>0.3029</td>
<td>5.1487</td>
<td>0.0233</td>
</tr>
<tr>
<td>Grading_Group_A</td>
<td>1 1</td>
<td>3.7650</td>
<td>5.0927</td>
<td>0.5466</td>
<td>0.0597</td>
</tr>
<tr>
<td>Grading_Group_A</td>
<td>1 2</td>
<td>2.2127</td>
<td>5.0744</td>
<td>0.1901</td>
<td>0.6629</td>
</tr>
<tr>
<td>Grading_Group_A</td>
<td>1 3</td>
<td>-0.0285</td>
<td>4.9770</td>
<td>0.0600</td>
<td>0.9964</td>
</tr>
<tr>
<td>GPA(\text{sq})\times\text{Grading_Group_A}</td>
<td>1 1</td>
<td>-0.6659</td>
<td>0.3730</td>
<td>3.1899</td>
<td>0.0742</td>
</tr>
<tr>
<td>GPA(\text{sq})\times\text{Grading_Group_A}</td>
<td>1 2</td>
<td>-0.3279</td>
<td>0.3622</td>
<td>0.8196</td>
<td>0.3653</td>
</tr>
<tr>
<td>GPA(\text{sq})\times\text{Grading_Group_A}</td>
<td>1 3</td>
<td>-0.00831</td>
<td>0.3441</td>
<td>0.0004</td>
<td>0.9907</td>
</tr>
</tbody>
</table>
The LOGISTIC Procedure

Response Variable CRT_Score
Number of Response Levels 4
Model generalized logit
Optimization Technique Fisher's scoring

Number of Observations Read 281
Number of Observations Used 281

Response Profile

Ordered Value CRT_Score Frequency
1 1 73
2 2 70
3 3 107
4 4 31

Logits modeled use CRT_Score='4' as the reference category.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Type 3 Analysis of Effects

Wald

Effect DF Chi-Square Pr > ChiSq
GPAsq 3 35.6629 <.0001
Grading_Group_B 3 11.6977 0.0085
GPAsq*Grading_Group_ 3 14.5369 0.0023

Analysis of Maximum Likelihood Estimates

Parameter CRT_Score DF Estimate Standard Error Chi-Square Pr > ChiSq Wald
Intercept 1 1 13.6093 3.6224 14.1151 0.0002
Intercept 2 1 13.0001 3.6166 12.9098 0.0003
Intercept 3 1 11.3663 3.6862 10.0334 0.0016
GPAsq 1 1 13.6093 3.6224 14.1151 0.0002
GPAsq 2 1 13.0001 3.6166 12.9098 0.0003
GPAsq 3 1 11.3663 3.6862 10.0334 0.0016
Grading_Group_B 1 1 1.2953 4.5410 0.0801 0.7772
Grading_Group_B 1 2 1.2953 4.5410 0.0801 0.7772
Grading_Group_B 1 3 1.2953 4.5410 0.0801 0.7772
GPAsq*Grading_Group_ 1 1 1.2953 4.5410 0.0801 0.7772
GPAsq*Grading_Group_ 1 2 1.2953 4.5410 0.0801 0.7772
GPAsq*Grading_Group_ 1 3 1.2953 4.5410 0.0801 0.7772
Algebra: A1 vs. A2
Grading Group A = 1

Algebra: A1 vs. A2
Grading Group A = 2

Algebra: B1 vs. B2
Grading Group B = 1

Algebra: B1 vs. B2
Grading Group B = 2