

AN ALTERNATIVE EXAMINEE-CENTERED STANDARD SETTING STRATEGY

Patrick M. Irwin
Center for Educational Testing and Evaluation
University of Kansas

Chad W. Buckendahl
Buros Center for Testing
University of Nebraska

Andrew J. Poggio
Center for Educational Testing and Evaluation
University of Kansas

AN ALTERNATIVE EXAMINEE-CENTERED STANDARD SETTING STRATEGY

Abstract

This study examined an alternative examinee-centered standard setting method that is a modification of the Contrasting Groups method called the Modified Contrasting Groups method (MCGM). The alternative method was designed to be affordable, not too complex computationally, and done in a reasonable amount of time. The study investigated how well the alternative method produced results that are approximately equivalent to other standard setting methods. The results generated by the MCGM were compared to the results produced by the Angoff method, Contrasting Groups method, and the Borderline Group method.

Two variations of the alternative method (MCGM 1 and MCGM 2) were explored in the study. MCGM 1 and MCGM 2 are based on having teacher predictions of student performance and the students' actual performance or scores on an assessment. The MCGM 1 and MCGM 2 follow different but standard mathematical procedures that are easy to follow. The calculation for the MCGM 1 requires finding the median score for each performance category, then calculating the average of the two adjacent performance categories. MCGM 2 is based on combining the adjacent performance categories and calculating a median value of the scores in the two adjacent categories.

This study found that MCGM 1 and MCGM 2 produced cut scores higher than the Angoff and Contrasting Groups method on average, but lower than the Borderline Group method. The study also found the MCGM 1 and MCGM 2 consistently produced different cut scores using the same data sets.

The MCGM 1 and MCGM 2 appear to be viable standard setting methods for many school districts in Nebraska to use to meet state and federal reporting requirements. For some districts, either or both of these variations could be added to what is currently in place to provide a range of plausible cut scores. For other districts, this method could be adopted as the sole cut score method.

This study is part of a larger study performed to assist Nebraska schools in finding a way to meet the state and federal requirements. The information presented on this page is intended to provide the rationale and intent of the study.

Nebraska's current assessment model encourages individual school districts to develop their own assessments to measure the state standards for state and federal reporting. The Nebraska Department of Education (2002) has developed a set of process standards with which each school district's assessment system must comply. These process standards are referred to as the quality criteria for assessment. There are six quality criteria that school districts must meet. They are: (a) the assessments reflect the state or state-approved local standards, (b) the students have an opportunity to learn the content, (c) the assessments are free from bias, (d) the level is appropriate for students, (e) there is consistency in scoring, and (f) the mastery levels are appropriate. Because the sixth criterion requires school districts to explain how they set their mastery levels or cut scores, school districts are expected to conduct a standard setting study and determine cut scores for each of their assessments.

With over 300 school districts in Nebraska, standard setting alone will require substantial district resources because of the requirements of the Nebraska Department of Education (NDE) to show that mastery levels were appropriately defined. With limited budgets and more budget cuts anticipated in the near future, the cost to both maintain and renew district assessments and to set a large number of cut scores will place a heavy burden on the majority of school districts. All Nebraska school districts have an assessment system in place, and they have set cut scores for reading and mathematics for grades 4, 8, and 11. However, by 2005-06, school districts needed

to establish cut scores for science in grades 4 or 5, 8, and 11. And, by the 2006-07 school year, districts will also need to establish cut scores for social studies in grades 4 or 5, 8, and 11.

In order for Nebraska school districts to comply with state and federal requirements, they need to use a method for determining cut scores (or mastery levels) that: (a) meets the state's requirements for setting mastery levels, (b) is affordable, (c) is not too computationally complex, and (d) can be completed in a reasonable amount of time. If school districts are not able to comply with state and federal requirements, both their funding and accreditation are in jeopardy.

The purpose of this study is to examine an alternative approach to the most frequently used standard setting methods that appear in the literature. In order to be considered acceptable, this alternative must place most students into the same performance categories as do the more complex methods. It also needs to be affordable, not too computationally complex, and able to be done in a reasonable amount of time. This alternative method should be scientifically sound and meet state and federal expectations. The study examined the ability of the two variations of the alternative method to produce results that are approximately equivalent to other standard setting methods.

Examinee-centered standard setting: An alternative approach

As formal assessments in educational settings have proliferated, the need for setting appropriate cut scores has also increased. Important decisions about students' performance on these assessments range from remedial or gifted program assignment to graduation eligibility. Additional policy considerations are also embedded within the requirements of NCLB. Strategies for setting cut scores are well-researched and debated with respect to their utility and appropriateness. Prevalent methodologies for establishing cut scores for these educational assessments include test-centered (e.g., Bookmark, Angoff) and examinee-centered strategies (e.g., Borderline group, Contrasting groups). Determining an appropriate standard setting strategy is dependent on considerations about the purpose of setting the standard, the characteristics of the population being tested, and the characteristics of the judges (Kane, 1998).

Although these different methodologies have their own assumptions, examinee-centered strategies may be preferred in education because of the tasks associated with the judgments. For examinee-centered methods, standard setting judges make decisions about the ability of their students in the context of performance level descriptors rather than judgments about students' likely performance on test items. These decisions are not unlike the tasks that a teacher performs in her/his classroom when making decisions about mastery or grades. Shepard, Glaser, Linn, and Bohrnstedt (1993) suggested using examinee-centered methods as opposed to the item-by-item approach used by many of the prominent test-centered methodologies. As Jaeger (1991) suggests, a qualified standard setting judge will have expertise in the content measured on the assessment, but will also have experience with the population of examinees being tested. For

standard setting judges in an educational setting this does not appear to be problematic.

However, results from examinee-centered methods have their own challenges.

This study evaluated two variations of a proposed, alternative examinee-centered standard setting strategy using actual data from a Midwestern school district. To evaluate the proposed methodology, the alternative must classify most students into the same performance categories as do the more complex methods. It also needs to be affordable, computationally palatable, and efficient for practitioners. This alternative method should also be scientifically sound and meet state and federal expectations. This study examined whether two variations of the proposed, alternative method produced results that were approximately equivalent to those obtained during a previous standard setting conducted on the same tests. The proposed alternative method is a modification of the original Contrasting Groups method described by Livingston & Zieky (1982). Two variations of the proposed method were used in this study. The variations differ in the calculations used to derive a final cut score. The method and variations are described in detail below.

METHODOLOGY

This study tested the two variations of the proposed alternative examinee-centered standard setting strategy using data from a variety of grade levels and content areas. The proposed method is a modification of the Contrasting Groups method called the Modified Contrasting Groups method (MCGM). Two variations of this method were applied to these data. These variations differ in the way the cut score is calculated. The cut score derived from the MCGM was compared to the cut scores from Contrasting Groups, Borderline Group, and Modified Angoff methods that were previously conducted using the data.

Data

The data used in the study were actual standard setting results from a Midwestern school district. The school district is a K-12 district with about 1,500 students per grade level and approximately 20,000 total students. There are 23 elementary schools, 6 middle schools, and 4 high schools with approximately 1,400 certified classroom teachers across the district.

The assessments selected for this study have met the following requirements: are in a core subject area (mathematics, reading, science, and social studies), consist of dichotomously scored multiple-choice items written to measure the school district content standards, and have been through a formal standard setting study. The reliability of these assessments ranged from 0.87 to 0.94, calculated using coefficient alpha. Extensive validation work was conducted on these assessments, including substantial review by teachers and others to match the test content to the district's content standards. The results of the various reviews also indicate a strong relationship between the district's criterion referenced test and the commercially available norm referenced tests used by the school district.

For each assessment used in the study, there are several cut scores. The first is a Modified Angoff cut score used to distinguish the below proficient from the barely proficient student. The second is a Borderline Group cut score also used to distinguish the below proficient from the barely proficient student. The third is a set of three Contrasting Groups cut scores used to distinguish the four performance categories (below proficient, barely proficient, proficient, and advanced).

Table 1 lists the assessments that were selected to test the proposed method, along with the years they were developed, and the reliability coefficient for each assessment. These assessments met the following criteria: they were in one of the core subject areas (reading, math, science, or social studies), they contained over 150 matched teacher ratings and student scores with at least 2% of the students in the below proficient category, and there was a range of cut scores from a standard setting study. There are two files for the ninth grade reading and eleventh grade science assessments; the earliest form was retired and the later form replaced it operationally. Both forms met the criteria and were included in the study.

Modified Contrasting Groups Method

The calculations required to compute the proposed method (MCGM) are relatively simple. Two variations of this method were used in the study. They are referred to as MCGM 1 and MCGM 2. The two variations differ in the calculations used to compute the cut score. The rationale for developing this method was for school districts to be able to set mastery levels or cut scores that are affordable, that are not extremely complicated, that can be done in a reasonable amount of time, and that can place students into essentially the same performance

categories as they would be placed in if another method was used. Both variations of the method meet the criteria.

The first step in calculating the MCGM 1 cut score is to merge the teacher predictions (global classifications) with the students' actual performance on the assessment. Then, the median score for each performance category is calculated. At this point, there is a median score for each of the four performance categories; below proficient, barely proficient, proficient, and advanced. The next step is to calculate the average or mean of the two adjacent performance categories. To illustrate these calculations, data were created and used to generate Table 2. Based on the data in Table 2, the median score for the below proficient level is 72, barely proficient 81.5, proficient 84, and advanced 92. The next step is to calculate the cut score or the value that differentiates one performance level from the adjacent performance level. This is done by taking the average (mean) of the median score for the two adjacent performance levels. For example, to calculate the single value that determines the difference between the below proficient level and the barely proficient level, calculate the average (mean score) of the two medians. The median score for the below proficient level is 72 (in the example in Table 2) and the median score from the barely proficient level is 81.5. The MCGM 1 score that distinguishes the below proficient group from the barely proficient group would be 76.75. The same process is followed to calculate the cut score between barely proficient and proficient, and then between proficient and advanced.

The first step in calculating the MCGM 2 is the same as the first step for MCGM 1. The teacher predictions (global classifications) are merged with the students' actual performance on the assessment. The next step is to combine adjacent performance categories and calculate a

median value for the two categories. A median value would be calculated for the below proficient and the barely proficient, a median value for the barely proficient and proficient, and a median value for the proficient and advanced. These median values become the cut scores. The same data used to produce Table 2 was used to create Table 3. These data are the cut scores produced by MCGM 2. Based on the data in Table 3, the cut score that separates the below proficient from the barely proficient is 77. This value is the median value of the scores for the below proficient (68 & 76) and the barely proficient (78 & 85) combined. The median value that separates the below proficient from the proficient is 84, and the median value that separates the proficient from the advanced is 87.50.

The difference between MCGM 1 and MCGM 2 is the calculations used to arrive at the cut score. To calculate the cut score using MCGM 1, there are two calculations needed while MCGM 2 requires only one calculation. Both variations of the method are simple and require only basic mathematical skills to calculate a cut score.

The rationale for using both variations of the method is because MCGM 1 operates at the performance category level and MCGM 2 operates at the score level. The MCGM 1 weighs the performance categories equally. In the example used for Tables 2 and 3, there are eight scores in the proficient category and two scores in the advanced category. The MCGM 1 calculates a median for the proficient category and a median for the advanced group. Then, the mean of these two medians is used to calculate the cut score. In this example, the two scores for the advanced group are treated or considered equally as important as the 8 scores in the proficient group. The MCGM 2 puts all 10 scores together (the 8 proficient and the 2 advanced) and calculates the

median value for the 10 scores, treating each score equally and not placing value on the performance category.

Cut Score Comparison

The cut scores from the MCGM 1 and MCGM 2 were compared to the Modified Angoff cut score, the Borderline Group cut score, and the three Contrasting Groups cut scores. It should be noted that only the Contrasting Groups and the MCGM methods produced all three score points (below proficient and barely proficient, barely proficient and proficient, and proficient and advanced). The other methods, the Borderline Group and the Modified Angoff, only produced one cut score that distinguishes between the below proficient and the barely proficient student. This is a function of the “type” of student the school district chose to identify.

Each of the methods used relies on a different set of assumptions and produces different cut scores. As discussed earlier, the reason for using more than one standard setting method was to have a range of possible cut scores to choose from. The results from the proposed method were compared to the cut scores produced by the other three methods. It is possible to have a different cut score for each of the four methods, even when using the same judges and the same facilitator.

The first step in looking at how the MCGM 1 and MCGM 2 compare to the other methods is to investigate the ranking of the cut points. There are 16 assessments in the study and, for each assessment, there are five cut points (Angoff, BLG, CGM, and the MCGM 1 and MCGM 2) to distinguish the below proficient from the barely proficient. The rank ordering of these scores was examined to see if there are any consistent patterns. To address whether the MCGM 1 and MCGM 2 yield similar results compared to the other methods, the percent of

students who would score below the cut score was compared for each of the methods. Because the Contrasting Groups method is the basis for the new method and also has the same number of cut scores, the three cut scores from the MCGM 1 and MCGM 2 were compared to the Contrasting Groups method. The cut score from MCGM 1 and MCGM 2 that separated the below proficient from the barely proficient was also compared with the Angoff and Borderline Group cut scores to investigate differential impact.

Table 1

Data Set

| Grade | Content Area | Number of Items | Alpha | Data Points | Year |
|------------------|----------------|-----------------|-------|-------------|------|
| 1 st | Language Arts | 50 | .88 | 426 | 2000 |
| 2 nd | Math | 75 | .93 | 431 | 2000 |
| 4 th | Math | 88 | .94 | 606 | 1999 |
| 5 th | Science | 90 | .88 | 271 | 2001 |
| 5 th | Social Studies | 80 | .87 | 308 | 2001 |
| 6 th | Reading | 58 | .89 | 604 | 2005 |
| 7 th | Math | 69 | .91 | 440 | 1999 |
| 7 th | Reading | 75 | .90 | 227 | 1999 |
| 8 th | Science | 79 | .89 | 449 | 2001 |
| 8 th | Social Studies | 82 | .87 | 315 | 2001 |
| 9 th | Reading | 76 | .92 | 374 | 1999 |
| 9 th | Reading | 76 | .90 | 165 | 2003 |
| 10 th | Math | 70 | .91 | 353 | 1999 |
| 11 th | Science | 78 | .90 | 365 | 2001 |
| 11 th | Science | 82 | .90 | 213 | 2003 |
| 11 th | Social Studies | 85 | .94 | 297 | 2001 |

Table 2

Modified Contrasting Groups Method 1 (MCGM 1) Example

| Student | Teacher rating | Test Score | Median Score | Cut Score |
|---------|-------------------|------------|--------------|-----------|
| A | Below Proficient | 68 | | |
| B | Below Proficient | 76 | 72.0 | |
| C | Barely Proficient | 78 | | 76.75 |
| D | Barely Proficient | 85 | 81.5 | |
| E | Proficient | 73 | | |
| F | Proficient | 76 | | |
| G | Proficient | 77 | | |
| H | Proficient | 83 | | 82.75 |
| I | Proficient | 85 | | |
| J | Proficient | 90 | | |
| K | Proficient | 92 | | |
| L | Proficient | 93 | 84.0 | |
| M | Advanced | 90 | | 88.00 |
| N | Advanced | 94 | 92.0 | |

Table 3

Modified Contrasting Groups Method 2 (MCGM 2) Example

| Student | Teacher rating | Test Score | Median Score | Cut Score |
|---------|-------------------|------------|--------------|-----------|
| A | Below Proficient | 68 | | |
| B | Below Proficient | 76 | | |
| C | Barely Proficient | 78 | 77.00 | 77.00 |
| D | Barely Proficient | 85 | | |
| E | Proficient | 73 | | |
| F | Proficient | 76 | | |
| G | Proficient | 77 | | |
| H | Proficient | 83 | 84.00 | 84.00 |
| I | Proficient | 85 | | |
| J | Proficient | 90 | | |
| K | Proficient | 92 | | |
| L | Proficient | 93 | | |
| M | Advanced | 90 | 87.50 | 87.50 |
| N | Advanced | 94 | | |

RESULTS

Overview

This section contains the results for the study split into two parts. In part one, the cut scores produced by the four methods (i.e. Angoff, Borderline Group, Contrasting Groups, and the two variations of the Modified Contrasting Groups) are examined. Part one also addresses the following questions: How do the two variations of the Modified Contrasting Groups method cut scores that distinguishes the below proficient student from the barely proficient student compare to the cut scores produced by the Angoff, Borderline Group, and Contrasting Groups methods? Are there consistent patterns or systematic differences in the cut scores produced by these five methods?

Part two compares the two variation of the Modified Contrasting Groups method to the Contrasting Groups method and compares variation one (MCGM 1) to variation two (MCGM 2). Part two addresses the following questions: How do the three cut scores from the two variations of the Modified Contrasting Groups compare to the three cut scores from the Contrasting Groups method? How do the two variation of the Modified Contrasting Groups method compare with each other? In addition, part two examines the issue of differential impact.

Part One

This section investigates the Modified Contrasting Groups method cut scores for the two variations that distinguishes the below proficient student from the barely proficient student and compares those cut scores with the cut scores produced by the Angoff, Borderline Group, and Contrasting Groups methods for each of the 16 assessments. The cut scores were examined for consistent patterns or systematic differences across the four methods. For each of the 16

assessments (in Table 1), all five methods were applied to produce the one cut score that distinguishes the below proficient student from the barely proficient student. To determine how the Modified Contrasting Groups method (MCGM 1 and MCGM 2) compares to the other methods (Angoff, Borderline Group, and Contrasting Groups), the cut scores were ranked from lowest to highest based on their relative positions for each assessment.

The information presented in Table 4 provides the frequency of the ranking of the standard setting methods across the 16 assessments. The position column titles refer to the ranking of the method by assessment (position 1 is the lowest cut score and position 5 is the highest). For example, all five methods were applied to the second grade math assessment. The Contrasting Groups method produced a cut score of 54, the Angoff a 55, the MCGM 1 a 59.5, the MCGM 2 a 64.00, and the Borderline Group method produced a cut score of 66. The scores were then ranked from lowest to highest. In this example, the Contrasting Groups method would be in position 1 (lowest), the Angoff would be position 2, the MCGM 1 would be in position 3, the MCGM 2 would be in position 4, and the Borderline Group would be in the position 5. This ranking was done on all 16 assessments in order to examine the rankings of the cut scores.

Across the 16 assessments, the Contrasting Groups method was in position 1 (ranked lowest) ten times, in position 2 four times, in position 3 once, and in position 4 once. The Angoff method was in position 1 five times, position 2 ten times, and position 5 (the highest) once. The MCGM 1 was never in position 1 and was in position 2 twice, position 3 thirteen times, and position 4 once. The MCGM 2 was never in position 1 or in position 2; it was in position 3 twice, position 4 thirteen times, and position 5 once. The Borderline Group method was in position 1 once, was never in position 2 or 3, was in position 4 once, and in position 5 fourteen times. On

average, the Contrasting Groups method produced the lowest cut score, the Angoff method produced the second lowest cut scores, the MCGM 1 produced third highest cut scores, the MCGM 2 produced the fourth highest cut scores, and the Borderline Group method tended to produce the highest cut scores across these five methods.

In 13 out of the 16 assessments, the first two positions were held by either the Contrasting Groups or the Angoff method. For nine out of the 13 assessments, the Contrasting Groups method was in position 1 (yielded the lowest cut score) and the Angoff method was in position 2 four times. Of those 13 assessments, the difference in cut score was 1 point (six times), 2 points (once), 3 points (twice), 4 points (twice), 9 points (once), and 11 points (once). These results are presented in Table 5.

The results support the research regarding the impact of method choice on cut scores. There appears to be a distinct pattern in the ranking of these cut scores. On average, the Contrasting Groups method produces the lowest cut scores followed by the Angoff method. Based on the results presented above, 11 out of 16 assessments produced results consistent with the findings of Mills (1983) in which he found the Contrasting Groups and the Angoff methods produced similar cut scores when the same judges were used for both methods. The same judges provided the data used to compute all four cut scores for each grade and test used in this study.

As shown in Table 4, the results display a relatively consistent pattern in ranking the methods. There are some irregularities in the rankings for each of the four methods. For example, the Borderline Group method was in position 5 for 14 assessments, and in position 4 for one assessment, but was in position 1 for only one of the assessments.

The results presented in Part one, supports the following findings: (a) method has an impact on cut score, (b) there are semi-consistent patterns in the data, and (c) on average, the MCGM 1 and MCGM 2 produce a cut score higher than the Angoff and Contrasting Groups method but lower than the Borderline Group method.

Part Two

In this section, three cut scores (below proficient from barely proficient, barely proficient from proficient, and proficient from advanced) are compared across the MCGM 1, MCGM 2, and the Contrasting Groups methods. The results for the cut scores separating the below proficient from barely proficient and the barely proficient from proficient are consistent with the results in the previous section where, on average, the MCGM 1 and MCGM 2 produced a higher cut score than the Contrasting Groups method. The Contrasting Groups method produced a higher cut score that separates the proficient student from the advanced student compared to the two versions of the Modified Contrasting Groups method.

The difference between the cut scores on the MCGM 2 and the Contrasting Groups method are larger than the differences between the MCGM 1 and the Contrasting Groups method. One would expect there to be a difference due to the results presented in the previous section, where on average the MCGM 2 produced a higher cut score than the MCGM 1 and the Contrasting Groups method.

The practical implications of using either the MCGM 1 or the MCGM 2 to set cut scores on an assessment would be to identify a higher percentage of students as below proficient, barely proficient, and advanced, but a lower percentage of students as proficient. If the Contrasting Groups method was used, there would be a lower percent of students identified as below

proficient, barely proficient, and advanced, but a higher percent as proficient. The MCGM 1 identified fewer students as below proficient, more students as barely proficient, more students as proficient, and fewer students as advanced compared to the MCGM 2.

In order to compare the difference of performance group placement across the four performance levels for the 16 assessments, the percent of students in a performance category was calculated for each method and then the Contrasting Groups percent was subtracted from the Modified Contrasting Groups percent. A positive score indicates the Modified Contrasting Groups method placed more students in the performance category compared to the Contrasting Groups method. And a negative score indicates the Contrasting Groups method placed more students in the performance category compared to the Modified Contrasting Groups method. In comparing the MCGM 1 to the Contrasting Groups method, 14 of the 16 difference scores for the below proficient category were positive and the remaining two had no difference or a score of zero. Across the 16 difference scores for the barely proficient category, there were four negative scores and 12 positive scores. There are two positive scores and 14 negative scores for the proficient category and one negative score, two no difference (zero) scores, and 13 positive scores for the advanced category. The MCGM 1 placed more students in the below proficient, the barely proficient, and the advanced performance levels, and the Contrasting Groups method placed more students in the proficient performance category.

In comparing the MCGM 2 to the Contrasting Groups method, all 16 of the difference scores for the below proficient category were positive, and across the 16 difference scores for the barely proficient category, there were four negative scores and 12 positive scores. There was one positive score and 15 negative scores for the proficient category and one negative score, one no

difference (zero) score and 14 positive scores for the advanced category. The MCGM 2 placed more students in the below proficient, the barely proficient, and the advanced performance levels, and the Contrasting Groups method placed more students in the proficient performance category.

To this point, the two variations of the Modified Contrasting Groups method have been compared to the Contrasting Groups method. In order to investigate the score difference between the two variations of the Modified Contrasting Groups method, a difference score was calculated by subtracting the MCGM 2 cut scores from the MCGM 1 cut scores. If the difference score was positive, the MCGM 1 produced a higher cut score, and if the difference was negative, the MCGM 2 produced a higher cut score. The results for the first cut score (the below proficient from the barely proficient) ranged from -4.50 to 1.50. Fourteen out of the 16 differences were negative, indicating that the MCGM 1 cut scores were lower than the MCGM 2 cut scores. The second cut scores (the barely proficient from the proficient) were all negative, ranging from -5.00 to -0.50. And the third cut scores (the proficient from the advanced) ranged from -4.00 to 2.25; there was one negative score, three scores of zero, five scores greater than zero and less than one, and six scores between 1 and 2.25. The third difference score (advanced cut score) had the smallest differences between the two cut scores. The difference for each cut score was summed across the 16 assessments calculating a total difference between the two variations. The total difference score for cut score 1 was -27.25, the difference for cut score 2 was -30.00, and cut score 3 was 9.25. These summed difference scores indicate a large negative differences for cut score 1 and cut score 2 and a smaller positive difference for cut score 3. These results indicate the MCGM 2 produced a higher cut score value for cut score 1 and 2, but a lower cut score for

cut score 3. Although the pattern was consistent between these two variations (MCGM 1 produced lower value for the first and second cut score compared to the MCGM 2, and the MCGM 2 produced a lower value for cut score three compared to MCGM 1), the average for the difference scores was less than two points across the 16 assessments.

Several practically significant results were obtained from the study. First, method choice has an impact on cut score. Second, there are consistent patterns in the data. Next, the barely proficient cut scores produced by the MCGM 1 and MCGM 2 were higher than the cut scores from the Angoff and Contrasting Groups methods on average, but lower than those from the Borderline Group method. Also, the MCGM 1 and MCGM 2 placed more examinees in the below proficient, barely proficient, and advanced categories, but fewer examinees in the proficient performance category compared to the Contrasting Groups method. Further, there is a slight, but consistent difference between the MCGM 1 cut scores and the MCGM 2 cut scores based on these data.

The Modified Contrasting Groups method produced results that are consistent with the other well known standard setting methods, is not too computationally complex, can be done in a reasonable amount of time, and appears to be affordable.