

Nebraska Department of Education

State Writing Assessment: Grades 4, 8, and 11

Standard Setting Study

Final Report

Teresa Eckhout, M.A.

Chad W. Buckendahl, Ph.D.

Abdullah Ferdous, M.A.

Jennifer Nebelsick, B.A.

Buros Institute for Assessment Consultation and Outreach

A Division of the Oscar and Luella Buros Center for Testing

University of Nebraska-Lincoln

May 2005

Acknowledgments

We would like to acknowledge our appreciation to several people who assisted us with this Standard Setting Workshop. Sue Anderson, state writing assessment coordinator, Faith Camargo and Jackie Naber at the Nebraska Department of Education were very helpful in organizing this workshop. The success of the workshop was due, in large part, to their efforts.

We also want to thank the teachers whose recommendations contributed to the outcome of the standard setting workshop. Teachers who participated in the workshop provided classifications of student performance that resulted in the recommended cut scores. Sue Anderson selected the anchor papers and assisted in conducting the standard setting workshop. Without their efforts and diligence, there would have been no study.

Nebraska Department of Education

State Writing Assessment: Grades 4, 8, and 11

Standard Setting Study

Final Report

Introduction

The purpose of this report is to document the procedures and analyses undertaken to recommend performance standards for the Nebraska Department of Education's *State Writing Assessment* administered in grades 4, 8, and 11. The report summarizes the procedures and the results of the standard setting studies and provides recommendations for the establishment of a minimum passing, or cut score for each grade level.

Background

As part of the state assessment and accountability system, Nebraska administers Writing Assessments across the state at selected grade levels (4, 8, and 11). These assessments will be used to distinguish between students who have met the state writing standards and those that have not met the state writing standards and may need additional instruction in writing. Because the Writing Assessments are to be used to classify students in terms of their level of performance in writing, the Department of Education has recognized the importance of using psychometrically accepted methods for setting these performance standards (minimum passing scores).

The writing assessments give students an opportunity to provide a writing sample in response to a narrative (4th grade), descriptive (8th grade), or persuasive (11th grade) prompt. The student writes to the prompt that is provided in a given year. The prompts are scored holistically across six traits on an augmented 10-point scale. Two trained scorers score each paper and the student's score on the paper is the sum of the two scorers' scores. If the two scorers disagree by more than one score point, a third scorer scores the response and an average of the two closest scores is computed.

The purpose of this study was to provide a range of defensible cut scores to the Nebraska Department of Education (NDE) for the *State Writing Assessment* in grades 4, 8, and 11. This report focuses on the results of the standard-setting studies for these three grade levels. The report provides an overview of the methods and procedures for the study. It includes a recommendation for a range of cut scores within which NDE may identify a defensible cut score that will help decide which students in the state have met the writing content standards.

Methods and procedures

Overview of Procedures

Two methods for estimating a cut score were used. Each method relies on different assumptions and the use of these independent methods is intended to provide a more defensible range of possible cut scores within which NDE may determine the final cut score. These methods are a) an analytical judgment method and b) a professional judgment method. Each of these methods is described briefly below.

Each of the methods took place on April 21, 2005 in a workshop facilitated by the Buros Center for Testing. The workshop began with an orientation and training activity that included an extended discussion of the test specifications and a description and discussion of the following student performance levels that were developed by NDE and provided to us for use in the workshop.

1. Beginning: Writing is still under development. Extensive revision or editing would be necessary.
2. Proficient: Writing has more strengths than weaknesses. Some revision or editing would be necessary.
3. Advanced: Writing has many strengths. Only minor revision or editing would be necessary.

The Analytical Judgment Method

One standard setting method used in the standard setting studies is a modification of a method proposed by Hambleton and Plake (2000). This method required panelists to read a set of 50 papers (described below) and sort the papers into the three broad performance classifications defined above (Beginning, Proficient, or Advanced). After the initial sorting was completed, teachers identified three papers from the “Beginning” papers that were the closest to being in the next higher classification (Proficient). Teachers also identified three papers classified as Proficient that were closest to being Beginning. That is, teachers identified the three best papers in the Beginning classification category and the three worst papers in the Proficient category. Teachers did not know the scores on the papers; instead each paper had a code corresponding to a specific score. The cut score for a teacher was that teacher’s mean or median score of the six specific papers that were closest to next higher or lower category. The overall cut score was the average of the individual teacher cut scores.

The 50 papers were selected using a stratified random sample from the total set of papers. The sampled papers met the following criteria.

1. All score points had at least 2-3 papers with more papers with scores between 2- and 3+ being included.
2. Selected papers were scored correctly and accurately. The basis for scoring was not to be an issue.
3. Selected papers are written legibly and darkly enough that they could be photocopied.

The Professional Judgment method

This method entailed asking teachers to estimate the percentage of tested students in their classes this year who would be classified as Beginning (i.e., the percentage of students who would “fail” the test). This was done after all training activities and before participants completed the analytical judgment method. Special forms that also included demographic information to document the level of experience of the teachers were used for this method.

Specific Procedures

The Analytical Judgment Method

The standard setting workshop took place in Lincoln, NE at the Downtown Holiday Inn on April 21, 2005. A total of 43 teachers participated with 15 at 4th grade, 16 at 8th grade, and 12 at 11th grade. All teachers were currently teaching or had recently taught English at their respective grade level and had been exposed to the six-trait writing method used to score the Writing Assessment. Some of the teachers had also participated in the scoring process and/or participated in their respective grade level’s writing assessment standard setting previously.

Following introductory comments to all teachers across grade levels, an orientation and training session was conducted. This session articulated the purpose of the standard setting workshop and detailed the steps to be taken to complete the standard setting process. Training included a discussion of the performance categories (defined above) and a discussion of each of the six traits. Sue Anderson described the six writing traits to the participants. After the large group orientation, the panelists were subdivided into their grade level teams for further training.

In these grade level teams, there was an extensive discussion of the student who was Proficient as the target student. Using performance descriptions derived from the previous standard setting workshops, the teachers discussed the skills and performance characteristics of the target student in each of the six traits and holistically. They added to and modified the performance descriptions to better clarify their conception of the Proficient student. These descriptions for each grade level are included as Appendix A. Teachers were advised that they would be reading a large number of papers (50) and would be making holistic classifications for these papers. These holistic classifications would result in three stacks of papers, those that represented work that was a) Beginning, b) Proficient, and c) Advanced. The decision to place a paper in the stack was to be based on both the descriptions of the student who was Proficient considering holistically the student work that is consistent with the definitions of the broad classifications.

Panelists were provided a set of ten papers to practice the process. All panelists received the same papers to rate. These papers were selected such that there were papers that spanned the score range. Teachers made two sorting decisions using these ten papers. First the papers were classified as being Beginning, Proficient or Advanced. After that sorting decision was made, teachers identified the paper from the Beginning papers that was closest to being in the Proficient category. They also selected the paper in the Proficient category that was closest to being in the Beginning category. After these selections were made, there was a show of hands regarding how each paper was

classified. This was followed by a discussion of why teachers made their classification decision.

The training was followed by the operational analytical judgment method. The teachers in each grade level team were provided with copies of 50 papers selected as described above. Teachers made the initial sort into the three broad categories and then selected the three best of the papers classified as being Beginning and the three papers they felt were the worst of those in the Proficient category. Papers were collected and data entered.

Professional Judgment Method

After the analytical judgment method was completed the Professional Judgment method was undertaken. This method entailed having the teachers estimate the percent of students in their classes this year who would be classified as being Beginning.

Results

Analytic Judgment Method

The minimum passing scores are based on the judgments of panelists who made holistic ratings on the 50 papers. Each teacher's individual cut score was computed. This involved computing both a mean and a median of the six papers that were just above and just below the performance of the student who was "Barely" Proficient (the target student). The reason for examining both the average and the median is because the median would compensate for teachers whose scores were very inconsistent or extreme.

Grade 4

For this grade the recommended cut score using the mean was 3.73. The closest score point to this mean value would be 3.67. If the median were used, the value would be 4.00. The panelists' median recommended cut score (4.00), and a range of cut scores plus and minus 1 score point are shown in Table 1. The approximate percent of 4th grade Nebraska students who would be below the cut point is also shown in the Impact column.

For the professional judgment method, teachers' estimated percent of students who will be classified as being Beginning ranged from a low of 10% to a high of 30%, with an average (mean and median) of 20%. The closest score point associated with these impact values is 4.00.

Table 1. Analytic Judgment-based cut score and impact and cut scores and impacts within a one score point range for 4th grade.

<u>Range</u>	<u>Cut score</u>	<u>Impact (% below)</u>
1 Score Below	3.67	12.65
Average	4.00	17.41
1 Score Above	4.33	23.74

Grade 8

For this grade the recommended cut score using the mean was 3.74. The closest score point to this mean value would be 3.67. If the median were used, the value would be 4.00. The median cut score (4.00), and a range of cut scores plus and minus 1 score point are shown in Table 2. The approximate percent of 8th grade Nebraska students who would be below the cut point is also shown in the Impact column.

For the professional judgment method, teachers' estimated percent of students who will be classified as being Beginning ranged from a low of 2% to a high of 40%, with a mean of 14% and a median of 10%. The closest score points associated with these impact values range from 3.67-4.00.

Table 2. Analytic Judgment-based cut score and impact and cut scores and impacts within a one score point range for 8th grade.

<u>Range</u>	<u>Cut score</u>	<u>Impact (% below)</u>
1 Score Below	3.67	8.50
Average	4.00	11.22
1 Score Above	4.33	15.56

Grade 11

For this grade the recommended cut score using the mean was 4.31. The closest value using the mean would be 4.33. If the median were used, the value would be 4.00. The mean cut score (4.33) and a range of cut scores plus and minus 1 score point is shown in Table 3. The approximate percent of 11th grade Nebraska students who would be below the cut point is also shown in the Impact column.

For the professional judgment method, teachers' estimated percent of students who will be classified as being Beginning ranged from a low of 0% to a high of 15%, with a mean of 9% and a median of 10%. The closest score point associated with the mean and median impact values ranged from 4.00-4.33.

Table 3. Analytic Judgment-based cut score and impact and cut scores and impacts within a one score point range for 11th grade.

<u>Range</u>	<u>Cut score</u>	<u>Impact (% below)</u>
1 Score Below	4.00	7.57
Average	4.33	10.64
1 Score Above	4.67	13.94

Some teachers were inconsistent in their ordering of papers. That is, the highest scoring paper that was classified as one of the best in the Beginning category had a higher score than the highest scoring paper that was listed as one of the lowest scoring papers in the Proficient category, or vice versa. By using the median value, though, the impact of these inconsistencies is reduced.

Evaluation Data

At the conclusion of the workshop, panelists completed an evaluation form consisting of four parts. Part 1 focused on the orientation and training; Part 2 addressed the panelists' levels of comfort and confidence in their Professional Judgment ratings; Part 3 was parallel to Part 2, but focused on the confidence and comfort levels for the Analytical Judgments. Part 4 consisted of closed and open-ended items asking about the overall success of the workshop and about recommended changes that might be made to improve future workshops. Evaluation comments are shown in Appendix B.

Grade 4

Part 1: Training

On a scale ranging from 1 - 6, where 1 = Very Unsuccessful and 6 = Very Successful, all mean ratings fall between 5.7 and 5.8. (Orientation mean = 5.8, Training on Analytical Judgments Method mean = 5.7, Description of target students mean = 5.7, Practice with Analytical Judgments Method mean = 5.8, and Overall Training mean = 5.8).

Panelists also rated the adequacy of the time provided for training and orientation. On a six-point scale, where 1 = Totally Inadequate and 6 = Totally Adequate, all mean rating exceeded 5.00. (Orientation mean = 5.8, Training on Analytical Judgments Method mean = 5.8, Description of target student mean = 5.7, Practice with Analytical Judgments Method mean = 5.8, and Overall Training mean = 5.6).

When asked to rate the amount of time allocated to training, the mean rating was 2.1 where a value of 2 was “The right amount of time was allocated to training.” Only one of the thirteen teachers felt that too much time was allocated to training.

Part 2: Professional Judgment Method

The mean panelists' confidence and comfort in making estimate using the Professional Judgment method were 3.5 and 3.8, respectively on a four-point scale (1 = Not Confident and 4 = Confident).

The mean rating for the allocation of time for making the professional judgments was 3.6 -- no teacher indicating that the time was insufficient. All teachers felt that there was enough or more than enough time for making these judgments.

Part 3: Analytical Judgments Method

The mean panelists' confidence in classifying papers into three categories was 3.6 on a four-point scale (1 = Not Confident and 4 = Confident). None of the teachers felt “not confident” about their judgment. The mean Comfort rating on the same 4-point scale (1= Not Comfortable and 4= Comfortable) for this process was also 3.6.

The final item in Part 3 asked about the adequacy of time allocated for making the analytical judgments. On the four-point scale (1 = More time needed and 4 = More than enough time was allotted), the mean rating was 3.7. None had said that more time was needed or there was just barely enough time.

Part 4: Overall

The first item in Part 4 asked about the teachers' confidence in the standard that would result from this process. The mean confidence was 3.8 on a four-point scale (1 = Not at all Confident and 4 = Confident). Thus, overall teachers were "Confident" about the appropriateness of the passing standard. None of the teachers had a confidence rating of less than 3.

Two questions asked panelists to rate the success and organization of the workshop (1 = Totally Unsuccessful and 4 = Totally Successful). The mean ratings on these items were 3.7 and 3.9, respectively.

Panelists were given an opportunity to provide comments they felt would be helpful in planning future standard setting studies. Five teachers made comments. The comments are attached in Appendix B.

Grade 8

Part 1: Training

On a scale ranging from 1 - 6, where 1 = Very Unsuccessful and 6 = Very Successful, all mean ratings exceeded 5.0 except for the Description of target student mean, which was 4.8. (Orientation mean = 5.3, Training on Analytical Judgments Method mean = 5.1, Description of target students mean = 4.8, Practice with Analytical Judgments Method mean = 5.1, and Overall Training mean = 5.1).

Panelists also rated the adequacy of the time provided for training and orientation. On a six-point scale, where 1 = Totally Inadequate and 6 = Totally Adequate, all mean rating exceeded 5.00. (Orientation mean = 5.7, Training on Analytical Judgments Method mean = 5.4, Description of target student mean = 5.1, Practice with Analytical Judgments Method mean = 5.2, and Overall Training mean = 5.5).

When asked to rate the amount of time allocated to training, the mean rating was 1.9 where a value of 1 was "Too little time was allocated to training" and a value of 2 was "The right amount of time was allocated to training." Only two of the sixteen teachers felt that too little time was allocated to training.

Part 2: Professional Judgment Method

Panelists' mean confidence and mean comfort in making judgment was the same 3.4 on a four-point scale (1 = Not Confident and 4 = Confident).

The mean rating for the allocation of time for making the professional judgments was 3.4 -- no teacher indicating that the time was insufficient. All teachers felt that there was enough or more than enough time for making these judgments.

Part 3: Analytical Judgments Method

The average panelists' confidence in classification of papers into three categories was 3.6 on a four-point scale (1 = Not Confident and 4 = Confident). None of the teachers felt "not confident" about their judgment. The mean Comfort rating on the 4-point scale (1 = Not Comfortable and 4 = Comfortable) for the process of making the analytic judgments was also 3.6 with no responses of 1 and 2.

The final item in Part 3 asked about the adequacy of time allocated for making the analytical judgments. On the four-point scale (1 = More time needed and 4 = More than enough time was allotted), the mean rating was 3.4.

Part 4: Overall

The first item in Part 4 asked about the teachers' confidence in the standard that would result from this process. The mean level of confidence was 3.4 on a four-point scale (1 = Not at all Confident and 4 = Confident). Thus, the mean indicates that overall, teachers were Confident about the appropriateness of the passing standard. None of the teachers had a confidence rating of less than 3.

Two questions asked panelists to rate the success and organization of the workshop (1 = Totally Unsuccessful and 4 = Totally Successful). The mean ratings on these items were 3.3, and 3.4 respectively.

Panelists were given an opportunity to provide comments they felt would be helpful in planning future standard setting studies. Eight teachers made comments. The comments are attached in Appendix B.

Grade 11

Part 1: Training

On a scale ranging from 1 - 6, where 1 = Very Unsuccessful and 6 = Very Successful, all mean ratings fall between 5.1 and 5.7. (Orientation mean = 5.6, Training on Analytical Judgments Method mean = 5.4, Description of target students mean = 5.1, Practice with Analytical Judgments Method mean = 5.7, and Overall Training mean = 5.5).

Panelists also rated the adequacy of the time provided for training and orientation. On a six-point scale, where 1 = Totally Inadequate and 6 = Totally Adequate, all mean rating exceeded 5.00. (Orientation mean = 5.6, Training on Analytical Judgments Method mean = 5.5, Description of target student mean = 5.1, Practice with Analytical Judgments Method mean = 5.5, and Overall Training mean = 5.7).

When asked to rate the amount of time allocated to training, the mean rating was 2.3 where a value of 2 was "The right amount of time was allocated to training." All of the twelve teachers felt the right amount of time was allocated to training.

Part 2: Professional Judgment Method

The mean panelists' confidence in making estimate was 3.8 on a four-point scale (1 = Not Confident and 4 = Confident). The rating of comfort level was however 3.9.

The mean rating for the allocation of time for making the professional judgments was 3.6 -- no teacher indicating that the time was insufficient. All teachers felt that there was enough or more than enough time for making these judgments.

Part 3: Analytical Judgments Method

The mean panelists' confidence in classifying papers into three categories was 3.7 on a four-point scale (1 = Not Confident and 4 = Confident). None of the teachers felt "not confident" about their judgment. The mean Comfort rating on the 4-point scale (1=

Not Comfortable and 4= Comfortable) for the process of making the analytic judgments was 3.8 with only one teacher responding a 2.0.

The final item in Part 3 asked about the adequacy of time allocated for making the analytical judgments. On the four-point scale (1 = More time needed and 4 = More than enough time was allotted), the mean rating was 3.9. All responses were either 3 or 4.

Part 4: Overall

The first item in Part 4 asked about the teachers' confidence in the standard that would result from this process. The mean level of confidence was 3.7 on a four-point scale (1 = Not at all Confident and 4 = Confident). Thus, overall teachers were "Confident" about the appropriateness of the passing standard. None of the teachers had a confidence rating of less than 3.

Two questions asked panelists to rate the success and organization of the workshop (1 = Totally Unsuccessful and 4 = Totally Successful). The mean ratings on these items were both 3.5.

Panelists were given an opportunity to provide comments they felt would be helpful in planning future standard setting studies. Eleven teachers made comments. The comments are found in Appendix B.

Conclusions and Recommendations

The teachers' recommendations for each grade level are based on considerations of both methods described in the body of this report. For 4th grade, both methods produced recommendations for cut scores of 4.00. If a cut score of 4.0 is adopted, approximately 17% of Nebraska's 4th grade students would be identified as having not met the writing standards and needing extra assistance.

At the 8th grade, both methods yielded recommendations that support a cut score of 4.00. If NDE adopts 4.00 as the recommended cut score, approximately 11% of Nebraska's 8th grade students would be identified as having not met the writing standards and needing extra assistance.

At the 11th grade, both methods produced recommendations that support a cut score of 4.33. If a cut score of 4.33 is adopted, approximately 11% of Nebraska's 11th grade students would be identified as having not met the writing standards and needing extra assistance.

References

Plake, B. S., & Hambleton, R. K. (2000). A standard-setting method designed for complex performance assessments: Categorical assignments of student work. Educational Assessment, 6(3), 197-215.

Appendix A

Performance Level Descriptions for Grades 4, 8, and 11

High School (Grade 11): Defining Proficiency for the 6 Traits

Advanced

Stronger	Weaker
Word Choice	
Organization	
Conventions	
Ideas	
Voice	
Sentence Fluency	

Proficient

Stronger	Weaker
Word Choice	Voice
Organization	Sentence Fluency
“Ideas” and “Conventions” - somewhere between strong and weak	

Beginning

Stronger	Weaker
May attempt “Word Choice”	Organization
Voice	Sentence Fluency
	Conventions
	Ideas – sketchy

Stronger	Weaknesses
<u>Ideas</u>	
Has details	Not fully developed
Commitment to the writer's position	
Clear	
<u>Organization</u>	
Attempt at paragraphing structure	
Sequencing logical	Not always complete
Evidence of Beginning, Middle, and End	Transition not always there
Clincher	
<u>Voice</u>	
Some conviction	Sometimes forced or mechanical
Appropriate tone for the audience	
<u>Word Choice</u>	
Clear and somewhat persuasive – appropriate for 11 th grade	Some trite, non-specific language
Details are attempted	Not fully developed or used appropriately in context
<u>Sentence Fluency</u>	
Some variety in structure and length	Phrasing may be more mechanical (related to flow)
Some transitions	Needs more development
Mostly flowing	
<u>Conventions</u>	
Errors don't detract from readability (some editing)	Few correct uses of stylistic punctuations.
Basic punctuation, usage (Uses ending punctuations -- . ? !)	
Reader can still understand the message	Some errors in spelling and grammatical usage

Grade 8: Defining Proficiency for the 6 Traits

Advanced

Stronger	Weaker
Word Choice	
Organization	
Conventions	
Ideas	
Voice	
Sentence Fluency	

Proficient

Stronger	Weaker
Word Choice	Voice
Organization	Sentence Fluency
Conventions	
“Ideas” - somewhere between strong and weak	

Beginning

Stronger	Weaker
May attempt “Word Choice”	Organization
Voice	Sentence Fluency
	Conventions
	Ideas – sketchy

Stronger	Weaknesses
<u>Ideas</u>	
Relevant to topic with some details	Not fully developed
Creative	Not always apparent
Clear	
<u>Organization</u>	
Some paragraphs	
Sequencing logical	Not always complete
Evidence of Beginning, Middle, and End	Transition not always there
Hook & Conclusion are attempted	
<u>Voice</u>	
Some personality with audience consideration	Sometimes forced or mechanical
<u>Word Choice</u>	
Word used correctly	Some trite, non-specific language
Some sensory details are apparent	Not fully developed or used appropriately in context
<u>Sentence Fluency</u>	
Some variety in structure and length	Phrasing may be more mechanical (related to flow)
Some transitions	Needs more development
Mostly flowing	
<u>Conventions</u>	
Errors don't detract from readability (some editing)	Few attempts to use stylistic punctuations.
Basic punctuation, usage (Uses ending punctuations -- . ? !)	
Reader can still understand the message	

Grade 4: Defining Proficiency for the 6 Traits

Advanced

Stronger	Weaker
Word Choice	
Organization	
Conventions	
Ideas	
Voice	
Sentence Fluency	

Proficient

Stronger	Weaker
Word Choice	Voice
Organization	Sentence Fluency
Conventions	
“Ideas” - somewhere between strong and weak	

Beginning

Stronger	Weaker
May attempt “Word Choice”	Organization
Voice	Sentence Fluency
	Conventions
	Ideas – sketchy

Stronger	Weaknesses
<u>Ideas</u>	
Has details	Not fully developed
Creative	Not always apparent
Clear	
<u>Organization</u>	
Attempt at paragraphing	
Sequencing logical	Not always complete
Evidence of Beginning, Middle, and End	Transition not always there
Hook & Conclusion are attempted	
<u>Voice</u>	
Some personality	Sometimes forced or mechanical
<u>Word Choice</u>	
Clear and somewhat descriptive – appropriate for 4 th grade	Some trite, non-specific language
Sensory details are attempted	Not fully developed or used appropriately in context
<u>Sentence Fluency</u>	
Some variety in structure and length	Phrasing may be more mechanical (related to flow)
Some transitions	Needs more development
Mostly flowing	
<u>Conventions</u>	
Errors don't detract from readability (some editing)	Few attempts to use stylistic punctuations.
Basic punctuation, usage (Uses ending punctuations -- . ? !)	
Reader can still understand the message	

Appendix B. Comments from Standard Setting Workshop Evaluation

Grade 4

- Awesome job!
- It may have been helpful to keep the papers from the first set of 10 handy to go back to as a reference point.
- I felt this was very well organized. I did find it difficult to narrow down to three papers in each category, would've been easier for me to leave it at 5.
- It was a great learning experience. I hope to participate again.
- Thank you for the opportunity to do this.

Grade 8

- The only part I would change would be not to give us 30 minutes to determine characteristics of Beginning students. 15 minutes would have been adequate.
- I would consider using the term "not proficient" or "beginning/progressing" rather than beginning so that people who also did the actual scoring or are familiar internally with the state's rubric are not influenced (either way) by this term used differently in the writing (scoring context) with these same papers. Beginning has a definite connotation compared to 'not proficient'. I would also suggest having the papers numbered. 1, 2, 3, etc. rather than the student # so it would be easier to refer to the same paper. Overall, it was well done and emphasized the "barely proficient" student which was good.
- This is a very insightful process.
- I feel teachers who work with a specific grade level can better evaluate where the cut off should be.
- This is a good activity. I wish more educators would agree to help with this. May be moving location to the middle of the state?
- Allow more introduction - I didn't know who people were, and it would be beneficial to know people and where they're from, level of expertise. This could be something quick and painless just so we know who we're working with. You might want to construct groups based on a variety of things (age, teaching background, location) so there's a solid mix to groups when discussing the analytical judgments and how to refine the differences between beginning and proficient. This was an extremely educational and rewarding experience! I enjoyed the format and appreciated the organization of a complex task! Excellent work!

Thanks for a fabulous day!

- Overview/introduction was good. Could have had more specific directions when we practiced and shared choices. Would have been nice to have a hard copy of our 6 traits charts - lighting and hand writing made them hard to read and use - may be a bit of 'retaining' before using them since it came right after lunch break.
- More time spent defining the target students what was beginning.

Grade 11

- On purple sheet (char. Of proficient) the "stronger" & "weaknesses" columns were confusing. This is because much of 6 trait training uses those terms differently. Could start on time & go a bit faster. Overall, good job. Josh[?] did a nice job!
- I would like to the essays more mixed. It became more of a guessing game with them in somewhat of an order.
- We spent a good deal of time on semantics when defining targets - necessary to discuss, yes; need for exact wording, probably not. Understanding, or having consensus on the essence is sufficient.
- Unfortunately because most of us are under contract our time is not paid. However, it is greatly appreciated that expenses incurred are covered by Buros.
- Chad - thanks for being patient with our gabby, over - analytical group. You controlled the group discussions well.
- The only question I have at this time is, "how were these papers selected"? Was it completely random or were they selected because of the categories they represented? That question might be addressed in the future.
- Well done! This is a tough process, but I do see consistent improvement in student writing each year so it must be successful.
- It helped that I have been involved in this process (or similar) a number of times. It also helped having been involved in the state screening process. It was good day.
- Start on time, please. Pay participants for travel time. Opening overview too lengthy. Very well explained, friendly staff, no pressure, and good food.

- This was well organized and very helpful to a beginner.
- Those who participate in scoring papers more easily grasp the standard setting process and come to consensus. Recruit from the scorers.