# Arizona

## Arizona English Language Learner Assessment (AZELLA)

### Standard Setting Report: Kindergarten Placement Test

Submitted to the
Arizona Department of Education
August 2013

## Table of Contents


Executive Summary ..... 1
General Standard Setting Procedures ..... 3
Panels ..... 3
Performance Level Descriptors ..... 3
Methodology Overview ..... 4
Data ..... 4
Security ..... 5
Staff ..... 5
Detailed Standard Setting Procedures ..... 6
Opening Session ..... 6
Introduction ..... 6
Performance Level Descriptors ..... 7
Experience the Test ..... 7
Development of Borderline Student Descriptors ..... 7
Standard Setting Methodology Training ..... 9
Practice Round Ratings ..... 9
Round 1 Ratings ..... 10
Round 2 Ratings ..... 10
Round 3 Ratings ..... 11
Standard Setting Evaluation Form ..... 12
Performance Level Descriptors Refinement ..... 12
Standard Setting Results ..... 12
References ..... 17
Appendix A: Standard Setting Participants ..... 18
Appendix B: Table Leader Information Sheet ..... 19
Appendix C: Table Leader Training Slides ..... 20
Appendix D: Performance Level Descriptors ..... 26
Appendix D.1: Preliminary Performance Level Descriptors ..... 26
Appendix D.2: Borderline Student Descriptors ..... 28
Appendix D.3: Instruction for the Performance Level Descriptor Refinement 29
Appendix D.4: Final Performance Level Descriptors after Refinement Process ..... 30
Appendix E: Standard Setting Script ..... 32
Appendix F: Opening Session Presentation Slides ..... 45
Appendix G: Breakout Session Presentation Slides ..... 56
Appendix H: Standard Setting Methodology Training Slides ..... 72
Appendix I: Standard Setting Panelist Readiness Form ..... 77
Appendix J: Example Standard Setting Rating Sheet ..... 78
Appendix K: Standard Setting Evaluation Forms with Responses ..... 79
Appendix L: Standard Setting Results ..... 83
Appendix L.1: Rating Distribution at Round 1 ..... 83
Appendix L.2: Rating Distribution at Round 2 ..... 96
Appendix L.3: Round by Round Raw Score Cut Summaries ..... 109

## List of Tables

Table 1. Final Recommended Raw Score Cuts ..... 2
Table 2. Final Scale Score Ranges ..... 2
Table 3. Proficiency Level Distribution for All PHOLTE Students ..... 2
Table 4. Median Raw Score Cuts by Round ..... 13
Table 5. Scale Score Cuts by Round ..... 15
Table 6. Impact Data for PHLOTE Students Based on Median Cuts by Round ..... 15
Table L.3.1 Minimum, Median, and Maximum Raw Cut Scores by Round .. 109
Table L.3.2 Median, Minimum, and Maximum Raw Cut Scores by Group after Round 1 ..... 109
Table L.3.3 Median, Minimum, and Maximum Raw Cut Scores by Group after Round 2 ..... 110
Table L.3.4 Median, Minimum, and Maximum Raw Cut Scores by Group after Round 3 ..... 110

## List of Figures

Figure 1. Graphical representation of borderline students. ..... 8
Figure 2. A sample worksheet for a multiple point item for the Proficiency cut. ..... 9
Figure 3. Raw score cut distributions of cuts at Round 1. ..... 13
Figure 4. Raw score cut distributions of cuts at Round 2. ..... 14
Figure 5. Raw score cut distributions of cuts at Round 3. ..... 14
Figure 6. Proficiency level distributions for PHLOTE, non-PHLOTE, and subgroups of PHLOTE by gender, and Hispanic status at the final round. .... 16

# Executive Summary

This document provides information about the procedures that were implemented to establish performance standards for the Arizona English Language Learner Assessment (AZELLA) Kindergarten Placement Test, developed by the Arizona Department of Education (ADE). The standard setting took place on July 11 and 12, 2012 at the Black Canyon Conference Center in Phoenix, Arizona. The AZELLA Kindergarten Placement Test is designed to assess the level of English proficiency for entering kindergarten students who have reported a primary or home language other than English. Students who do not have sufficient English proficiency to do regular classwork in English will be provided with language services.

The AZELLA Kindergarten Placement Test is an individually administered test that takes approximately 20 minutes. The blueprint of the AZELLA Kindergarten Placement Test was established based on the English Language Proficiency Standards focusing primarily on the Listening and Speaking sections at the lower skill levels and on the Arizona Early Learning Standards. The performance level descriptors (PLDs) for the AZELLA Kindergarten Placement Test were also developed prior to the standard setting.

The AZELLA Kindergarten Placement Test classifies the students into the following English proficiency levels:

1) Pre-emergent/Emergent
2) Basic/Intermediate
3) Proficient

During the standard setting meeting, the standard setting panelists engaged in the following activities. Please refer to the *Detailed Standard Setting Procedures* section for the details of each activity.

1. Opening session
2. Review performance level descriptors
3. Experience the test
4. Develop borderline student descriptors
5. Standard setting methodology training
6. Practice round of ratings
7. Round 1 ratings
8. Round 1 feedback and discussion
9. Round 2 ratings
10. Round 2 feedback and discussion
11. Round 3 ratings
12. Performance level descriptor refinement
13. Complete standard setting evaluation

The performance standards for the AZELLA Kindergarten Placement Test were established using the Modified Angoff method (Angoff, 1971; Plake &

Cizek, 2012). The final results after three rounds of ratings are presented below. The final recommended raw score cuts and percentage of points required for the *Basic/Intermediate* and *Proficient* levels are summarized in Table 1. To achieve the *Basic/Intermediate cut*, students should score at least 22 points out of 42 maximum possible points, which is approximately 52% of the points. For the *Proficient* cut, students should make 32 points or above (approximately 76% of the points).

Table 1.

*Final Recommended Raw Score Cuts*

| Basic/Intermediate | | Proficient | |
|---|---|---|---|
| Raw Score | % of Points | Raw Score | % of Points |
| 22 | 52% | 32 | 76% |

The final scale score ranges for each proficiency level are presented in Table 2. The scale score for the AZELLA Kindergarten Placement Test ranges from 100 to 300.

Table 2.

*Final Scale Score Ranges*

| Pre-emergent/Emergent | Basic/Intermediate | Proficient |
|---|---|---|
| 100-207 | 208-234 | 235-300 |

The percentage of students at each proficiency level based on the final recommended cut scores is summarized in Table 3. The percents in the table are based on students with a Primary Home Language Other Than English (PHLOTE) who participated in the field test. Based on these cut score ranges, approximately 60% of PHLOTE students would be *Proficient*. The percentage of PHLOTE students at *Pre-emergent/Emergent* and *Basic/Intermediate* would be 15% and 25%, respectively.

Table 3.

*Proficiency Level Distribution for All PHOLTE Students*

| Pre-emergent/Emergent | Basic/Intermediate | Proficient |
|---|---|---|
| 15% | 25 % | 60% |

# General Standard Setting Procedures

## *Panels*

The ADE invited Arizona educators to participate in the standard setting for the AZELLA Kindergarten Placement Test. Arizona educators have had experience with the curriculum, content, and performance standards, as well as with the student groups and grade level for which standards were set. Participating educators represented the diverse demographics of students educated across the state. The input of these educators ensured that standard setting reflected what students should know and be able to do.

The ADE recruited panelists based on the following characteristics:

- Be subject matter experts
- Understand the examinee population
- Understand what contributes to item difficulty
- Have knowledge of the instructional environment
- Appreciate the consequences of the standards
- Be representative of all the stakeholder groups

There was one panel that consisted of 13 panelists for the meeting (Please refer to Appendix A for the panelists' background information.). The panelists were divided into three table groups of 4 or 5 panelists each. One panelist from each table was assigned as a table leader. Prior to the standard setting meeting, the table leaders met on the morning of the meeting to go over the table leader information sheet (Appendix B) and table leader PowerPoint training (Appendix C) for their roles and responsibilities during the standard setting.

## *Performance Level Descriptors*

The Performance Level Descriptors (PLDs) for the AZELLA Kindergarten Placement Test were preliminarily developed in July 2012 prior to the standard setting meeting (Appendix D.1). There are three English language proficiency levels for this assessment:

1) *Pre-emergent/Emergent*
2) *Basic/Intermediate*
3) *Proficient*

There are two sections on the PLD document. On the top of the document, there is a concise description about what a student at each proficiency level should be able to do. The bullets at the bottom provide a highlighted list of Performance Indicators (PIs) by domain for each proficiency level. The PLDs were reviewed and refined by the panelists after the standard setting. The top half of the final PLDs is on the AZELLA Kindergarten Placement Test student score report.

## *Methodology Overview*

To date, a number of methods for standard setting have been proposed and established. The Modified Angoff method (Angoff, 1971; Plake & Cizek, 2012), approved by the Technical Advisory Committee formed by ADE, was used to establish the performance standards for the AZELLA Kindergarten Placement Test. The original Angoff method has been modified in many ways, and a version based on the original method is typically referred to as the modified Angoff method. The modified Angoff method has been widely used in education and is recognized as a well-known standard setting method (Plake & Cizek, 2012). In the standard setting for the AZELLA Kindergarten Placement Test, the panelists were asked to determine the percentage of students at the barely proficiency level for one point items and the mean score of barely proficient students for multiple point items. The panelists made the rating for each item for the *Proficient* level and the *Basic/Intermediate* level. In other words, they rated 38 items for 2 proficiency cuts, which resulted in 76 ratings for one round. The ratings for all items by each panelist were aggregated by summing them up, which would be an expected raw score for the student at the borderline of each proficiency level. This served as the recommended raw score cut for each proficiency level for each panelist. Then, the median raw score cut across all panelists was considered as the committee's recommended cut for the proficiency level. The panelists engaged in three rounds of ratings. During Rounds 1 and 2 the panelists rated each item and at Round 3 they were given the option of changing the cut score resulting from their Round 2 ratings to use as their Round 3 recommended raw score cut. The recommended raw score cuts for the *Proficient* level and the *Basic/Intermediate* level were then mapped on to the reporting scale using the raw score to theta table to determine the recommended cuts on the reporting scale. Please refer to *Standard Setting Methodology Training* under *Detailed Standard Setting Procedures* for the training the panelists received.

## *Data*

Data collected from the field test administration in February 2012 was used for all analyses. The Rasch model (Rasch, 1960) was used for one point items, including one point oral response items, one point physical response items, and the Partial Credit model (Masters, 1982) was used for multiple point items, including two point physical response items and three point oral response items, to scale the AZELLA Kindergarten Placement Test. The operational form of the test, which consisted of 38 items, was constructed after the field test administration. The operational form was reviewed by the panelists to set the performance standards for the test during the meeting. A raw score to theta conversion table was developed for the operational form as a part of the calibration and scaling of the new test. The raw score frequency distributions for overall PHLOTE students as well as subgroups were used to determine what percent of students were expected to fall into

each proficiency level. The impact data was presented to the panelists after Round 2 of the standard setting meeting. Please refer to *Detailed Standard Setting Procedures* for more details about the impact data.

## *Security*

Maintaining the security and confidentiality of test items and student responses is of utmost importance. Pearson has experience providing for and working in secure environments and has established procedures for maintaining the confidentiality of student responses and the security of test forms and materials. These procedures were implemented at each standard setting meeting session.

As the panelists arrived, Pearson staff registered them and had them sign a security agreement form. Upon registration, each panelist received a unique identification number. All materials received throughout the standard setting meeting possessed the identification numbers, so strict inventory control could be implemented and maintained. The facilitator had the panelists sign-in all materials at the end of each day to make sure that the secured materials were returned at the dismissal of the meeting.

## *Staff*

The following psychometric and content staff supported the AZELLA Kindergarten Placement Test standard settings:

Dr. Steven Fitzpatrick received his Ph.D. in Educational Psychology with a specialization in Quantitative Methods from the University of Texas at Austin and has been employed at Pearson since 2002. He is a Principal Research Scientist and serves as the lead Research Scientist on the AZELLA program. He has nearly 30 years of experience in the psychometric field and is nationally renowned for his extensive experience and technical skill. Dr. Fitzpatrick oversaw the standard setting and data analysis in support of the standard setting activities during the standard setting meeting. He also served as the facilitator of the meeting.

Ms. Beverly Nedrow received her M.S. in Curriculum and Instruction with specializations in English as a Second Language and Reading from Texas A & M at Corpus Christi and has been employed with WestEd for 6 years. She has taught English Language Learners from the elementary level through college. She is the Content Lead on the AZELLA program. She has nearly 25 years of experience in the development of English Language Arts and English Language Learner assessments and is nationally recognized for her content expertise. Ms. Nedrow participated as the content expert in support of the standard setting activities during the standard setting meeting.

Dr. Hirotaka Fukuhara received his Ph.D. in Measurement and Statistics from the Florida State University and has been employed at Pearson since 2011.

He is a research scientist and serves on the AZELLA program. Dr. Fukuhara served as the data analyst during the standard setting meeting.

Psychometric and content staff members from the ADE were also available.

## Detailed Standard Setting Procedures

In this section, a more detailed description of activities that took place throughout the standard setting meeting is provided. The facilitator led all activities according to the script approved by the ADE (Appendix E). Please refer to Appendix F for the presentation slides of the opening session, Appendix G for the presentation slides of the break out session, and Appendix H for the presentation slides of the standard setting methodology training.

### *Opening Session*

The standard setting meeting began by welcoming the panelists. The meeting facilitator introduced staff members of ADE, Pearson, and WestEd, explained the roles of ADE, Pearson, WestEd, and the panelists, explained the purpose of meeting, and provided a brief overview of the standard setting process. The ADE also gave a presentation about the historical background of the test and the purpose of the test. Logistics and security of the meeting were also addressed.

### *Introduction*

After a break, the panelists were asked to introduce themselves by describing their educational and professional backgrounds. The following questions were provided to aid the panelists in introducing themselves:

- Name
- Where are you from?
- How long you have been in your current position/field?
- What educational roles you have fulfilled?
- Have you participated in a standard setting before?
- Tell us something interesting about yourself

When the introductions were completed, the facilitator went over the agenda for the rest of the day for panelists to understand what needed to be accomplished on Day 1. The facilitator mentioned to the panelists that the time allocated for each activity on the agenda might deviate from what might be actually spent, depending on the pace of activities and additional discussion that might be required for some activities.

## *Performance Level Descriptors*

Next, the panelists had an opportunity to review the PLDs. The aim of this exercise was for the panelists to become familiar with the PLDs and to have a group discussion about skills described for each proficiency level. The facilitator reminded the panelists that they would have more time for in-depth discussion about the PLDs later on.

## *Experience the Test*

After a brief discussion about the PLDs, the Panelists were asked to go over the AZELLA Kindergarten Placement Test. An efficient way to help panelists become familiar with test content and gain the appreciation of the test is to have them actually take the test under simulated testing administration conditions. However, since the AZELLA Kindergarten Placement Test is individually administered, it is difficult to mimic a live administration. Thus, instead of taking the test, the panelists were asked to walk through the items on the test independently. The score rubrics were provided to the panelists to understand how each item would be scored. After the panelists spent approximately 45 minutes to complete the task, the following questions were brought up to them for discussion:

- What are your general impressions about the test?
- Did the test generally cover the depth and breadth of the language proficiency standards?
- Does the test generally have a range of item difficulties (e.g., easier items, moderate items, difficult items)?

Although some discussion about individual items occurred, the facilitator made sure that the panelists focused on the discussion around the questions above. The facilitator also encouraged the panelists to record any comments about the items to share with the ADE.

## *Development of Borderline Student Descriptors*

After lunch, the facilitator led the panelists toward discussion about borderline students, defined as the students just barely at the proficiency level. First, the panelists revisited the PLDs for more in-depth discussion by identifying three or four key characteristics that distinguish performance at a given proficiency level from that of adjacent proficiency levels for a skill set.

Next, the panelists at each table were asked to develop concrete descriptions of what students at *Proficient* or *Basic/Intermediate* should be able to do. For the *Proficient* level, the table groups were asked:

- What should the students at *Proficient* be able to do?
- What skills should the students at *Proficient* possess?

- What should the students at *Proficient* know?
- What language skills are necessary to access the mainstream curriculum in English?

Similarly, for the *Basic/Intermediate* level, the table groups were asked:

- What should the students at *Basic/Intermediate* be able to do?
- What skills should the students at *Basic/Intermediate* possess?
- What should the students at *Basic/Intermediate* know?
- What English language skills demonstrate that the students are at *Basic/Intermediate*?

One panelist from each table group was appointed as a recorder to write the comments from the table discussion on a flip chart. The table group discussion was then shared with the entire group.

After the panelists had a good understanding of the distinguishing characteristics between the adjacent proficiency levels based on the PLDs, the facilitator defined what the borderline students were and presented graphically who they were (Figure 1). First, the table groups were asked to identify three characteristics or behaviors that most distinguish the students who are just barely at *Proficient* from the students who are at *Basic/Intermediate*. Then, the table groups were also asked to identify three characteristics or behaviors that most distinguish the students who are just barely at *Basic/Intermediate* from the students who are at *Pre-emergent/Emergent*. Again, each group recorded the work on a flip chart. Once all table groups completed the task, they reconvened as a single large group for the committee level discussion. The facilitator captured the discussion on a flip chart, brought it to a data analyst to type it up, and shared the printed copy with the panelists as the final draft of borderline student descriptors. These borderline student descriptors are shown in Appendix D.2.

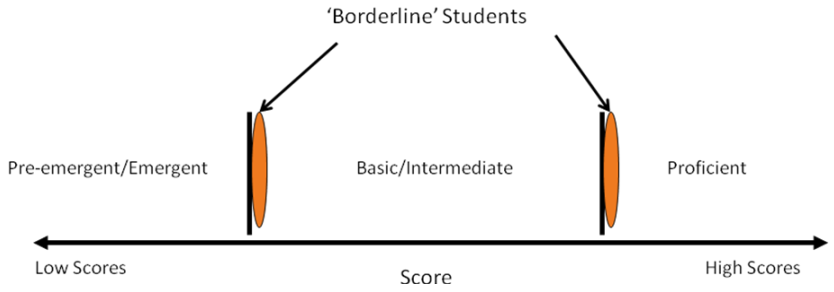



*Figure 1. Graphical representation of borderline students*

## *Standard Setting Methodology Training*

The facilitator gave the methodology training to the panelists (Appendix H for the training slides). Under the modified Angoff method (Angoff, 1971; Plake & Cizek, 2012) used for this meeting, the panelists were instructed to determine what percentage of students at the just barely proficient level should be able to answer correctly for one point items. For multiple point items, they were instructed to determine the average score of the students at the just barely proficient level. In order to come up with the average score of multiple point items, the panelists were instructed to determine what percent of students at the just barely proficient level should be able to score each score point and work out the computation on the worksheet in Figure 2. The panelists were instructed to rate percents for each item in increments of 5.

| Item 9 | Proficient | | | |
|---|---|---|---|---|
| Score | 0 | 1 | 2 | |
| 1. Percentage | | | | Sum |
| 2. Score × Percentage | | | | |

*Figure 2.* A sample worksheet for a multiple point item for the Proficiency cut

The panelists were told that a panelist's recommended cut would be determined by aggregating his or her ratings of all items. They were also told that a committee's recommended cut would be determined by taking the median of all panelists' recommended cuts. The facilitator instructed the panelists to consider the following four tools when they determined the rating of each item: 1) the Arizona ELP content standards, 2) the borderline student descriptors, 3) the item, and 4) the knowledge, skills, and abilities students developed. The panelists were instructed to rate the items for the *Proficient* cut and then continued with the *Basic/Intermediate* cut. The facilitator mentioned to the panelists that three rounds of ratings would take place, and some feedback information would be provided to them after each round.

## *Practice Round Ratings*

After going through the standard setting training, the panelists worked through a practice set of 10 items that were on the field test forms but not on the operational form for the *Proficient* level. The purpose of this exercise

was for the panelists to get comfortable with the rating process without feeling the pressure of reviewing real items.

### *Round 1 Ratings*

Prior to the Round 1 ratings, the facilitator made sure to address questions regarding the process that the panelists had. The table leaders also confirmed with the panelists in the table groups that they were willing and prepared to begin the Round 1 ratings. The panelists were also asked to fill out the Readiness Form (Appendix I), expressing they were ready to do the Round 1 ratings. After all panelists submitted the Readiness Form, the facilitator reminded them that they would begin the ratings with the *Proficient* cut and move on to the *Basic/Intermediate* cut and that this would be an individual task. As the panelists completed the Round 1 ratings on the rating sheet (Appendix J), the facilitator spot checked their rating sheets to make sure that they filled out the sheet correctly. After the panelists turned in the materials used for activities on Day 1 and signed the material sign in sheet, they were allowed to leave for the day. After all panelists left the meeting, the ADE, Pearson, and WestEd staff met to discuss the activities of the day and the results from the Round 1 ratings.

### *Round 2 Ratings*

At the beginning of Day 2, the facilitator started the meeting by sharing the empirical item difficulties and the results of the Round 1 ratings. The empirical item difficulty was an average score based on all PHLOTE students who participated in the field test administration. The rating distributions of each item for the *Proficient* cut and the *Basic/Intermediate* cuts were presented in bar graphs to the panelists. The median cut scores at the table group level as well as the committee level for the *Proficient* cut and the *Basic/Intermediate* cut were also shared with the panelists. The Round 1 rating sheet with the panelist's recommended cuts was also returned to each panelist.

Given the feedback the panelists received, the facilitator opened the discussion round with the following questions regarding the ratings of items:

- How similar are your ratings compared to the group (i.e., are there panelists who are more lenient or stringent that the other panelists)?
- Do panelists have different conceptualizations of the 'just-barely' students at the proficiency level?
- How similar are your ratings compared to the empirical item difficulty?

After the panelists reviewed the rating distributions of items and had discussions about each item, the facilitator led the panelists in a discussion of their recommended cut scores from Round 1. The panelists were asked:

- How similar are your cut scores compared to the group median cut scores?

The facilitator informed the panelists that consensus on their judgments was not a requirement.

Following the discussion, the facilitator reminded the panelists of the process for making their ratings for Round 2. They were told to begin the ratings with the *Proficient* cut and move on to the *Basic/Intermediate* cut. The facilitator instructed the panelists to reflect on the discussion about the Round 1 feedback as they determined whether they wanted to modify their Round 1 ratings. The facilitator checked with the panelists that they were ready to work on the Round 2 ratings and asked them to fill out the Round 2 readiness form. After all panelists had marked their readiness form, the Round 2 ratings took place.

### *Round 3 Ratings*

After the Round 2 ratings the panelists took a break while Pearson staff performed the analyses on the ratings. When the analyses were complete, the Round 2 rating sheet was returned to the panelists with their Round 2 recommended cuts for the *Proficient* cut and the *Basic/Intermediate* cut. The rating distributions for each item as well as the median cut scores at the committee level from Round 2 were shared with the panelists. Then, the panelists engaged in similar discussion regarding the feedback above that they had after Round 1.

Next, impact data was presented to the panelists. The facilitator informed the panelists that the impact data was the percent of students in the field test administration who would be classified into each proficiency level based on the Round 2 recommended cuts at the committee level. The impact data included the results for PHLOTE, non-PHLOTE, male, female, Hispanic, and non-Hispanic students. Please refer to *Standard Setting Results* for the impact data for Round 2. The panelists were asked if the impact data aligned with their expectations. The facilitator informed the panelists that the impact data was just a reality check and that they still should make judgments about items based on the knowledge, skills, and abilities required for the items.

As the panelists understood the feedback after Round 2 and had discussions about it, the facilitator led them to the Round 3 ratings. For Round 3, the panelists were asked to determine the recommended cuts for the Proficient cut and the Basic/Intermediate cut instead of rating each item as had occurred for Rounds 1 and 2. The facilitator made sure that the panelists understood the process for Round 3 and had them indicate so on the Round 3 readiness form. After all panelists indicated their readiness on the form, they made their Round 3 ratings.

After a short break, the final recommended cut scores based on the median cut scores from Round 3 and the corresponding impact data were presented to the panelists.

### *Standard Setting Evaluation Form*

The panelists filled out the standard setting evaluation form upon the completion of the meeting. The questions and responses to the evaluation form are summarized in Appendix I.

### *Performance Level Descriptors Refinement*

After the standard setting meeting, the panelists participated in a performance level descriptors refinement meeting. The facilitator distributed the instructions for refining the preliminary PLDs and went over them with the panelists. Please refer to Appendix D.2 for the instructions for modifying the performance level descriptors. The panelists were told that the preliminary PLDs were created by a committee of educators formed by ADE prior to the standard setting meeting. The panelists were informed that there were two parts on the PLD documents. The top of the PLD documents provided concise statements that would be placed on the student score report. The bullets at the bottom of the PLD documents listed the primary Performance Indicators (PIs) by domain from the ELP content standards, some of which might be combined to a single statement. The panelists were told that the refined PLDs produced by them would be given as their recommendation to ADE, to be reviewed and finalized by ADE.

After the facilitator gave the overview of the PLD document, the panelists were instructed to have discussion within the table group by following the instruction document. The panelists were asked to take notes on their recommendations as they would be shared with the other table groups for further discussion at committee level.

Following the table group discussion, the panelists started the large group discussion by sharing what they discussed at each table. The facilitator used the track changes facility to apply the recommended edits to the existing PLD document. The refined PLDs can be found in Appendix D.3. After the panelists went over the refinement of PLDs for each proficiency level and finalized their recommendations, they were dismissed from the meeting.

## Standard Setting Results

In this section, the results from each round are summarized. Note that the results after each round were internally reviewed by ADE, Pearson, and WestEd staff before they were shared with the panelists. Also note that some

results in this section were not shared with the panelists but with ADE, Pearson, and WestEd staff. Please refer to *Detailed Standard Setting Procedures* for the feedback the panelists received after each round. Please see detailed standard setting results such as the rating distributions of items by round in Appendix L.

The median raw score cuts for the Proficient and Basic/Intermediate levels by round are summarized in Table 4. The median raw score cuts for the Proficient and Basic/Intermediate levels at Round 1 were 22 and 32 (out of 42 maximum possible points), respectively. The recommended cuts did not change at Rounds 2 and 3 although the distributions of recommended cut raw scores were somewhat different across the three rounds (Figures 3-5).

Table 4.

*Median Raw Score Cuts by Round*

| Raw Score Cuts | Basic/Intermediate | Proficient |
|---|---|---|
| Round 1 | 22 | 32 |
| Round 2 | 22 | 32 |
| Round 3 | 22 | 32 |

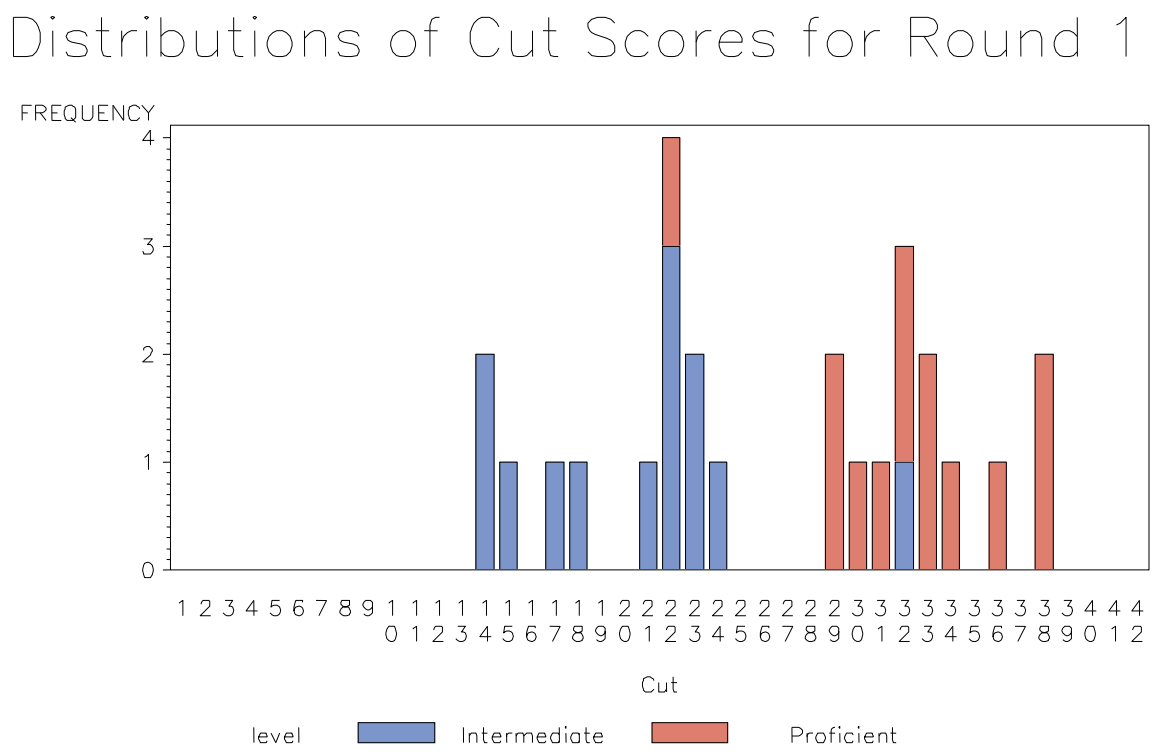



*Figure 3.* Raw score cut distributions of cuts at Round 1

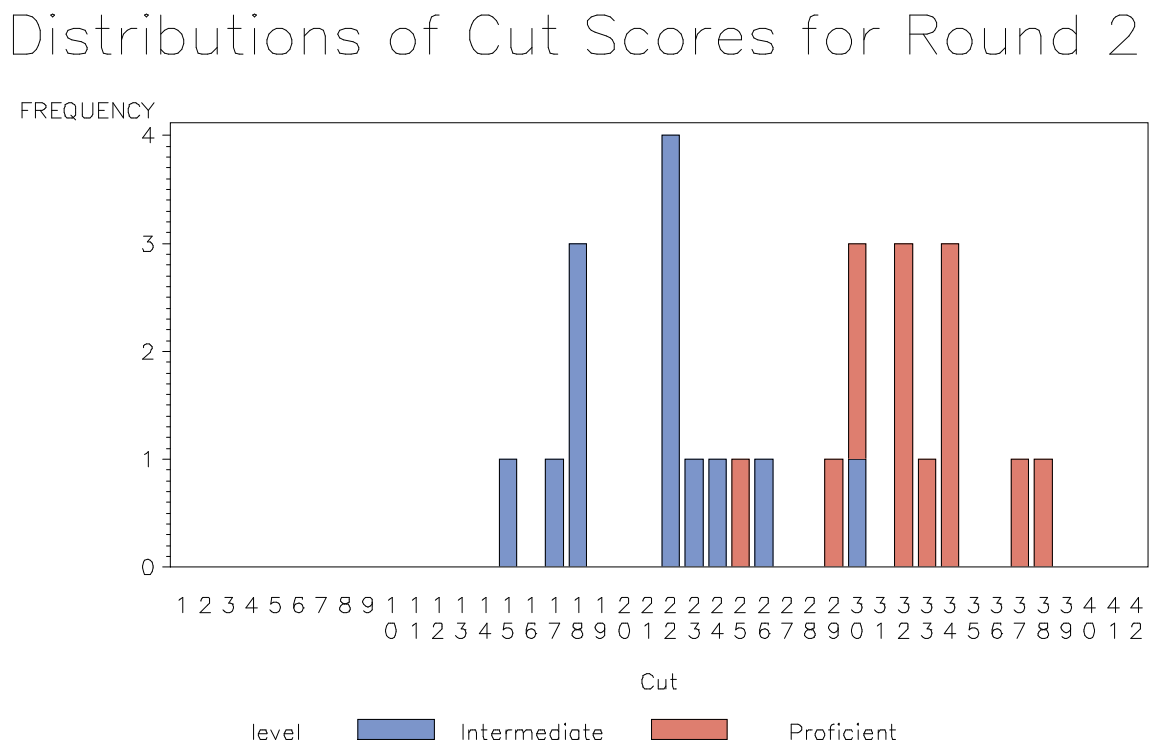



*Figure 4.* Raw score cut distributions of cuts at Round 2

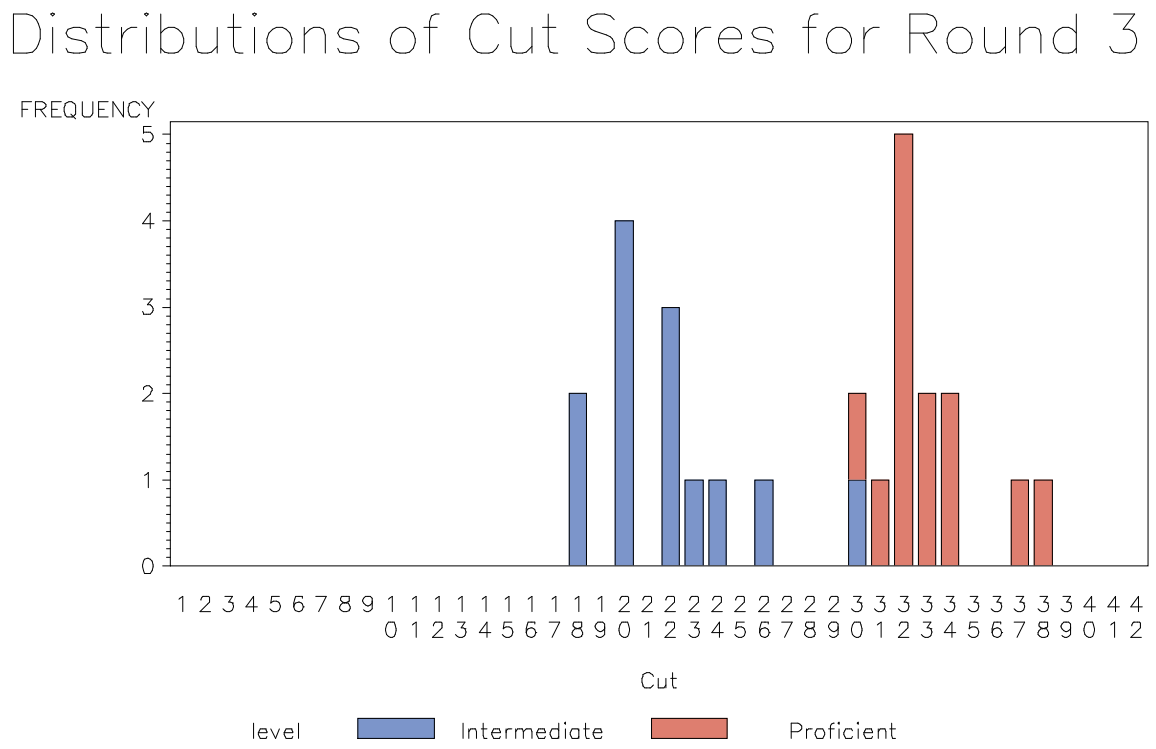



*Figure 5.* Raw score cut distributions of cuts at Round 3

Scale score cuts based on the median raw score cuts for the Proficient and Basic/Intermediate levels by round are presented in Table 5. Since the median raw score cuts did not change for Rounds 1 – 3, the scale score cuts stayed the same for all rounds. Note that the scale score of AZELLA Kindergarten Placement test ranges from 100 to 300.

Table 5.

*Scale Score Cuts by Round*

| Scale Score Cuts | Basic/Intermediate | Proficient |
|---|---|---|
| Round 1 | 208 | 235 |
| Round 2 | 208 | 235 |
| Round 3 | 208 | 235 |

The PHLOTE student proficiency level distribution by round is presented in Table 6. Approximately 15%, 26%, and 60% of PHLOTE students who participated in the field test administration would be classified as Pre-emergent/Emergent, Basic/Intermediate, and Proficient, respectively. Again, the impact data did not change from Round 1 through Round 3.

Table 6.

*Impact Data for PHLOTE Students Based on Median Cuts by Round*

| Impact | Pre-emergent/Emergent | Basic/Intermediate | Proficient |
|---|---|---|---|
| Round 1 | 14.7% | 25.5% | 59.8% |
| Round 2 | 14.7% | 25.5% | 59.8% |
| Round 3 | 14.7% | 25.5% | 59.8% |

Finally, the proficiency level distributions for the PHLOTE, and non-PHLOTE students, and subgroups of PHLOTE students by gender and Hispanic status at Round 3 are summarized in Figure 6. 88% of non-PHLOTE students who participated in the field test administration would be classified as Proficient. The performance of students within the subgroups was similar for gender and Hispanic status.

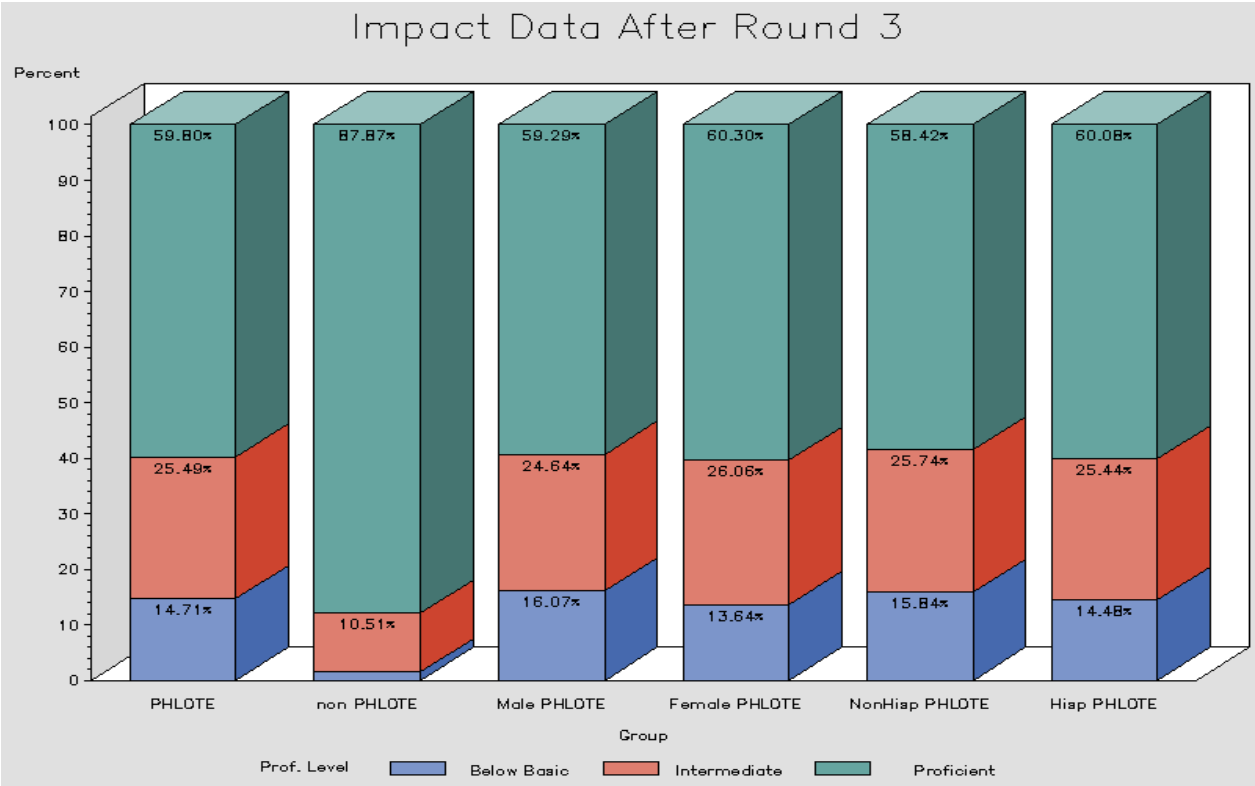



*Figure 6.* Proficiency level distributions for PHLOTE, non-PHLOTE, and subgroups of PHLOTE by gender, and Hispanic status at the final round

## References

Angoff, W.H. (1971). Scales, norms and equivalent scores. In R.L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp.508-600). Washington DC: American Council on Education.

Masters, G.N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.

Plake, B. S. & Cizek, G. J. (2012). Variations on a theme: The modified Angoff, extended Angoff, and Yes/No standard setting methods. In Cizek, G. J. (Eds.), *Setting performance standards: Foundations, Methods, and Innovations*. (pp.181-199). New York, NY: Routledge.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmark Pedogogiske Institute. (Reprinted 1980 by University of Chicago).

# Appendix A: Standard Setting Participants

| Occupation | Title of Current Position | Years in Current Position | Educational Experience | Certification | Endorsement | Highest Degree | Ethnicity | Gender | Taught SpEd? | Taught ELL/ESL? | District Size | District Location | School/ Institution SES |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Teacher | Kindergarten Teacher | 6 | Teaching K for 6 yrs. Been on interview team. Grade level chair for 2 yrs. District K core team | Elementary Ed. | Early Childhood, Reading, SEI | Masters | W | F | N | N | M | Urban | Upper Middle |
| Teacher | Kindergarten Teacher | 7 | Teaching K for 7 yrs | | SEI, Early Childhood | Masters | H | F | N | N | L | Suburban | Lower Middle |
| Teacher | Kindergarten Teacher | 36 | Taught K, 1, 2, and 6. Literacy coordinator. | Elementary K-8, Principal K-8, Administration K-8 | SEI | Masters | W | F | N | N | M | Suburban | Lower |
| Teacher | Kindergarten Teacher | 4 | Teaching K for 4 yrs. Taught Gr. 1 and SpEd PreK | Early Childhood Ed. | SEI | Masters | W | F | N | N | L | Suburban | Upper Middle |
| Education Non-teacher | ELL Coordinator/ Program Coach | 3 | Taught Gr. 4 for 4 yrs, Gr.5 for 7 yrs, K for 3 yrs, K-2 for 3 yrs, 1-2 BLE for 4 yrs, 1-3 BLE. ELL specialist for 5 yrs. ELL program coach/coordinator for 3 yrs. | Elementary K-8 | K-12 ESL, BLE, Reading, SEI | Masters | W | F | N | Y | L | Urban | Lower Middle |
| Teacher | ELL Specialist | 13 | Taught K-6 ELL | Elementary, Early Childhood Ed. | ELL, SEI, Reading | Bachelor | A | F | N | Y | S | Suburban | Middle |
| Teacher | Kindergarten Teacher | 4 | Taught K, 1, 2, 4 | Early Childhood Ed. | SEI | Masters | W | F | Y | Y | L | Suburban | Middle |
| Education Non-teacher | Speeh language pathologist/ Research technician | 37 | SLP in public schools preK-12 since 1975. Diagnostics primarily since 1990. Research technician for Uof A and WMU. Supervision of grad students, some quasi administrative experience. | AZ regular speech pathologist, Licensed SLP | | Masters | W | F | N | N | L | Urban | Lower Middle |
| Teacher | Kindergarten Teacher | 4 | Taught K for 4 yrs, ELD for 1 yr, PreK for 1 yr | Early Childhood Ed. | SEI | Bachelor | W | F | N | Y | M | Rural | Lower Middle |
| Education Non-teacher | ELD Coordinator | 1 | Taught K-3, SEI Gr. 1-2, SEI Gr. 2-3, K-2 ESL pull out/push in | | SEI, ESL | Bachelor | W | F | N | Y | S | Suburban | Middle |
| Teacher | Kindergarten Teacher/ Grade Level Chair | 11 | Teaching K for 11 yrs. Grade level chair. K ESL/SEI class | | Bilingual, Reading, Early childhood | Masters | H | F | N | Y | S | Rural | Lower |
| Teacher | Kindergarten Teacher | 5 | Taught K, 1, 3, 4. K SEI class for 1 yr. | K-8, Principal | SEI, Reading, Gifted, Early Childhood | Masters | W | F | N | Y | S | Rural | Lower Middle |
| Teacher | Kindergarten Teacher | 3 | Teaching K for 3 yrs. Instructional assistant with ELL students for 7 yrs. Worked with K-4 students. | | SEI, K-8 | Bachelor | W | F | N | N | S | Rural | Lower Middle |

# Appendix B: Table Leader Information Sheet

**ARIZONA ENGLISH LANGUAGE LEARNER ASSESSMENT (AZELLA)**
**TABLE LEADER INFORMATION SHEET**
**KINDERGARTEN PLACEMENT TEST**

**BLACK CANYON CONFERENCE CENTER**
**PHOENIX, ARIZONA**

**Role Description**

- Facilitate discussion
- Keep process on track
- Vote as one of the table members
- Monitor group discussion
- Watch the clock and monitor time
- Cut off discussion or diplomatically resolve differences between members

**Specific Tasks:**

1. Before all rounds
    a. Make sure participants put ID numbers on forms
    b. Check that participants complete readiness forms
    c. If someone puts a NO on readiness form, see if you can help explain. If participant is still unsure, inform Pearson facilitators
    d. Ensure that table members understand activity
    e. Notify group leaders of any problems
2. After Round 1
    a. Check that participants rated all items for each proficiency cut (Proficient, Basic/Intermediate) on the rating sheet
    b. Collect all table members' rating sheets and give to Pearson facilitators
3. Round 1 Feedback and Discussion
    a. Ensure that all members participate in discussion and encourage all points of view
    b. Check that participants understand agreement data
    c. Lead a discussion on items with greatest judgment disagreement
4. Rounds 2 and 3 Feedback and Discussion
    a. Ensure that all members participate in discussion and encourage all points of view.
    b. Check that participants understand agreement data AND impact data
    c. Lead a discussion on items with greatest judgment disagreement
5. Before breaks and at end of day
    a. Remind participants to leave all secure materials on the table
    b. Remind participants to initial the Materials Sign-in Sheet at end of each day.
    c. Collect all materials and verify that all have been received
6. After collection at the end of the day
    a. Turn in all materials to Pearson facilitators
    b. Participate in a debriefing session with ADE at end of Day 1

## Appendix C: Table Leader Training Slides

# Standard Setting on the Arizona English Language Learner Assessment (AZELLA)

### Kindergarten Placement Test

Table Leader Training

July 11, 2012

Phoenix, Arizona

ALWAYS LEARNING PEARSON

### Purpose of Standard Setting

- The purpose of this standard setting is to establish 'recommended' cut scores on the Kindergarten Placement Test.
- You were selected to serve on the committee for a variety of reasons:
  - Familiarity with the knowledge and skills required to "master" the English Language Proficiency standards at various Proficiency levels
  - Representation of various jurisdictions and demographic characteristics
- You were selected to be a table leader because of your experience, ability to lead, and strong communication skills.

2

ALWAYS LEARNING PEARSON

## Standard Setting Roles

- Lead Research Scientist
- Statistical Analyst
- Content Specialist
- Program Management
- ADE Staff
- Table Leader
- Participants

3

ALWAYS LEARNING PEARSON

## Table Leader Roles

- Facilitate discussion
- Keep process on track
- Vote as one of the table members
- Monitor group discussion
- Watch the clock and monitor time
- Might need to cut off discussion or diplomatically resolve differences between members

4

ALWAYS LEARNING PEARSON

## Table Leader Tasks

- Provide instructions
  - ID numbers
  - How to fill out rating form
  - How to collect and return materials
- Lead discussion at table and across tables
  - Ensure that all participants engage in discussion
- Verify understanding
  - Process
  - Feedback
- Verify completeness
  - Readiness forms
  - Rating forms
- Materials collection and audit
- Notify facilitator of problems
- Participate in a debriefing session with ADE at end of Day 1

5

ALWAYS LEARNING PEARSON

## Standard Setting Overview

Panelists will:

- Review and discuss the test.
- Develop a shared understanding of each Performance Level (PLDs).
- Develop "Borderline Student" Descriptors.
- Receive Standard Setting Training and Practice.
- Participate in three rounds of ratings –
  - Round 1: Independent
  - Round 2: Independent, but with table discussion
  - Round 3: Independent, but with table & full group discussion
- Finalize Performance Level Descriptors.

6

ALWAYS LEARNING PEARSON

## Table Leader Role in Gaining an Understanding of the Proficiency Levels

- Within each table group, ask, "What should students know and be able to do at each level?"
  - "Basic/Intermediate", "Proficient"
- Appoint a recorder to write on the flip chart.
- Suggestions should be:
  - Concrete.
  - Clearly related to the PLDs.
- Note: This concept will be presented by the facilitator, but the table leader will facilitate the conversation at his/her table.

7

ALWAYS LEARNING PEARSON

## Table Leader Role in Gaining an Understanding of the Proficiency Levels

- Ask Table Members to describe concretely students who are at "Proficient."
  - What should they be able to *do*?
  - What *skills* should they possess?
  - What should they *know*?
  - What *academic behaviors* demonstrate that they are at "Proficient"?
- Repeat the process for "Basic/Intermediate."

8

ALWAYS LEARNING PEARSON

## Table Leader Role in Borderline Student Descriptors

- Ask table members to think about the borderline students at "Proficient."
  - Identify three characteristics or behaviors that MOST distinguish a student who just barely "Proficient."
  - Record the three responses on your flipchart.
- Repeat the process for "Basic/Intermediate."

9

ALWAYS LEARNING PEARSON

## Three Rounds of Ratings

- Round 1 Ratings
  - Independently
- Round 2 Ratings
  - Independently, but after discussion with your table group
- Round 3 Ratings
  - Independently, but after discussion with your table group and entire committee

10

ALWAYS LEARNING PEARSON

## Standard Setting Item Map and Rating Sheet

- Each panelist will be provided with an item list that provides information about each item.
- Each panelist will record his/her recommended average item score on a rating sheet.
- The table leader will help panelists with questions about how to use these documents.

11

ALWAYS LEARNING PEARSON

## Table Leader Role Before All Rounds

- Make sure participants put ID numbers on the forms.
- Check that participants complete the readiness forms.
- If someone puts a "NO" on the readiness form, see if you can help explain. If the participant is still unsure, inform the Pearson facilitator.
- Ensure that table members understand activity.
- Notify the facilitator of any problems.

12

ALWAYS LEARNING PEARSON

## Table Leader Roles After Round 1

- Check that participants recorded ratings correctly on Rating Sheets.
- Collect all table members' rating sheets and give to the facilitator.

13

ALWAYS LEARNING PEARSON

## Table Leader Roles at Round 1 Feedback and Discussion

- Ensure that all members participate in the discussion and encourage all points of view.
- Check that participants understand agreement data.
- Lead discussion on items with greatest judgment variability.

14

ALWAYS LEARNING PEARSON

## Table Leader Roles at Rounds 2 and 3 Feedback and Discussion

- Ensure that all members participate in the discussion and encourage all points of view.
- Check that participants understand the agreement data AND impact data.
- Lead a discussion on items with greatest judgment disagreement at table level as well as committee level.

15

ALWAYS LEARNING PEARSON

## Table Leader Roles Before Breaks and at End of Day

- Remind participants to leave all secure materials on the table.
- Remind participants to initial the Materials Sign-in Sheet at end of each day.
- Collect all materials and verify that all have been received.

16

ALWAYS LEARNING PEARSON

## Recap of Table Leader Tasks

- Provide instructions
- Lead discussion at table and across tables
- Verify understanding
- Verify completeness of forms
- Materials collection and audit
- Notify facilitator of problems
- Participate in a debriefing session with ADE at end of Day 1

Click here for: Table Leader Handout

17

ALWAYS LEARNING PEARSON

# Appendix D: Performance Level Descriptors

## Appendix D.1: Preliminary Performance Level Descriptors

## Arizona English Language Learner Assessment Performance Level Descriptors
## Kindergarten Placement Test

**Proficient**—Students at this level listen and respond appropriately to spoken English. They have an expanded English vocabulary to orally communicate basic needs and ideas with English words, phrases, and sentences. They have a limited understanding of sounds in words. They can use pictures to retell events from a story heard, identify pictures with the same first sound, and can add details to drawings. This student demonstrates the skills necessary to access mainstream curriculum.

**Basic/Intermediate**—Students at this level have a general ability to understand spoken English, but do not have the vocabulary to respond consistently. They orally communicate basic needs and ideas with gestures and isolated English words. They can use pictures to recall objects from a story heard, and show minimal control when adding details to drawings. This student does not demonstrate the sufficient skills in English to access mainstream curriculum and demonstrates the need for specific support in English Language Development instruction.

**Pre-Emergent/Emergent**—Students at this level has no ability or a very limited ability to communicate, retell stories heard, or add details to drawings in English. This student does not demonstrate the sufficient skills in English to access mainstream curriculum and demonstrates the need for specific support in English Language Development instruction.

| **Pre-Kindergarten**<br>Students scoring **Proficient** at this grade generally demonstrate the following skills, knowledge, and abilities drawn from the Intermediate level of the ELP standards. | **Pre-Kindergarten**<br>Students scoring **Basic/Intermediate** at this grade generally demonstrate the following skills, knowledge, and abilities drawn from Pre-Emergent and Emergent levels of the ELP standards. |
|---|---|
| • Speaking/Listening<br>○ Responds appropriately to social interactions<br>○ Consistently uses correct pronunciation<br>○ Uses correct developmentally–appropriate grammatical structures<br>○ Uses basic verbs and adjectives<br>○ Follows multiple-step directions<br>○ Sequences events in a story<br><br>• Pre-reading<br>○ Uses pictures to retell events from a story heard<br>○ Identifies pictures that begin with the same first sound<br><br>• Pre-Writing<br>○ Adds details to drawings | • Speaking/Listening<br>○ Minimally responds to social interactions<br>○ Repeats individual words and short phrases<br>○ Uses basic nouns<br>○ Follows single-step directions<br>○ Identifies objects from a story<br><br>• Pre-reading<br>○ Uses pictures to recall people or objects from a story heard<br><br>• Pre-writing<br>○ Demonstrates minimal control when adding details to drawings |

## Appendix D.2: Borderline Student Descriptors

| Borderline Proficient Student Descriptors |
| --- |
| 1. Ask for clarification<br>2. Confident enough to take risk - Becoming more verbal, Using complete sentence, Increasing participation, Attempts to answer, Takes time to think<br>3. Using prior knowledge to make connections between world and self, Using verbal and non-verbal social cues to communicate, Carrying on conversing with peers and adults<br>4. Using details when speaking and retelling and writing<br>5. Speaking simple and grammatically correct sentences, incomplete mastery of grammatical structure<br>6. Has basic vocabulary, but doesn't experiment with new vocabulary<br>7. Limited fluency in speaking narratively<br>8. Reading - knows most letters and some sounds, not consistent in identifying initial sounds, Can retell one or two ideas from story, Refers to characters by a common noun,<br>9. Can communicate with short phrases and sentences using appropriate pronunciation and grammatical structure<br>10. Can retell a story including a minimal one detail or event<br>11. Can perform a minimum of two step directions |

| Borderline Basic/Intermediate Student Descriptors |
| --- |
| 1. Uses gestures and one word or short phrases to respond or express ideas<br>2. Uses pictures to express wants/needs<br>3. Minimal details in speaking and retelling of stories and drawing pictures<br>4. Follows one step direction<br>5. Uses one word response and questions<br>6. Follows one step direction with visual support<br>7. Non-verbal response to familiar objects<br>8. Repeat phrases and simple sentences<br>9. Draw familiar objects<br>10. Recalls familiar items, generalizes characters and objects from a story<br>11. Can respond one word answer in English<br>12. Can repeat a minimum number of individual words<br>13. Can identify basic nouns |

# Appendix D.3: Instruction for the Performance Level Descriptor Refinement

## Refinement of AZELLA Kindergarten Placement Test PLDs

Revised July 6, 2012

**Instructions to tables for refinement of the AZELLA Kindergarten Placement Test PLDs:**

### Background

- The top part of the PLDs presents the three performance levels and is a generalized reflection of the bullets on the bottom. This narrative piece is used for student reports.

- The bullets at the bottom are designated as highlighted PIs from the English Language Proficiency Standard, and several PIs may have been combined into single bullets. The bullet text and PI verbiage are usually not verbatim.

### Procedures

1. The ELP standards must be available as a reference for this activity.
2. Begin with the bullets at the bottom. Determine if any bullets should move from one performance level to another. If a bullet (PI) is moved and the objective is noted in the narrative at the top of the PLD, the objective must also be moved to the new performance level in the narrative.
3. Since some bullets are a combination of PIs, it may be necessary to break apart the bullet to place the separate parts in different performance levels. If needed, make the appropriate adjustments to the narrative.
4. Note the bullet's beginning action verb. The verb, along with the rest of the text, may be changed and kept at the original performance level or moved to another.
5. New bullets may be added if appropriate and necessary; however, removal of bullets is not recommended. All assessments must conform to the test blueprint, and although not all the bullets will be covered in the current assessment, over time, the future assessments will include all the performance indicators identified in the bullets.
6. Adjust the narrative accordingly.
7. Table Leaders will share their tables' recommendations, and Track Changes will be made to the existing document.

# Appendix D.4: Final Performance Level Descriptors after Refinement Process

## Arizona English Language Learner Assessment Performance Level Descriptors
## Kindergarten Placement Test

**Proficient**—Students at this level listen and respond appropriately to spoken English. They have an expanded English vocabulary to orally communicate basic needs and ideas with English words, phrases, and sentences with correct pronunciation. They use pictures or words to retell events from a story heard, identify pictures with the same first sound, and add relevant details to drawings. This student demonstrates the skills necessary to access mainstream curriculum.

**Basic/Intermediate—**Students at this level generally understand spoken English, but do not have the vocabulary to respond consistently. They orally communicate basic needs and ideas with gestures and isolated English words. They use pictures to recall objects from a story heard, repeat words that begin with the same first sound and add minimal details to drawings. This student does not demonstrate the sufficient skills in English to access mainstream curriculum and demonstrates the need for specific support in English Language Development instruction.

**Pre-Emergent/Emergent**—Students at this level lack the English skills to communicate, retell stories heard, or add details to drawings. This student does not demonstrate sufficient skills in English to access mainstream curriculum and demonstrates the need for specific support in English Language Development instruction.

| **Pre-Kindergarten**<br>Students scoring **Proficient** at this grade generally demonstrate the following skills, knowledge, and abilities drawn from the Kindergarten ELP standards. | **Pre-Kindergarten**<br>Students scoring **Basic/Intermediate** at this grade generally demonstrate the following skills, knowledge, and abilities drawn from the Kindergarten ELP standards. |
|---|---|
| • Speaking/Listening<br>○ Responds appropriately to social interactions<br>○ Consistently uses correct pronunciation<br>○ Uses correct developmentally–appropriate grammatical structures<br>○ Uses basic verbs and adjectives<br>○ Follows multiple-step directions<br>○ Sequences events in a story<br><br>• Pre-Reading<br>○ Uses pictures or words to retell events from a story heard<br>○ Identifies pictures that begin with the same first sound<br>○ Uses pictures to make predictions<br><br>• Pre-Writing<br>○ Adds relevant details to drawings | • Speaking/Listening<br>○ Responds to social interactions with gestures and simple words<br>○ Repeats individual words and short phrases<br>○ Uses basic nouns<br>○ Follows single-step directions<br>○ Identifies objects from a story<br><br>• Pre-Reading<br>○ Uses pictures to recall people or objects from a story heard<br>○ Repeats words that begin with the same first sound<br><br>• Pre-Writing<br>○ Adds minimal details to drawings |

# Appendix E: Standard Setting Script

**ARIZONA ENGLISH LANGUAGE LEANER ASSESSMENT (AZELLA)**
**STANDARD SETTING SCRIPT FOR FACILITATORS**
**KINDERGARTEN PLACEMENT TEST**

**JULY 11-12, 2012**
**BLACK CANYON CONFERENCE CENTER**
**PHOENIX, ARIZONA**

## OVERVIEW OF STANDARD SETTING TASKS

The standard setting for AZELLA KPT will take place from July 11-12 at the Black Canyon Conference Center. The Standard Setting will consist of the following activities. Each of these will be described in detail in this standard setting script which is intended for the standard setting facilitators.

**Day 1**

A. Table Leader Training
B. Opening Remarks
C. Overview of the Test and Standard Setting
D. Review of Performance Level Descriptors and Scoring Rubric
E. Review and discuss the test
F. Borderline Student Descriptors Development
G. Standard Setting Training
H. Practice Round of Ratings
I. Round 1 Ratings

**Day 2**

J. Round 1 Feedback and Discussion
K. Round 2 Ratings
L. Round 2 Feedback and Discussion
M. Round 3 Ratings
N. Round 3 Feedback
O. Standard Setting Evaluation
P. Performance Level Descriptor Refinement

## JULY 10, 2012

**5:00pm – 7:00pm Ensure arrival of materials/supplies (Hotel)**

- Take Inventory of Supply Box
- Take inventory of Facilitator Binder
- Take inventory of other materials
- Take inventory of participant folders (See Table 1)

## JULY 11, 2012

**6:30am – 7:30am Room Set-Up (Steve, Hiro, Erica)**

- Verify Set-Up of Room against room diagram in Facilitator Binder.
- Place seating cards at chair locations (Table leaders are in positions 1, 7, and 13)
  - Table 1: Participants 1-6
  - Table 2: Participants 7-12
  - Table 3: Participants 13-18
- Set up projector
- Set out Pencils at every place
- Set out post-its and highlighters in the middle of tables
- Remove all pads of paper
- Have room locked
- Go to breakfast area

**7:30am – 8:00am Table Leader Training (Steve)**
**Materials:**

- **Participant Folder**
- **Table Leader PowerPoint Presentation**
- **Table Leader Handout**

Three table leaders will be assigned by ADE. Table leaders are experienced educators and may have had a previous role with the assessment. The primary role of the table leader is to monitor the group interaction, keep the group focused on the task at hand and keep time for the group. The table leaders will be given a copy of the 1) Agenda, 2) Table Leader PowerPoint Presentation, and 3) Table Leader Handout. We will discuss their role and responsibilities during the standard setting meeting.

Pearson will explain to table leaders what their role will be in general and relative to each standard setting task. We will make sure they understand that they will be leading the discussions within their group. Therefore, they need to have a clear understanding of the process. Below is a bulleted list of information that we plan to share during the table leader training.

<u>Role Description</u>

- Facilitate discussion.
- Keep process on track.
- Vote as one of the table members.
- Monitor group discussion.
- Watch the clock and monitor time.
- Cut off discussion or diplomatically resolve differences between members when necessary.

<u>Specific Tasks</u>:

1. Before all rounds

   a. Make sure participants put ID numbers on forms
   b. Check that participants complete readiness forms
   c. If someone puts a NO on readiness form, see if you can help explain. If participant is still unsure, inform Pearson facilitators
   d. Ensure that table members understand each activity
   e. Notify group leaders of any problems
2. After Round 1
   a. Check that participants rated all items for each proficiency cut (Proficient, Intermediated/Basic) on the rating sheet
   b. Collect all table members' rating sheets and give to Pearson facilitators
3. After Round 1 table level agreement data are shared
   a. Ensure that all members participate in discussion and encourage all points of view
   b. Check that participants understand agreement data
   c. Check that participants mark items with the greatest range of ratings after table data are shared
   d. Lead discussion on what those items are measuring and whether a student who meets the minimum requirements should be able to answer them
4. After Rounds 2 and 3
   a. Ensure that all members participate in discussion and encourage all points of view.
   b. Check that participants understand agreement data AND impact data
   c. Check that participants mark items with the greatest range of ratings after table data and group data are shared
   d. Lead a discussion on what those items are measuring and whether a target student who meets the minimum requirements should be able to answer them
5. Before breaks and at end of day
   a. Remind participants to leave all secure materials on the table
   b. Remind participants to initial checkout materials sheet
   c. Collect all materials and verify that all have been received
6. After collection at the end of the day
   a. Turn in all materials to Pearson facilitators
   b. Participate in debriefing session with ADE (except last day)

### 7:30-8:00 Breakfast

### 8:00-8:30 Registration (Erica)

Participants should check in with Pearson staff. Upon arrival, each participant will be given a folder. Participants should write down their names on the cover of the folder. See Table 1 for items included in Participant Folder.

Table 1: Materials Included in Participant Folder

| Item | Location |
|---|---|
| Name Tent | Left pocket |
| Agenda | Left pocket |
| PowerPoint Training Presentation: General SS Session | Left pocket |
| Demographic Information Survey | Right pocket |
| Non-disclosure Form for ADE | Right pocket |

| | |
|---|---|
| Non-disclosure Form for Pearson | Right pocket |
| Reimbursement Form | Right pocket |

As indicated on the sign-in sheets that are included in the facilitator binder, the master copies for ADE and Pearson staff are labeled as A-I as defined below in Table 2. The panelists receive numbered copies of materials from 1-18. Table 1 will receive materials 1-6; Table 2 receives materials 7-12; Table 3 receives materials 13-18.

Table 2: Master Copies of Secure Materials

| | |
|---|---|
| A | Pearson: Steve Fitzpatrick |
| B | Pearson: Hiro Fukuhara |
| C | ADE: Roberta Alley |
| D | ADE: Charlie Bruen |
| E | ADE: Frank Brashear |
| F | ADE: Lee Scott |
| G - I | ADE |

**8:30-9:00 Opening Remarks – ADE and Pearson RS (Roberta and Steve)**

- **Welcome and Why You Are Here**
- **Review of Agenda**
- **Security Forms/Non-disclosure forms (Erica)**
- **Reimbursement forms (Erica)**

ADE formally welcomes participants and explains the purpose of the standard setting meeting.

Pearson RS introduces the Pearson staff involved and their role in the standard setting meeting. Go over the agenda and the security forms and administrative tasks. Emphasize that the secure materials are based on operational items and security is of paramount importance throughout the standard setting process.

**9:00-9:15 Overview of the Tests (ADE: Roberta)**

- **History**
- **Purposes**

ADE staff gives a brief overview of the AZELLA test; provide historical background of the test, purposes of the test, and implementation of the AZELLA test. Introduce key concepts of the test, the test blueprint, scoring rubric etc.

**9:15-9:30 Overview of Standard Setting (Steve)**

- **Purpose**
- **Modified Angoff Method**

Pearson RS goes over the powerpoint slides for the overview of standard setting. Present the purposes of standard setting, definitions of content standards and performance standards, and definition of the borderline students. Pearson RS also briefly introduces the modified Angoff method.

*Pearson staff should collect the signed confidentiality agreement form before participants break*.

**9:30-9:45 BREAK**

**9:45-10:00 Committee Introductions**

- Make sure everyone is in the correct room.
- Introduce yourself and give some background.

Once everyone is settled in the room, the participants are asked to introduce themselves and provide some information about their professional experience. Participants may share the following:

- Name.
- Where are you from?
- How long you have been in your current position/field?
- What educational roles you have fulfilled?
- Have you participated in a standard setting before?
- Tell us something interesting about yourself.

Remind the participants to write their names on their folders if they have not done so already. A review of the agenda for the rest of the day is provided in order for participants to develop a perspective of what is to be accomplished and the pace at which the meetings should proceed. Note that we might deviate from the time allotments on the agenda if we feel a topic requires additional discussion.

**10:00-10:15 Review of Performance Level Descriptors**

Distribute the PLDs and the scoring rubric. Allow time for the panelists to review them but do not begin the in-depth discussion that leads to the development of the borderline student descriptors. Tell them that they are going to review the test. Then they will discuss the PLDs, scoring rubric.

**10:15-11:00 Review the KPT**

- **Go over the test**
- **Review scoring rubrics**

Important Notes

- Hand out Test Booklets
- Hand out the answer document
- Verify that each panelist gets the correct security number
- This is an individual, independent activity (no discussion)
- When each panelist completes the test give them the scoring key

In order for participants to gain an appreciation of the assessment experience and the instrument's degree of difficulty, participants are asked to look through the operational test that will be used in the standard setting. Participants will spend approximately 45 minutes walking through the test booklet to understand how the test is administered and reviewing the scoring rubrics on the answer document. Since the KPT is individually administered, it is difficult to mimic a live administration. Thus, instead of taking the KPT, participants will walk through the KPT items independently.

### 11:00-11:45 Discuss the Test

Spend some time discussing the overall test experience. Ask questions such as:

1. What are your general impressions about the test?
2. Did the test generally cover the depth and breadth of the content standards?
3. Does the test generally have a range of item difficulties (e.g., easier items, moderate items, difficult items)?

Although some discussion about individual test items is normal, focus participants away from prolonged debate about the quality or appropriateness of the items. Ask participants to record any comments about the test items on the index cards provided and they will be passed on to ADE.

### 11:45-12:30 LUNCH

- Inform location of lunch.
- Remind them when to return.
- Have them place all material in a pile at their seating location.
- Do not leave your room until it is locked.
- Get door unlocked at 12:15 and then remain in room.

### 12:30-2:30 Borderline Student Descriptors Development

- **Discuss knowledge, skills, and abilities (KSAs) for each performance level**
- **Review and discuss characteristics of student response exemplars at each performance level**
- **Define the distinguishing characteristics of borderline student performance**

Careful notes need to be taken during performance level descriptors discussions. The participants will be split into three groups/tables, with six people per table. **One member per table should be appointed as note taker**.

Next, panelists will be familiarized with the performance level descriptors. To familiarize panelists with the performance level descriptors and to help foster a shared understanding of them, Pearson facilitators will distribute a document listing the three performance level descriptors and then use this document to work with panelists to help summarize these descriptors. The goal will be to help all panelists develop and share a strong, common understanding of each performance level with specific emphasis on the way those performance level descriptions relate to the relevant content and grade level of the appropriate AZELLA KPT test.

Panelists will be asked to identify the main topics and skill sets addressed by the PLDs and to identify the three to four key characteristics that distinguish **performance at a given level** from that of adjacent performance levels for each topic or skill set. Panelists will conduct these tasks first in small group discussions at their table and then in a single large group.

After panelists have a good understanding of the distinguishing characteristics between the levels of performance based on the PLDs, they will work on identifying three characteristics that most distinguish students that are at the **borderline of each performance level**. They will start with the borderline between "Proficient" vs. "Basic/Intermediate." Within each table group, panelists will be asked to identify three characteristics or behaviors that most distinguish students that are at the

borderline of "Proficient" from the top of "Basic/Intermediate." Each table group will record their responses on a flip chart. They will repeat the same activity to distinguish three characteristics that differentiate between "Basic/Intermediate" vs. "Pre-Emergent/Emergent". Once the table groups have completed this task, they will reconvene as a single large group. Each table will present their distinguishing characteristics and the facilitator will lead a discussion of the commonalities and differences across the table groups. The facilitator will capture the discussion on the group flip chart.

Have the panelists place all of their materials in a pile at their seating location before taking a break.

### 2:30-2:45 BREAK

- Facilitators type up borderline student descriptors and have the Research Assistant (RA) print them

### 2:45-3:15 Standard Setting Training (Steve)

- **Modified Angoff Method**
- **Item Notes sheet**
- **Ratings Forms**

Pearson Research Scientist (RS) introduces the modified Angoff method. Provide a review of the modified Angoff method for the participants. Instruct participants to consider four tools when determining % of students who answer correctly (One point items) or determining the average score (Multiple point items); 1) the Arizona ELP content standards, 2) the borderline student descriptors, 3) the item, and 4) the KSAs they developed. A formal PowerPoint presentation will be provided.

### 3:15-3:30 Practice Round

A practice test booklet will be distributed by the facilitator. This allows participants to practice the procedure without feeling the pressure of reviewing real items. Using these items, the panelist will rate 10 items on the practice test booklet for a "Proficient" cut. Participants will work on the practice round individually.

Important Notes

- Hand out **Rating Sheet: Practice Round**
- Hand out **Test Booklet: Practice Round**

### 3:30-4:30 Round 1 Standard Setting

- **Readiness Check**
- **Round 1 Ratings**
- **Materials Collection**

Important Notes

- Hand out Readiness Form
- Hand out Rating Sheet
- Hand out Test Booklet

- Verify security numbers match (sign-out)

The facilitator will provide a short summary of the procedure just discussed. When no more questions are asked, and after all participants complete the two readiness questions, Round 1 will begin. Once participants demonstrate that they understand how to determine the % of borderline student who would answer correctly for 1 point items or average score of borderline students for multiple point items to make their Round 1 cuts. Remind participants that this is an individual activity. Check with the table leaders that everyone is ready for Round 1. Each participant should work on the "Proficient" cut first followed by "Basic/Intermediate". When finished, the table leader will collect and verify that all materials are received. Participants will be reminded that the meeting will resume the next morning at 8:00 (with breakfast starting at 7:30).

**Important Notes**

**When a panelist completes Round 1**

- Collect Rating Sheets (group them by table).
- Spot check Rating Sheet.
- Sign in Rating Sheet.
- Place Rating Sheet in designated folder and give to RA once all sheets are collected.

**Collect (and Sign-In) All Other Secure Materials (Use Secure Material Sign-In Sheet)**

- Test Booklet
- Answer Document
- Any notes

**Closing the room**

- Prepare room for the next day
- Get security to lock the room.

### 5:00-5:15 Table Leaders Debrief

The table leaders will meet with Pearson and/or ADE staff to discuss the activities of the day.

**END OF DAY 1**

## DAY 2 – JULY 12, 2012

**6:45 – 7:15 Room Set-Up**

- Place **all** materials and supplies on the table where the panelists were sitting from the day before.
- Set up projector and computer.
- Load the excel workbooks onto your computer.

**7:30-8:00 Breakfast**

**8:00-9:00 Round 1 feedback and discussion**

- **Discussion of rating frequency**
- **Discussion of recommended cuts**
- **Handouts**
  1. **Rating Sheets**
  2. **Item Means (Item Performance Data)**
  3. **Median cut scores**

**Important Notes**

- Go over agenda for the day.
- Hand out Rating Sheets.
- Hand out Item List with Item Mean.
- Hand out Median cut scores.
- Go through item rating charts
- Discussions will occur at committee level

Start on Round 1 feedback discussion. The Rating Sheets from Round 1 with the panelist's raw score cuts will be returned to the panelists. Median raw score cuts at committee level will be provided as a handout. Item map with item means as empirical item difficulty will also be provided as a handout to the participants. Explain that item mean is based on all the non-native English speaking students who took the field test in spring 2012, not just the borderline students at "Proficient", "Basic/Intermediate", and "Pre-Emergent/Emergent". Explain that the participants should use the item mean to check their estimates of how difficult an item is. Rating frequency of each item for entire committee will also be shown on the screen. In reviewing the rating frequency, participants will be asked to think about the following:

- How similar are their ratings to that of the group (i.e., is a given participant more lenient or stringent than the other participants)?
- If so, why is this the case?
- Do participants have different conceptualization of the borderline students?

After reviewing the rating frequencies, the discussion will be shifted to the cut scores. The participants will be asked to discuss the following:

- How similar are their cut scores compared to the group?

Lead the participants to discuss their "Proficient" ratings first, then move to the "Basic/ Intermediate," ratings.

During the Round 1 discussion, inform participants that we do not intend for them to come to consensus on their judgments, but we do want them to discuss differences to get a feel for why differences exist. Let them know that we want them to try to better understand the reasons for the differences. Are there underlying differences in

what the participants believe these borderline students can /can not do? Do they implement different procedures to assign ratings?

**9:00-9:30 Round 2 Standard Setting**

- **Readiness Check**
- **Round 2 Ratings**

Participants will be reminded that data are intended to inform, but not dictate their item ratings. When participants indicate that they understand the data they have been provided, have them fill out the readiness survey. When everyone answers "yes" to the Round 2 questions on the readiness survey, participants can start working on their round 2 ratings.

**Important Notes**

**When a panelist completes Round 2**

- Collect Rating Sheets (group them by table).
- Spot check Rating Sheet.

**When all have completed Round 2 ratings**

- Place in designated folder and give to RA.

**9:30-10:00 BREAK**

Over break, Pearson staff members enter data for Round 2 and generate feedback reports.

**10:00-10:45 Round 2 Feedback and Discussion**

- **Group discussion of recommended cut scores**
- **Group discussion of impact data**
- **Handouts**
    1. **Rating Sheets**
    2. **Median cut scores**
    3. **Impact Data**

At Round 2 feedback and discussion, the rating sheet will be returned to the participants with his/her Round 2 recommended cut scores. Participants will be provided the median cut scores for the committee as a handout. The facilitator leads the discussion with all tables combined. Remind the participants that consensus is not required. The participants will be asked to discuss about the following:

- How similar are their cut scores compared to the group?

Finally, participants will be provided a graphical display of the impact data using the median cuts for all students.

The impact data graphic representation provides participants with information on what percentages of students are at each performance level for the populations of interest (all students, female/male, and Hispanic/non-Hispanic, and Non-ELL).

Participants will be given time to discuss, within the Group, the appropriateness of the group level ratings given the proportion of students that would fall in each level.

Let participants know that they should make these decisions based on what they know about students in the state, the requirements of the test, and the standards. Recommendations:

- Do not change the ratings based solely on how you believe the impact data will be perceived. Think about whether the percentages represented by the impact data are an accurate reflection of how students currently should be distributed given the proficiency level descriptions and the content/skills measured by the test. Try to balance your concerns on what you believe to be appropriate given the content of the test and what others (in the state) will regard as acceptable.
- If you do not believe the proportion of students falling in each level is appropriate, do not arbitrarily modify the ratings (e.g, add 5% to each proportion in a given level). You have already given the items and the ratings, as well as conceptualization of the borderline students, a lot of thought, so don't throw that all away.
- How does a participant modify the ratings to influence proportion of students in a given proficiency level?

After participants have completed their discussions and indicate that they understand the impact data and the other data associated with Round 2, they will respond to the readiness survey. When participants answer "yes" to all of these questions, they will make their Round 3 Ratings.

**10:45-11:00 Round 3 Standard Setting**

- **Readiness Check**
- **Round 3 Ratings**

Check with the table leaders that everyone is ready for Round 3. At Round 3, the participants will be asked to determine their cut scores. Make sure the participants understand that they don't rate each item. Each participant should start with the rating for "Proficient" first followed by "Intermediate/Basic". Remind participants that the rating is always an independent activity. Collect the Rating Sheets as participants complete them.

**Important Notes**

**When a panelist completes Round 3**

- Collect Rating Sheets (group them by table)
- Spot check Rating Sheet

**When all have completed Round 3 ratings**

- Place in designated folder and give to RA

**11:00-11:30 Break**

RA will do analysis of Round 3 ratings

**11:30-11:45 Present Round 3 Results**

- **Final Recommended Cut Scores**
- **Impact Data**

No handouts. Present results on screen only.

**11:45-12:00 Standard Setting Closure**

- **Complete Survey on the Standard Setting Process**
- **Materials Collection**

Participants will be given evaluation forms to complete and return. The participants' ratings of the standard setting process and their comments will be solicited. Remind the participants that after they complete the forms, they need to leave all secured materials that have not already been collected (e.g., scratch paper etc.).

**Important Notes**

**Collect Secure Materials (**to be picked up**)**-

- Item Map
- Rating Sheet
- Test Booklet
- Supplemental Items Booklet
- Answer Document
- Answer Key
- Answer Key with P-Values
- Borderline Student Descriptors
- Any feedback data (charts and graphs provided after Rounds 1-3)

**12:00-1:00 LUNCH**

Performance Descriptor refinement will occur after lunch. Tell the panelists to meet back in the room after lunch at 1:00.

ADE will debrief with full committee and providing closing remarks during lunch.

**1:00-2:00 Performance Level Descriptor Refinement Discussion**
**2:00-2:30 Dismissal of Participants**
**2:30-4:00 Performance Level Descriptor Modification with Tableleaders**

The Facilitator presents instruction for refining PLDs. Panelists will be asked to discuss the definitions within each performance level, particularly with respect to the items immediately on either side of each bookmark and propose any final edits to the Borderline Student Descriptors that might be made to more clearly reflect the primary skill and knowledge attributes of students classified in each performance level. Discussion will take place within table groups and the table leader will take notes on the recommended changes. Once the table groups have completed their task, the table leaders will meet all together with ADE and Pearson to finalize the descriptors. Pearson RS will show the descriptors up on the screen and make the recommended changes as they are reported by the table leaders.

**Distribute the handout on Refinement of PLDs**

**Procedures**

1. The ELP standards must be available as a reference for this activity.
2. Begin with the bullets at the bottom. Determine if any bullets should move from one performance level to another. If a bullet (PO) is moved and the objective is noted in the narrative at the top of the PLD, the objective must also be moved to the new performance level in the narrative.

3. Since some bullets are a combination of POs, it may be necessary to break apart the bullet to place the separate parts in different performance levels. If needed, make the appropriate adjustments to the narrative.
4. Note the bullet's beginning action verb. The verb, along with the rest of the text, may be changed and kept at the original performance level or moved to another.
5. New bullets may be added if appropriate and necessary; however, removal of bullets is not recommended. All assessments must conform to the test blueprint, and although not all the bullets will be covered in the current assessment, over time, the future assessments will include all the performance objectives identified in the bullets.
6. **Adjust the narrative accordingly, but do not exceed the maximum amount of characters assigned to the space.**
7. Table Leaders will share their tables' recommendations, and Track Changes will be made to the existing document.

**4:00-4:30 ADE-Pearson Debrief**

# Appendix F: Opening Session Presentation Slides

## Standard Setting on the Arizona English Language Learner Assessment (AZELLA)

### Kindergarten Placement Test (KPT)

Opening Comments
July 11-12, 2012
Phoenix, Arizona

ALWAYS LEARNING PEARSON

## Welcome and Introductions

- Arizona Department of Education
  - Roberta Alley: Associate Superintendent
  - Leila Williams, Ph.D.: Deputy Associate Superintendent
  - Charlie Bruen, Ed.D.: Dir. of Data Analysis, Budget, & Technology
  - Frank Brashear: Dir. of Test & Item Development
  - Marlene Johnston: Dir. of English Language Learner Assessment
  - Irene Hunting: Dir. of State Test Administration
  - Lee Scott: Research Scientist
  - Linda Harvey AZELLA Administration Coordinator
  - Kimlee Buttacavoli Grant ELL Assessment Development Coordinator
- Pearson/WestEd
  - Steve Fitzpatrick, Ph.D.: Lead Research Scientist
  - Hiro Fukuhara, Ph.D.: Research Scientist
  - Erica Baltierra: Project Manager
  - Rich Young: Program Director
  - Beverly Nedrow: Content Specialist

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 2

ALWAYS LEARNING PEARSON

## Overview of AZELLA KPT

- History
- Purposes

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 3

ALWAYS LEARNING PEARSON

## Who is an ELL?

An ELL student is one that has been:

- Identified by the Home Language Survey
  - First language of student
  - Language spoken at home
  - Primary language used
- Taken the state assessment, AZELLA
- Scored less than Proficient

ALWAYS LEARNING PEARSON

## Demographics - Arizona ELLs

- Approximately 10% of Arizona's K-12 students are English Language Learners (ELL).
- Spanish is the most common non-English language home languages representing 80% of ELLs.
- Nearly 50% of all ELLs are in K-2.

ALWAYS LEARNING PEARSON

## Laws That Shape Arizona's ELL Programs

- Title VI of the Civil Rights Act of 1964
- Equal Educational Opportunities Act of 1974
- Lau v. Nichols (U.S.S. Ct. 1974)
- Castañeda v. Pickard 1981
- Arizona State Law – Proposition 203 (2002)
- Arizona State Law – HB 2064
- Title I
- Title III

ALWAYS LEARNING PEARSON

## SEI Model Key Components

- Entry and Exit based on AZELLA
- Students grouped by AZELLA proficiency level
- Specified time allocations
- ELLs are taught by Highly Qualified and trained teachers
- Classroom practice based on English Language Development (ELD) delineated by the Arizona English Language Proficiency Standards

ALWAYS LEARNING PEARSON

## ELD - *Definition*

- *"ELD"* means English language development, the teaching of English language skills to students who are in the process of learning English.
- It is distinguished from other types of instruction, e.g., math, science, or social science, in that the content of ELD emphasizes the English language itself.

ALWAYS LEARNING PEARSON

## What is AZELLA?

- If a student is identified as having a home language that is other than English, the student is required to be tested to determine if they need language instruction in order to have access to classroom instruction.
- In Arizona the test used is called AZELLA. Students receive proficiency level scores in Reading, Writing, Listening, Speaking, Oral and Comprehension.
- The AZELLA tests the Arizona English Language Proficiency Standards.

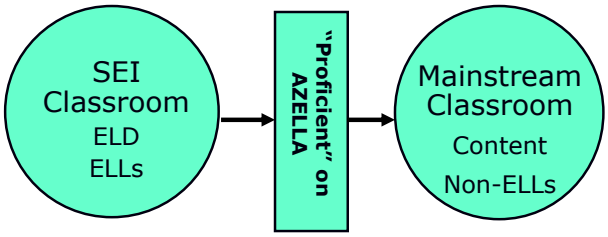
ALWAYS LEARNING PEARSON

## How is AZELLA Used

- To identify ELL students.
- To inform classroom placement.
- To inform instruction based on the Arizona English Language Proficiency Standards (ELPS).
- To measure annual progress in English.
- To exit ELL students from ELL programs.
- As a measurement and accountability tool.

ALWAYS LEARNING PEARSON

## *SEI Classroom v. Mainstream Classroom*




Entry and exit is based on AZELLA

ALWAYS LEARNING PEARSON

## Benefits of Ability Based Grouping

•Targeted instruction.

•Lesson is linguistically appropriate for ALL students.

•Students are not overwhelmed - otherwise proficient students do all of the talking.

•Accurate monitoring of student production.

•Specialists can be developed for each level.

•Minimizes students developing large language "gaps".

•Makes lesson planning easier for teachers.

ALWAYS LEARNING PEARSON

## ELL Assessment History

School Years 2005-2006

Statewide SELP implementation

School Years 2007-2009

AZELLA Form AZ-1

School Year 2009-2012

AZELLA Form AZ-2

School Year 2013

- Launch of AZELLA Kindergarten Placement Test
- Use of revised AZELLA for annual Spring end-of-year test
- AZELLA Form AZ-2 for Placement of Grades 1-12

ALWAYS LEARNING PEARSON

## AZELLA Kindergarten Placement Test

ALWAYS LEARNING PEARSON

## AZELLA Kindergarten Placement Test

- AZ educators asked for a test for Kindergarten PHLOTE students.
- Test time was targeted at 20 minutes. Short enough for engagement; long enough to gain meaningful information.
- Blueprint based on ELPS Listening and Speaking at the lower skill levels, and on the AZ Early Learning Standards.
- Item types were piloted in January 2012.
- Draft Performance Level Descriptors (PLDs) developed on February 24, 2012 .
- Field test was April 23-May 11.

ALWAYS LEARNING PEARSON

## AZELLA Kindergarten Placement Test Format

Approximate
Total Time: 14-22 minutes

| Part | Type of Administration | Materials Needed | Approximate Time |
|---|---|---|---|
| 1: Let's Get to Know Each Other! | one-on-one | •Student and Test Administrator Test Book (spiral bound)<br>•Student Response Sheet<br>•Stickers (Bag 1) | 2-4 minutes |
| 2: Picture Time! | one-on-one | •Student and Test Administrator Test Book (spiral bound)<br>•Student Response Sheet<br>•Laminated Activity Card<br>•Erasable Marker/Eraser | 4-6 minutes |
| 3: Let's Talk! | one-on-one | •Student and Test Administrator Test Book (spiral bound)<br>•Student Response Sheet | 4-6 minutes |
| 4: Story Time! | one-on-one | •Student and Test Administrator Test Book (spiral bound)<br>•Student Response Sheet<br>•Stickers (Bag 2) | 4-6 minutes |

## AZELLA Kindergarten Placement Test

*Administration Notes*

- All districts and charters will use this assessment beginning in the 2012-13 school year to assess non-English PHLOTE incoming Kindergarten students.
- There will be required online training for Test Administrators.
- Student responses will be captured, scored and uploaded via PearsonAccess to return proficiency level results.
- Test will be available late July or early August.
- The annual end-of-year AZELLA will be a separate test.

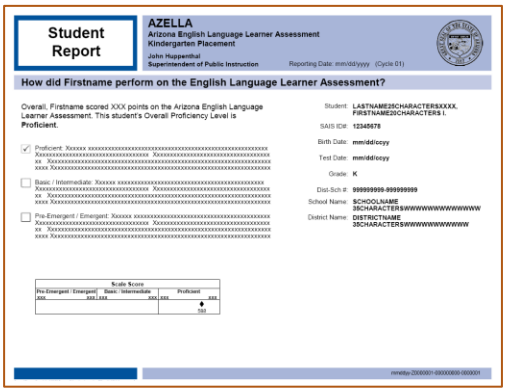
ALWAYS LEARNING PEARSON

## AZELLA Proficiency Levels for Kindergarten Placement Test

- Pre-Emergent/Emergent
- Basic/Intermediate
- Proficient

ALWAYS LEARNING PEARSON

## Kindergarten Placement Test
# Individual Student Report


Student Report

AZELLA
Arizona English Language Learner Assessment
Kindergarten Placement
John Huppenthal
Superintendent of Public Instruction
Reporting Date: mm/dd/yyyy (Cycle 01)

How did Firstname perform on the English Language Learner Assessment?

Overall, Firstname scored XXX points on the Arizona English Language Learner Assessment. This student's Overall Proficiency Level is **Proficient.**

Proficient: Xxxxxx xxxxx

## Why You Are Here

- The purpose of this standard setting is to establish 'recommended' cut scores on the AZELLA KPT.
- You were selected to serve on this committee for a variety of reasons:
  - Familiarity with the knowledge and skills required to "master" the English Language Proficiency standards at various proficiency levels
  - Representation of various jurisdictions and demographic characteristics

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 22

ALWAYS LEARNING PEARSON

## Standard Setting Overview

Panelists will:

- Review and discuss the test.
- Develop a shared understanding of the Performance Levels (PLDs).
- Develop "Borderline Student" Descriptors.
- Receive Standard Setting Training and Practice.
- Participate in three rounds of ratings –
  - Round 1: Independent
  - Round 2: Independent, but with table discussion
  - Round 3: Independent, but with table & full group discussion
- Finalize Performance Level Descriptors.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 23

ALWAYS LEARNING PEARSON

## Logistics

- Location of Meals and Breaks
- Security Forms
- Reimbursement Forms

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 24

ALWAYS LEARNING PEARSON

## Security

- DO NOT:
  - Remove any secure materials from the room on breaks or at end of day.
  - Discuss cut scores (yours or others) with anyone outside of the meeting.
  - Discuss secure materials with non-participants.
- Notes should be taken on our materials only.
- Write your Panelist ID number on all materials.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.


## What is Standard Setting?

- Process used to determine recommended cut scores on an assessment that will classify student performance into different categories
  - Provides a frame of reference for the interpretation of test scores
  - A semi-quantitative, semi-standardized judgment process
  - A routine, daily activity for teachers

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.


## What are Standards?

- English Language Proficiency Standards
  - Content standards specify the curriculum that ELL students are taught and expected to learn.
- Performance Standards
  - Performance standards specify the level of knowledge of that content that students must demonstrate to be categorized into a proficiency level.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

## Proficiency Levels for AZELLA KPT

- Pre-emergent/Emergent
- Basic/Intermediate
- Proficient

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 28

ALWAYS LEARNING PEARSON

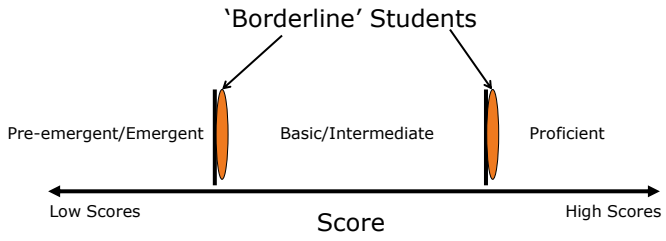
## Borderline Student Descriptions

- The cut score is set at the beginning of the proficiency level.
  - Basic/Intermediate
  - Proficient
- Create Borderline Student Descriptions for each of those levels.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 29

ALWAYS LEARNING PEARSON

## Three Proficiency Levels: Two Cuts




Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 30

ALWAYS LEARNING PEARSON

## Standard Setting: Modified Angoff Procedure

- For single point items:
  - Determine the percent of borderline students who would get the item correct
  - Example: "Given the knowledge, skills, and abilities that are required in this item, what would be the percent of borderline proficient students who would answer correctly?"
- For multiple point items:
  - Determine the average score of borderline students
  - Example: "Given the knowledge, skills, and abilities that are required in this item, what would be the average score of the item over a group of borderline proficient students?"
  - It can help to think about the percent of borderline students who would receive each score point
- Record your rating of each item for each proficiency cut (Basic/Intermediate and Proficient).

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 31

ALWAYS LEARNING PEARSON

## Recap of Activities

Panelists will:

- Review and discuss the test.
- Develop a shared understanding of the Proficiency Levels.
- Develop "Borderline Student" Descriptors.
- Receive Standard Setting Training and Practice.
- Participate in three rounds of ratings –
  - Round 1: Independent
  - Round 2: Independent, but with table discussion
  - Round 3: Independent, but with table & full group discussion
- Finalize Performance Level Descriptors.
- Provide an evaluation of the Process.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 32

ALWAYS LEARNING PEARSON

# Appendix G: Breakout Session Presentation Slides

**Standard Setting on**
**Arizona English Language Learner Assessment (AZELLA)**
**Kindergarten Placement Test**

**Breakout Session – Day1**

1

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

ALWAYS LEARNING PEARSON

## Committee Introductions

- Name
- Where are your from?
- How long you have been in your current position/field?
- What educational roles you have fulfilled?
- Have you participated in a standard setting before?
- Tell us something interesting about yourself.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 2

ALWAYS LEARNING PEARSON

## Day 1 Overview

- Discuss Performance Level Descriptions (PLDs)
- Create borderline student descriptions
- Experience the Kindergarten Placement Test
- Round 1 Rating

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 3

ALWAYS LEARNING PEARSON

## Performance Level Descriptors (PLDs)

- Performance Levels for Kindergarten Placement Test
  - Pre-emergent/Emergent
  - Basic/Intermediate
  - Proficient
- Read through the PLDs to get an idea of the skills described at each level.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 4

ALWAYS LEARNING PEARSON

## Understanding the Performance Levels

- Later we will have discussions about what distinguishes the performance levels
  - Compare "Basic/Intermediate" to "Proficient"
  - Compare "Pre-emergent/Emergent" to "Basic/Intermediate"
- And develop Borderline Descriptors
- But first….

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 5

ALWAYS LEARNING PEARSON

## Review Kindergarten Placement Test

- Gain an appreciation of the assessment
- Understand how the Kindergarten Placement Test is administered
- Understand how each item is scored
- Work independently
- Group discussion after everyone has completed and scored their test

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 6

ALWAYS LEARNING PEARSON

## Group Discussion About the Test

- What are your general impressions about the test?
- Did the test generally cover the depth and breadth of the content standards?
- Does the test generally have a range of item difficulties (e.g., easier items, moderate items, difficult items)?

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 7

ALWAYS LEARNING PEARSON

## Lunch Time!

- Please take 45-minute break for lunch.
- Reconvene in this room at 12:30.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 8

ALWAYS LEARNING PEARSON

## Performance Level Descriptors (PLD)

- Performance Levels for Kindergarten Placement Test
  - Pre-emergent/Emergent
  - Basic/Intermediate
  - Proficient
- Read the descriptors of the performance levels.
- What distinguishes each level?
  - Compare "Basic/Intermediate" to "Proficient"
  - Compare "Pre-emergent/Emergent" to "Basic/Intermediate"

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 9

ALWAYS LEARNING PEARSON

## Gaining an Understanding of the Performance Levels

- Within each table group, ask, "What should students know and be able to do at each level?"
  - "Basic/Intermediate," "Proficient"
- Appoint a recorder to write on the flip chart.
- Suggestions should be:
  - Concrete.
  - Clearly related to the PLDs.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 10

ALWAYS LEARNING PEARSON

## Understanding the Performance Level: "Proficient"

- Describe concretely the students who are at "Proficient."
  - What should they be able to *do*?
  - What *skills* should they possess?
  - What should they *know*?
  - What *language skills* are necessary to access mainstream curriculum in English?

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 11

ALWAYS LEARNING PEARSON

## Understanding the Performance Level: "Basic/Intermediate"

- Describe concretely the students who are at "Basic/Intermediate"
  - What should they be able to *do*?
  - What *skills* should they possess?
  - What should they *know*?
  - What English *language skills* demonstrate that they are at "Basic/Intermediate"?

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 12

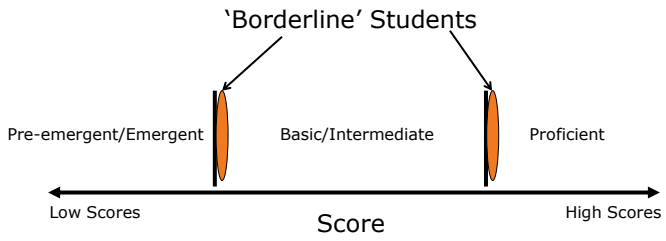
ALWAYS LEARNING PEARSON

## Establishing Recommended Cut Scores

- The cut score is set at the beginning of the performance level:
  - Basic/Intermediate
  - Proficient
- When determining cut scores, we need to think about the "Borderline Student" descriptors for that performance level:
  - The "borderline student" just barely makes it into the performance level.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.


## Establishing Recommended Cut Scores: Finding the Cut




Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.


## Examples of "Real World" Performance Levels

- Total Blood Cholesterol Level
  - Less than 200 mg/dL: Desirable
  - 200–239 mg/dL: Borderline-High Risk
  - 240 mg/dL and over: High Risk
- Blood Sugar Level
  - Normal Levels: 70 - 150 mg
  - High: above 150 mg

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

## Distinguishing "Proficient" from "Basic/Intermediate"

- Think about the borderline students at "Proficient"
  - Identify three characteristics or behaviors that MOST distinguish a student who just barely "Proficient"
  - Record the three responses on your flipchart at each table

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 16

ALWAYS LEARNING PEARSON

## Distinguishing "Basic/Intermediate" from "Pre-emergent/Emergent"

- Think about the borderline students at "Basic/Intermediate"
  - Identify three characteristics or behaviors that MOST distinguish a student who just barely "Basic/Intermediate"
  - Record the three responses on your flipchart at each table

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 17

ALWAYS LEARNING PEARSON

## Borderline Student PLDs: Group Discussion

- Reconvene as whole committee.
- Each table presents their examples of, "What should students know and be able to do at each level?"
- Each table describes the three distinguishing characteristics.
- Look for differences and commonalities across tables.
- The facilitator will capture the discussion on the group flip chart.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 18

ALWAYS LEARNING PEARSON

## Recap of Completed Activities So Far

- Reviewed the test
- Reviewed Performance Level Descriptors (PLDs)
  - Pre-emergent/Emergent
  - Basic/Intermediate
  - Proficient
- Developed the Borderline Student PLDs
  - Just 'barely' at Basic/Intermediate
  - Just 'barely' at Proficient

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

ALWAYS LEARNING PEARSON

## Time for a Break!

- Please take a 15-minute break.
- Reconvene in this room at 2:45.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

ALWAYS LEARNING PEARSON

## Practice Round

- Let's practice ratings of 10 items on the practice round test booklet.
  - 8 one point items
  - 1 two point item
  - 1 three point item
- Make a judgment about each item for a 'Proficient' cut
  - Determine a percent of borderline Proficient students who would answer each one point item correctly
  - Determine the average score of borderline Proficient students on each multiple point item
- Work on the ratings individually

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

ALWAYS LEARNING PEARSON

## Round 1 - Readiness Survey

- Consider the task we ask of you.
- Answer the questions on the Readiness Survey for Round 1.
- Table leaders give the thumbs up when everyone at table is ready to go.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 22

ALWAYS LEARNING PEARSON

## Round 1

- Start with a 'Proficient' cut
- Remember to consider the 'borderline' students when rating items
- Recall discussions about performance levels and 'borderline' student descriptors
- Once you rate all items for the 'Proficient' cut, start over and provide ratings for a 'Basic/Intermediate' cut
- Work on the ratings individually – please refrain from discussing ratings or items

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 23

ALWAYS LEARNING PEARSON

## Completion of Round 1

- Once completed, your table leader and/or facilitator will collect and check in all of your materials.
- See you back tomorrow morning at 7:30 for breakfast.
- Meeting starts at 8:00.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 24

ALWAYS LEARNING PEARSON

**Standard Setting on**
**Arizona English Language Learner Assessment (AZELLA)**
**Kindergarten Placement Test**

**Breakout Session – Day2**

2

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 25

ALWAYS LEARNING PEARSON

## Day 2 Overview

- Round 1 Feedback
- Round 2 Standard Setting
- Round 2 Feedback
- Round 3 Standard Setting
- Round 3 Results
- Revisit Performance Level Descriptors.
- Complete Survey.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 26

ALWAYS LEARNING PEARSON

## Round 1 Feedback

- We provide the following feedback:
  - Frequency of ratings for all items/cuts for the whole committee
  - Individual cut score at each performance level
  - Median cut scores for the whole committee
- Additional information
  - Empirical Item Difficulty

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 27

ALWAYS LEARNING PEARSON

## Empirical Item Difficulty

- Item mean is a indication of how difficulty an item is.
- For 1 point items, the range is from 0 to 1.
  - High item mean indicates that an item is easy
  - Low item mean indicates that an item is difficult
- For multiple point items, the range is 0 to the item's maximum point value.
- Data tell how students *DID* perform
- Data CANNOT tell how students SHOULD perform *nor* how students at the borderline of "Proficient," "Basic/Intermediate" perform.
- Data is based on all PHLOTE students who participated in the field test in Spring 2012

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 28

ALWAYS LEARNING PEARSON

## Round 1 Feedback and Discussion

- Let's talk about the ratings of each item for each performance level cut
  - How similar are your ratings compared to the group (i.e., are there panelists who are more lenient or stringent that the other panelists)?
  - Do panelists have different conceptualizations of the 'just-barely' students at the performance level?
  - How similar are your ratings compared to the empirical item difficulty?
- What about your cut scores compared to the group median cut scores?
- Remember that consensus is not a requirement
- All of you are experts – everyone in this committee should have an opportunity to participate in the discussion

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 29

ALWAYS LEARNING PEARSON

## Why Round 2?

- You are now an improved advisor.
- Consider judgments & views of your peers.
- Consider student achievement data.
- Goal: NOT consensus, but *reflection*

**YOU ARE NOW A *BETTER* ADVISOR, because you are a better informed advisor.**

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 30

ALWAYS LEARNING PEARSON

## Round 2 – What to Do?

- Reflect on earlier ratings – yours and those of your peers.
- Reflect on the table and committee discussion.
- Think about the panelist agreement data and item means.
- Decide if you want to modify your ratings.
- Remember to consider the 'borderline' students when rating items
- Recall discussions about performance levels and 'borderline' student descriptors

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 31

ALWAYS LEARNING PEARSON

## Round 2 - Readiness Survey

- Consider the task we ask of you.
- Answer the questions on the Readiness Survey for Round 2.
- Table leaders give the thumbs up when everyone at table is ready to go.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 32

ALWAYS LEARNING PEARSON

## Round 2

- Start with the 'Proficient' cut
- If you decide to keep the same rating as Round 1 for an item, you can leave it blank for the item.
- Once you rate all items for the 'Proficient' cut, start over and provide ratings for a 'Basic/Intermediate' cut
- Work on the ratings individually – please refrain from discussing ratings or items

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 33

ALWAYS LEARNING PEARSON

## Completion of Round 2

- Once completed, your table leader and/or facilitator will collect and check in all of your materials.
- See you back in this room at 10:00

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 34

ALWAYS LEARNING PEARSON

## Round 2 Feedback

- We provide the following feedback:
  - Frequency of ratings on all items/cuts at committee level
  - Individual cut score at each performance level
  - Median cut scores for the whole committee
- Additional information
  - Impact data

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 35

ALWAYS LEARNING PEARSON

## Impact Data

- The impact data show the percentage of students in each of the performance levels based on the current cut score recommendations.
- The current cut score recommendations are based on the median of committees' recommended cut scores.
- The impact data are based on the Spring 2012 field test administration - Same sample as the item mean.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 36

ALWAYS LEARNING PEARSON

## Impact Data

- Attach the impact data here

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 37

ALWAYS LEARNING PEARSON

## Round 2 Discussion on Impact Data

- As a group, let's discuss about the impact data
- How do the impact data align with your expectations?
- Reminder – the purpose of reviewing impact data is to provide a reality check for judgments

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 38

ALWAYS LEARNING PEARSON

## "How do I know if I'm right?"

- There is no "*right.*"
- Everybody is an expert.
- Remember to keep in mind:
  - "*Should*"
  - The *borderline students*
  - Our table and committee discussions

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 39

ALWAYS LEARNING PEARSON

## Round 3 - What to Do ?

- Reflect on earlier ratings – yours and peers
- Reflect on the table and committee discussions
- Think about the panelist agreement data and impact data

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 40

ALWAYS LEARNING PEARSON

## Round 3 - Readiness Survey

- Consider the task we ask of you.
- Answer the questions on the Readiness Survey for Round 3.
- Table leaders give the thumbs up when everyone at table is ready to go.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 41

ALWAYS LEARNING PEARSON

## Round 3

- You will determine cut scores instead of rating each item at this round
- Fill out your recommended cut scores on the top of rating sheet under 'Round 3'
- Start with a 'Proficient' cut and then 'Basic/Intermediate' cut

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 42

ALWAYS LEARNING PEARSON

## Completion of Round 3

- Once completed, your table leader and/or facilitator will collect and check in all of your materials.
- See you back in this room at 11:30

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.
43

ALWAYS LEARNING PEARSON

## Final Recommended Cuts and Impact Data

- Insert the recommended cuts and impact data here

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.
44

ALWAYS LEARNING PEARSON

## Standard Setting Evaluation

- Please complete the evaluation form.
- Table leader will help to coordinate the order of materials for easy check-in.
- Facilitator will pick up and check in materials.
- THANK-YOU!

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.
45

ALWAYS LEARNING PEARSON

## Lunch Time!

- Please return at 1:00.
- Reconvene in this room.
- Refinement of Performance Level Descriptors

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.
46

ALWAYS LEARNING PEARSON

## Steps in PLD Review and Revision

1. The ELP standards must be available as a reference for this activity.
2. Begin with the bullets at the bottom. Determine if any bullets should move from one performance level to another. If a bullet (PI) is moved and the objective is noted in the narrative at the top of the PLD, the objective must also be moved to the new performance level in the narrative.
3. Since some bullets are a combination of PIs, it may be necessary to break apart the bullet to place the separate parts in different performance levels. If needed, make the appropriate adjustments to the narrative.
4. Note the bullet's beginning action verb. The verb, along with the rest of the text, may be changed and kept at the original performance level or moved to another.
5. New bullets may be added if appropriate and necessary; however, removal of bullets is not recommended. All assessments must conform to the test blueprint, and although not all the bullets will be covered in the current assessment, over time, the future assessments will include all the performance indicators identified in the bullets.
6. Adjust the narrative accordingly.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.
47

ALWAYS LEARNING PEARSON

## Steps in PLD Review and Revision

- Table Leaders will share their tables' recommendations, and Track Changes will be made to the existing document.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.
48

ALWAYS LEARNING PEARSON

# Appendix H: Standard Setting Methodology Training Slides

## Standard Setting on Arizona English Language Learner Assessment (AZELLA)

### Kindergarten Placement Test

Standard Setting Training
July 11-12, 2012
Phoenix, Arizona

ALWAYS LEARNING PEARSON

### Purpose of Presentation

- The purpose of this session is to introduce you to the process that we will use to establish 'recommended' cut-scores on the Arizona English Language Learner Assessment (AZELLA) Kindergarten Placement Test

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 2

ALWAYS LEARNING PEARSON

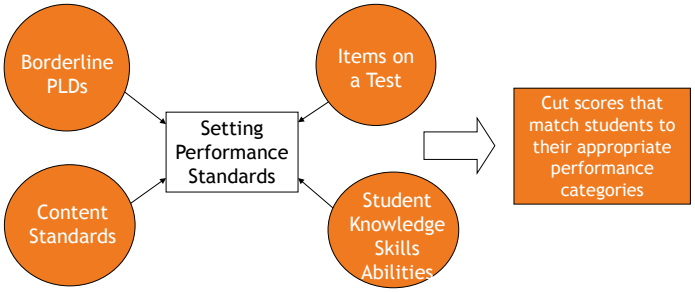
### What is Standard Setting?

- A process of deriving levels of performance on educational or professional assessments, by which decisions or classifications of persons will be made (Cizek, 2006)
- Test scores can be used to group students into meaningful performance levels
- Standard setting is the process whereby we "draw the lines" that separate the test scores into various performance levels

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 3

ALWAYS LEARNING PEARSON

## Setting Performance Standards




Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 4

ALWAYS LEARNING PEARSON

## Standard-Setting Method

- The Modified Angoff[1] method is the approach we are using to set performance standards
- Panelists consider the difficulty of each item and expectations of test-takers to render item-level ratings
- Ratings are quantified as a percentage for one point items and an average score for multiple point items.

[1]Angoff, W.H. (1971). Scales, norms and equivalent scores. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, DC: American Council on Education.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 5

ALWAYS LEARNING PEARSON

## Modified Angoff Method

- For single point items:
  - Determine the percent of borderline students who would get the item correct
  - Example: "Given the knowledge, skills, and abilities that are required in this item, what would be the percent of borderline proficient students who would answer correctly?"
- For multiple point items:
  - Determine the average score of borderline students
  - Example: "Given the knowledge, skills, and abilities that are required in this item, what would be the average score of the item over a group of borderline proficient students?"
  - It can help to think about the percent of borderline students who would receive each score point

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 6

ALWAYS LEARNING PEARSON

## Modified Angoff Method (Continued)

- A panelist's recommended cut will be determined by aggregating his/her ratings of all items.
- A committee's recommended cut will be determined by taking a median of all panelists' recommended cuts.
- Multiple rounds of judgments and delivery of information is designed to optimize decision making

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 7

ALWAYS LEARNING PEARSON

## Rating Process

- Record a percentage for each one point item
  - Record ratings in 5% intervals (i.e., 50, 55, 60, etc.) on your rating sheet
  - The rating should be between 0 and 100.
- Record an average item score for each multiple point item
  - Record ratings in 5 intervals (i.e., 120, 125, 130, etc.) on your rating sheet
  - The rating should be between 0 and an item's possible maximum point.
- Rate every item for a 'Proficient' cut point, then start over and rate each item again for a 'Basic/Intermediate' cut point
  - Each item will receive two ratings
  - A percentage or average item score for a Proficient cut <u>must be</u> greater than or equal to that for a Basic/Intermediate cut

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 8

ALWAYS LEARNING PEARSON

## Rating Sheet

**Participant ID:** 20

| Page | Question | Round 1 | | Round 2 | |
|---|---|---|---|---|---|
| | | Proficient | Basic/ Intermediate | Proficient | Basic/ Intermediate |
| | 1 | 90 | | | |
| | 2 | 75 | | | |
| | 3 | 80 | | | |
| | 4 | 85 | | | |
| | 5 | 75 | | | |
| | 6 | 70 | | | |
| | 7 | 85 | | | |
| | 8 | 90 | | | |
| | 9 | 125 | | | |
| | 10 | 150 | | | |

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 9

ALWAYS LEARNING PEARSON

## Rating Sheet

Participant ID: 20

| Page | Question | Round 1 | | Round 2 | |
|---|---|---|---|---|---|
| | | Proficient | Basic/ Intermediate | Proficient | Basic/ Intermediate |
| | 1 | 90 | 70 | | |
| | 2 | 75 | 50 | | |
| | 3 | 80 | 70 | | |
| | 4 | 85 | 65 | | |
| | 5 | 75 | 50 | | |
| | 6 | 70 | 60 | | |
| | 7 | 85 | 75 | | |
| | 8 | 90 | 75 | | |
| | 9 | 125 | 75 | | |
| | 10 | 150 | 100 | | |

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.


ALWAYS LEARNING PEARSON

## Worksheet for Multiple Point Items

| Item 9 | Proficient | | | |
|---|---|---|---|---|
| Score | 0 | 1 | 2 | |
| 1. Percentage | 25 | 25 | 50 | Sum |
| 2. Score × Percentage | | | | |

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.


ALWAYS LEARNING PEARSON

## Worksheet for Multiple Point Items

| Item 9 | Proficient | | | |
|---|---|---|---|---|
| Score | 0 | 1 | 2 | |
| 1. Percentage | 25 | 25 | 50 | Sum |
| 2. Score × Percentage | 0 | 25 | 100 | 125 |

Your rating

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.


ALWAYS LEARNING PEARSON

## Summary of Standard Setting Procedure

- Modified Angoff method will be used for this standard setting
- Panelists will make a judgment about each item considering 1) Borderline Students, 2) Content Standards, 3) Knowledge, Skills, Abilities Required, and 4) Item
- Panelists will determine
  - what percent of borderline students would answer an item correctly for (One point items)
  - what the average score of an item for borderline students would be (Multiple point items)
- Panelists will rate all items for a 'Proficient' cut, and then rate all items for a 'Basic/Intermediate' cut on a rating sheet.
  - Two ratings for each item
- Panelists will go through a multiple rounds of ratings.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 13

ALWAYS LEARNING PEARSON

## Any Questions?

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved. 14

ALWAYS LEARNING PEARSON

## Appendix I: Standard Setting Panelist Readiness Form

**ARIZONA ENGLISH LANGUAGE LERNER ASSESSMENT (AZELLA)**
**STANDARD SETTING**
**ROUND READINESS FORM**
**KINDERGARTEN PLACEMENT TEST**

Panelist ID: _____

Instructions: Please circle your response to the following questions.

| **Round 1** | | |
|---|---|---|
| I understand my task for Round 1. | **No** | **Yes** |
| I am ready to begin Round 1. | **No** | **Yes** |

| **Round 2** | | |
|---|---|---|
| I understand my task for Round 2. | **No** | **Yes** |
| I understand the data that was presented from Round 1. | **No** | **Yes** |
| I am ready to begin Round 2. | **No** | **Yes** |

| **Round 3** | | |
|---|---|---|
| I understand my task for Round 3. | **No** | **Yes** |
| I understand the data that was presented from Round 2. | **No** | **Yes** |
| I am ready to begin Round 3. | **No** | **Yes** |

## Appendix J: Example Standard Setting Rating Sheet

| Round 1: Cut Scores | | Round 2: Cut Scores | | Round 3: Cut Scores | |
|---|---|---|---|---|---|
| Basic/ Intermediate | | Basic/ Intermediate | | Basic/ Intermediate | |
| Proficient | | Proficient | | Proficient | |

| Page | Question | Round 1 | | Round 2 | |
|---|---|---|---|---|---|
| | | Proficient | Basic/ Intermediate | Proficient | Basic/ Intermediate |
| | 1 | | | | |
| | 2 | | | | |
| | 3 | | | | |
| | 4 | | | | |
| | 5 | | | | |
| | 6 | | | | |
| | 7 | | | | |
| | 8 | | | | |
| | 9 | | | | |
| | 10 | | | | |
| | 11 | | | | |
| | 12 | | | | |
| | 13 | | | | |
| | 14 | | | | |
| | 15 | | | | |
| | 16 | | | | |
| | 17 | | | | |
| | 18 | | | | |
| | 19 | | | | |

# Appendix K: Standard Setting Evaluation Forms with Responses

# Arizona English Language Learner Assessment (AZELLA)
## STANDARD SETTING
## DECISION MAKING FACTOR SURVEY
## KINDERGARTEN PLACEMENT TEST

Directions: Please respond to each statement by placing an "X" in the box corresponding to your opinion.

| **How much did each of the following factors influence your decisions on the cut score recommendations for the Arizona English Language Learner Assessment Kindergarten Placement Test?** | | Not at All | Somewhat | Moderately | Strongly | Very Strongly | No Response |
|---|---|---|---|---|---|---|---|
| 1 | Your experience in education | 0 | 0 | 1 | 2 | 10 | 0 |
| 2 | Prior to this standard setting meeting, your perceptions about students in each of the three performance levels | 0 | 1 | 4 | 5 | 2 | 1 |
| 3 | Your prior knowledge about standard setting | 4 | 4 | 1 | 4 | 0 | 0 |
| 4 | The orientation on standard setting | 0 | 1 | 3 | 7 | 2 | 0 |
| 5 | Your perception of the high stakes versus low stakes context of the AZELLA Kindergarten Placement Test | 3 | 1 | 3 | 2 | 4 | 0 |
| 6 | Your thinking about students in each performance level with whom you have had experience | 0 | 0 | 1 | 6 | 6 | 0 |
| 7 | The consequences of your decisions for No Child Left Behind (NCLB) | 3 | 3 | 1 | 3 | 3 | 0 |
| 8 | Your concerns about district or state political or economic issues | 4 | 2 | 0 | 3 | 4 | 0 |
| 9 | Your understanding of the performance level descriptors | 0 | 0 | 0 | 8 | 5 | 0 |
| 10 | Your understanding of the borderline performance level descriptors | 0 | 0 | 1 | 9 | 3 | 0 |
| 11 | The empirical item difficulty (item mean) presented after Round 1 | 0 | 2 | 5 | 4 | 2 | 0 |
| 12 | Frequency of ratings presented after Round 1 | 0 | 2 | 4 | 6 | 1 | 0 |
| 13 | Median cut scores presented after Round 1 | 1 | 2 | 4 | 5 | 1 | 0 |
| 14 | Median cut scores presented after Round 2 | 1 | 2 | 3 | 6 | 0 | 1 |
| 15 | The impact data presented after Round 2 | 1 | 1 | 4 | 7 | 0 | 0 |
| 16 | Your interactions with your fellow panelists before **Round 2** | 0 | 2 | 3 | 7 | 1 | 0 |
| 17 | Your interactions with your fellow panelists before **Round 3** | 1 | 2 | 2 | 6 | 2 | 0 |

Directions: Please respond to each statement by placing an "X" next to the category that best describes your school.

1. In general, my school/educational institution **mostly** serves students in the following socioeconomic status (choose one):

   _3_ Lower _6_ Lower/Middle _2_ Middle _2_ Upper Middle _0_ Upper

2. My educational institution is a charter school (choose one): _0_ Yes _13_ No

# ARIZONA ENGLISH LANGUAGE LERNER ASSESSMENT (AZELLA)
## STANDARD SETTING
## FINAL STANDARD SETTING EVALUATION FORM
## KINDERGARTEN PLACEMENT TEST

Directions: Please respond to each statement by placing an "X" in the box corresponding to your opinion. If you have any additional comments, please write them in the space provided at the end of this form.

**NOTE: SD=Strongly Disagree; D=Disagree; A=Agree; SA=Strongly Agree; NR = No Response**

| | Statement | SD | D | A | SA | NR |
|---|---|---|---|---|---|---|
| 1 | The workshop was well organized. | 0 | 0 | 2 | 11 | 0 |
| 2 | The training materials were helpful. | 0 | 0 | 4 | 9 | 0 |
| 3 | The modified Angoff method for providing the ratings was conceptually clear. | 0 | 1 | 4 | 8 | 0 |
| 4 | I had a good understanding of what the test was intended to measure. | 0 | 0 | 2 | 11 | 0 |
| 5 | I had a good understanding of Performance Level Descriptors. | 0 | 0 | 5 | 8 | 0 |
| 6 | Borderline Performance Level Descriptors helped me determine the rating of each item. | 0 | 0 | 6 | 7 | 0 |
| 7 | The practice round of ratings was helpful to understand what to do at Round 1. | 0 | 1 | 5 | 7 | 0 |
| 8 | After the first round of ratings, I felt comfortable with the standard setting procedure. | 0 | 0 | 6 | 7 | 0 |
| 9 | I found the feedback on empirical item difficulty (item mean) after Round 1 useful. | 0 | 0 | 3 | 10 | 0 |
| 10 | I found the feedback on the frequency of ratings after Round 1 useful. | 0 | 1 | 4 | 8 | 0 |
| 11 | I found the feedback on median cut scores after Round 1 useful. | 0 | 0 | 5 | 8 | 0 |
| 12 | I found the feedback on median cut scores after Round 2 useful. | 0 | 0 | 6 | 7 | 0 |
| 13 | I found the feedback on the percentage of the students tested that would be classified at each performance level (Impact Data) after Round 2 useful. | 0 | 0 | 8 | 5 | 0 |
| 14 | Discussion after Rounds 1 and 2 was open and honest. | 0 | 0 | 3 | 10 | 0 |
| 15 | I believe that my opinions were considered and valued by my group. | 0 | 0 | 4 | 9 | 0 |
| 16 | I am confident that my round 3 ratings for "*Basic/Intermediate*" reflect the knowledge, skills, and abilities described in the performance level descriptors. | 0 | 0 | 5 | 8 | 0 |
| 17 | I am confident that my round 3 ratings for "*Proficient*" reflect the knowledge, skills, and abilities described in the performance level descriptors. | 0 | 0 | 5 | 8 | 0 |
| 18 | I would defend the standards recommended by our committee. | 0 | 0 | 4 | 9 | 0 |
| 19 | Overall, I valued the workshop as a professional development experience. | 0 | 0 | 2 | 11 | 0 |

Please feel free to add comments on any of your responses above, make suggestions to improve future standard settings, and/ or tell us what you liked and did not like about this workshop. Thank you.

_____

_____

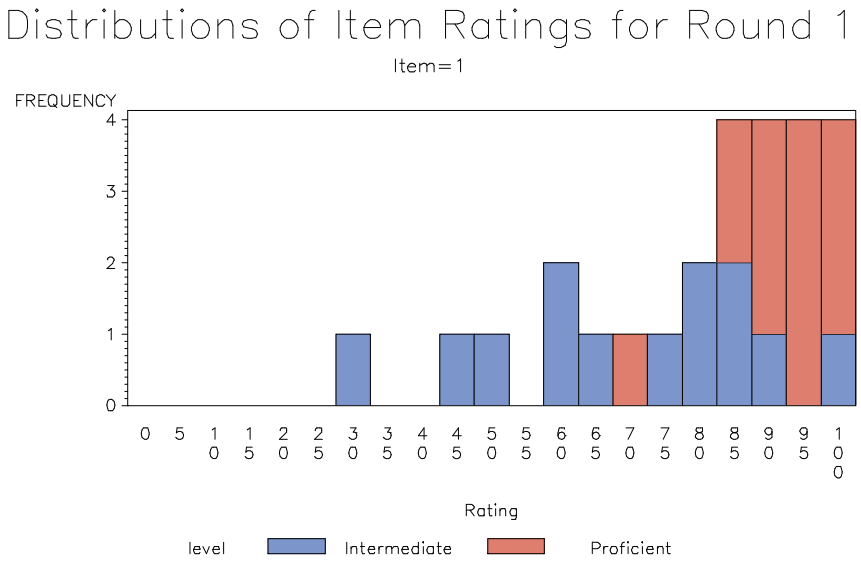
_____

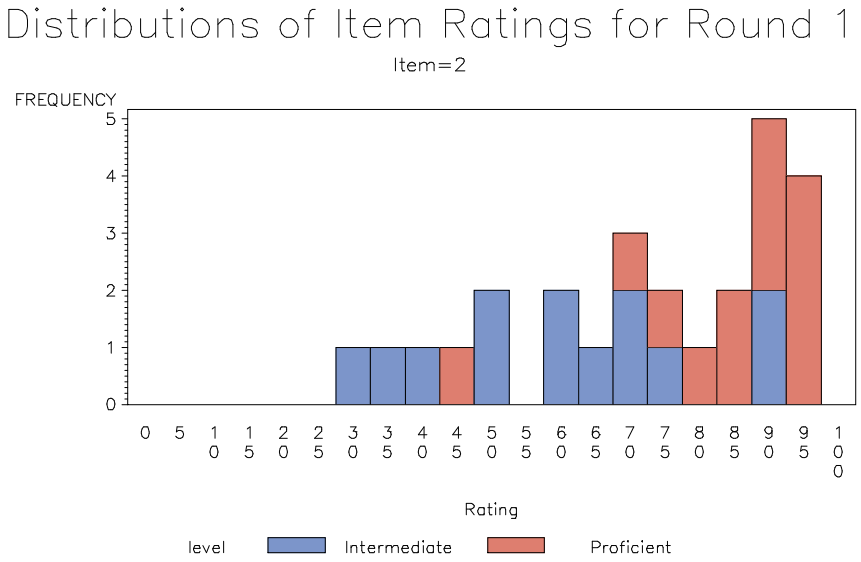
The Comment Made on the Evaluation Form by the Panelists

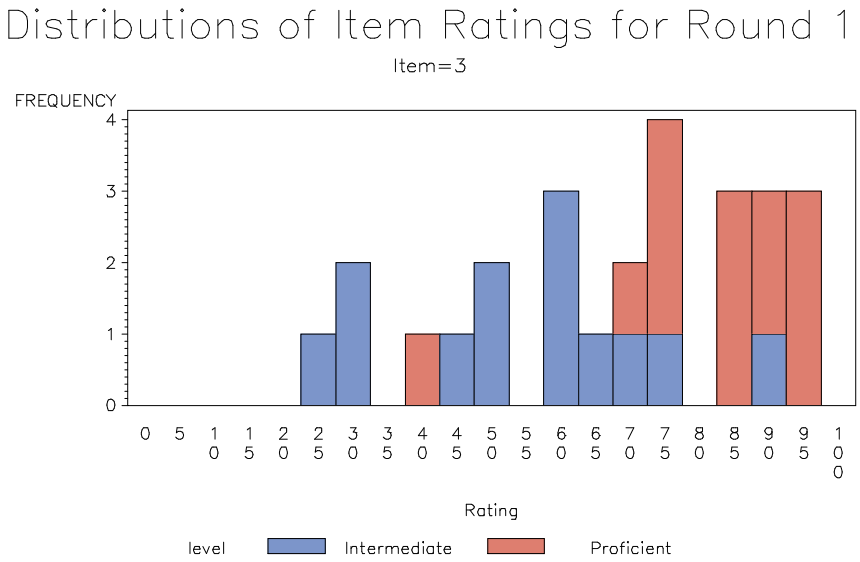
| |
|---|
| I learned a lot, would have liked to discuss individual items more - went through questions fast. |
| Enjoy being here! Very honored! Good Work By All! |
| This was a great experience. So often we just give tests and don't get to see what goes into setting score. |
| I really enjoyed my item on the committee. I felt that it was a wonderful learning experience too! |
| I felt very honored to be part of this. This conference was very organized, professional, and discussed items that are relevant to what Kindergartners need at the beginning of the year. |
| Thank you for the opportunity to participate. It was a productive experience in every regard. |
| It was a great experience getting to see how cut scores are determined. |

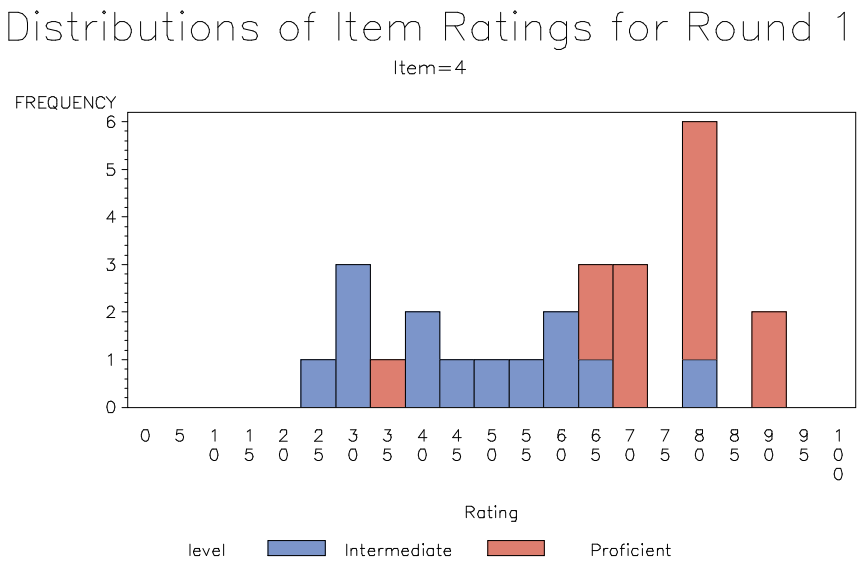
# Appendix L: Standard Setting Results

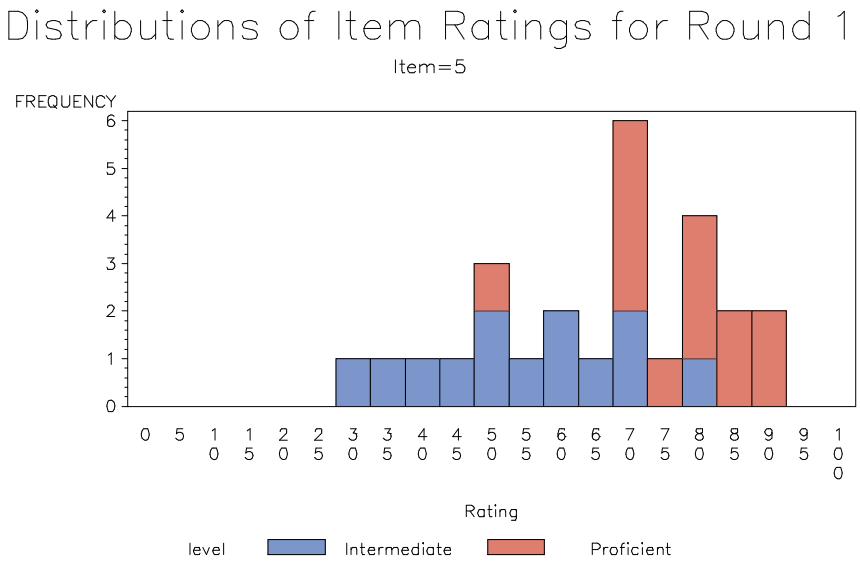
## Appendix L.1: Rating Distribution at Round 1

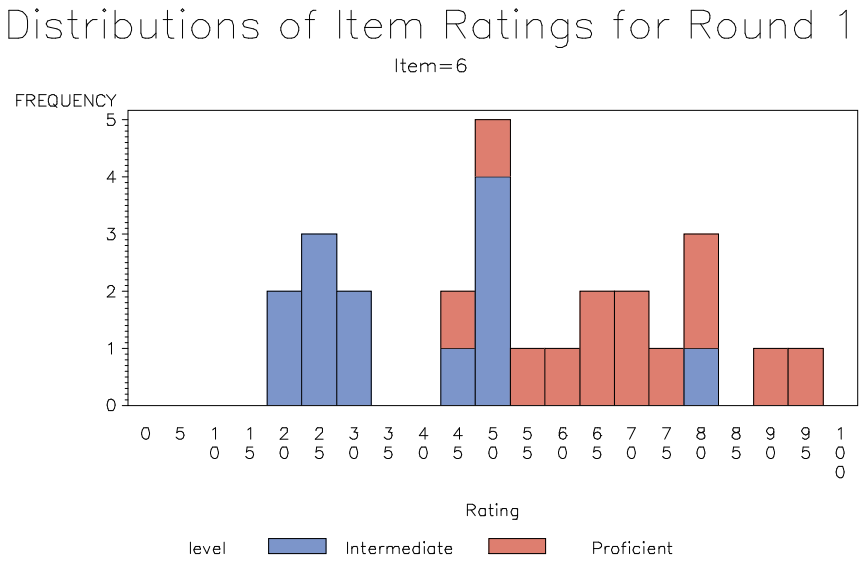


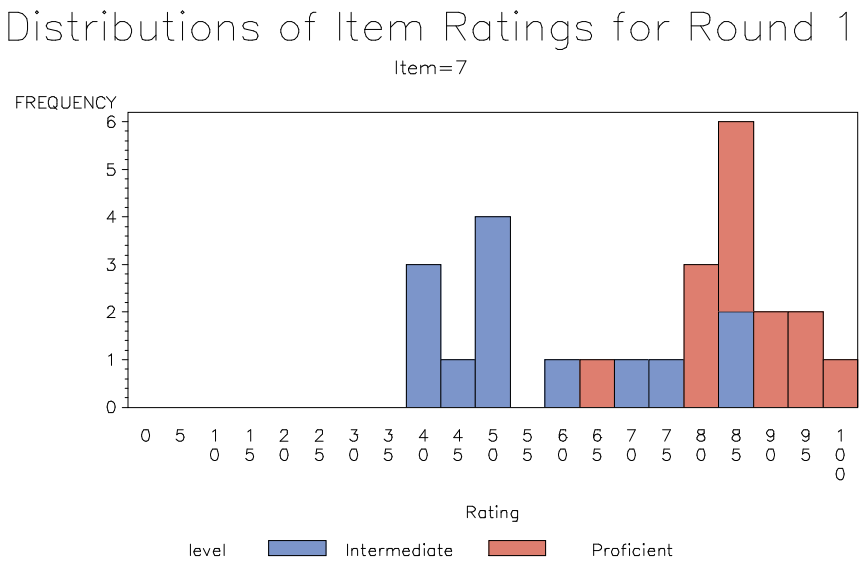



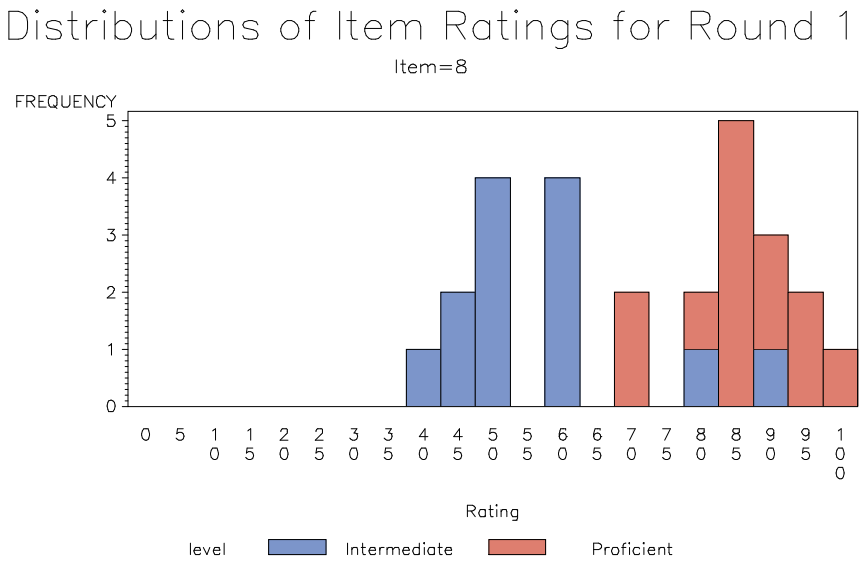

Distributions of Item Ratings for Round 1
Item=3
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

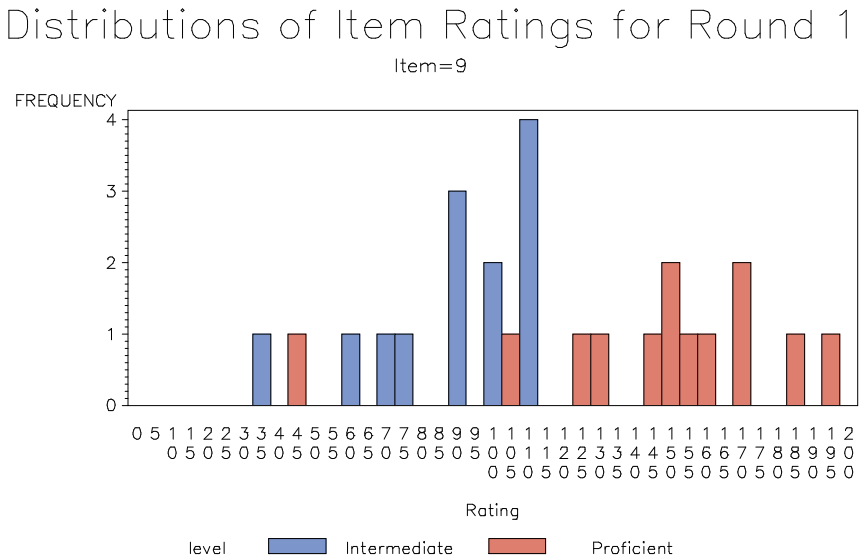

Distributions of Item Ratings for Round 1
Item=4
FREQUENCY
0
1
2
3
4
5
6
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

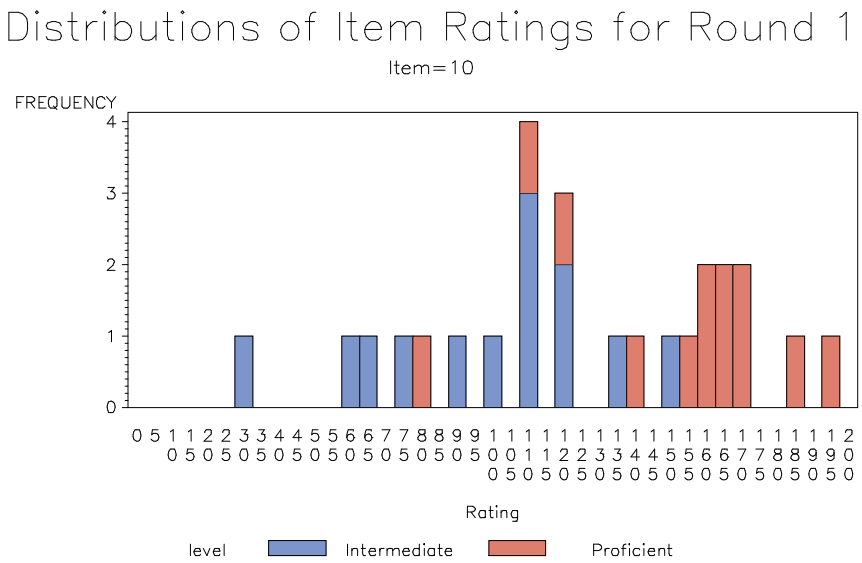

Distributions of Item Ratings for Round 1
Item=5
FREQUENCY
0
1
2
3
4
5
6
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

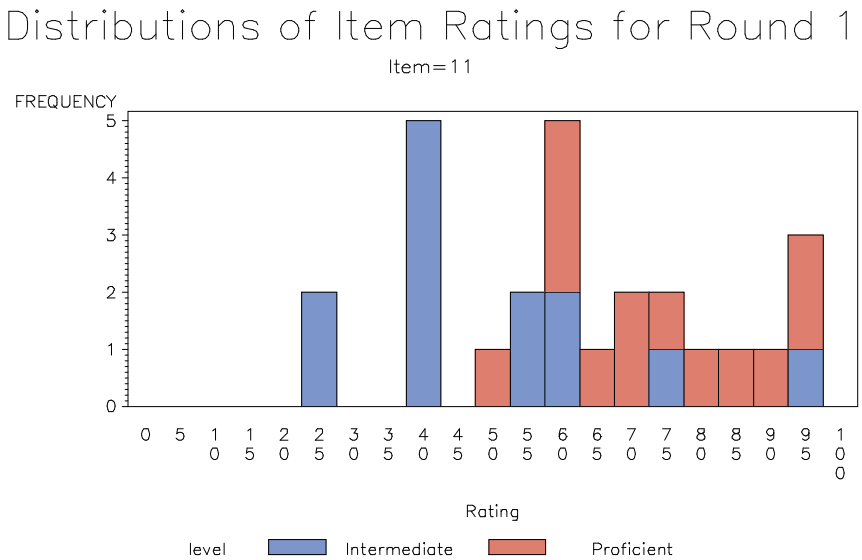

Distributions of Item Ratings for Round 1
Item=6
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

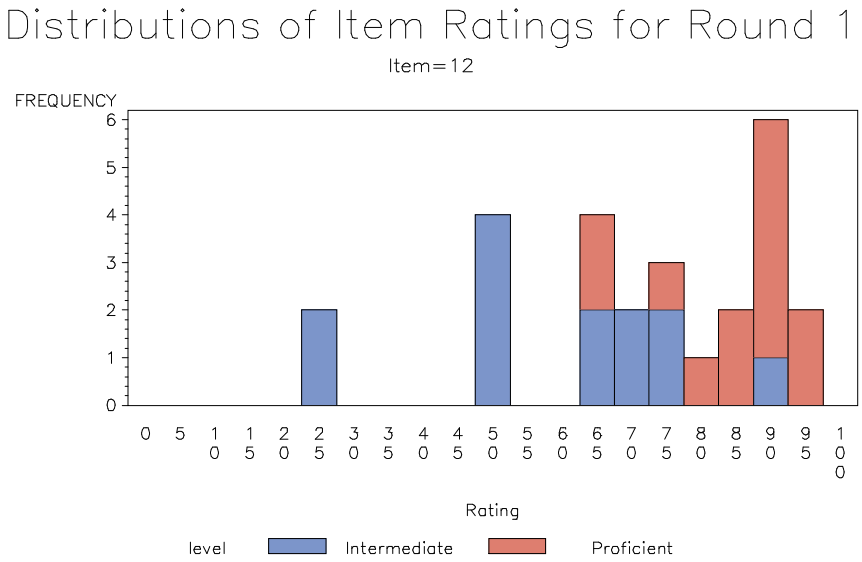

Distributions of Item Ratings for Round 1
Item=7
FREQUENCY
0
1
2
3
4
5
6
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

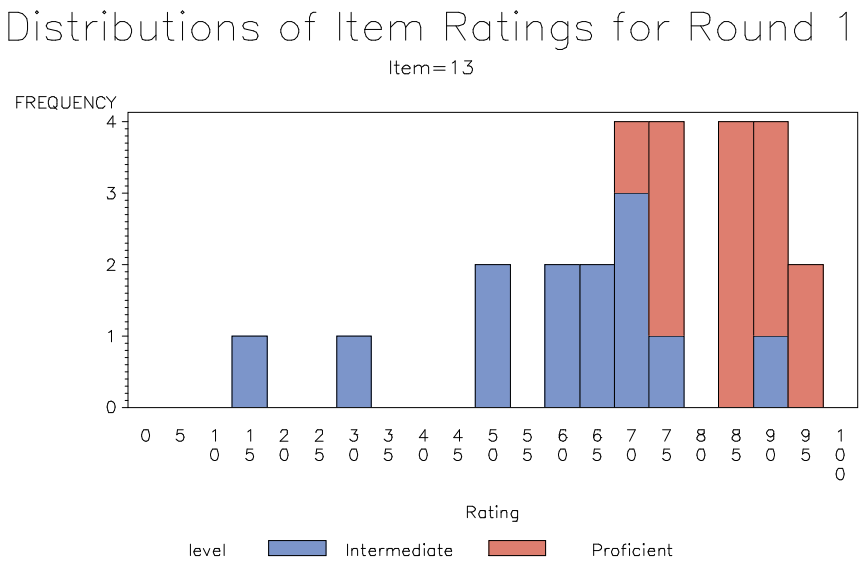

Distributions of Item Ratings for Round 1
Item=8
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

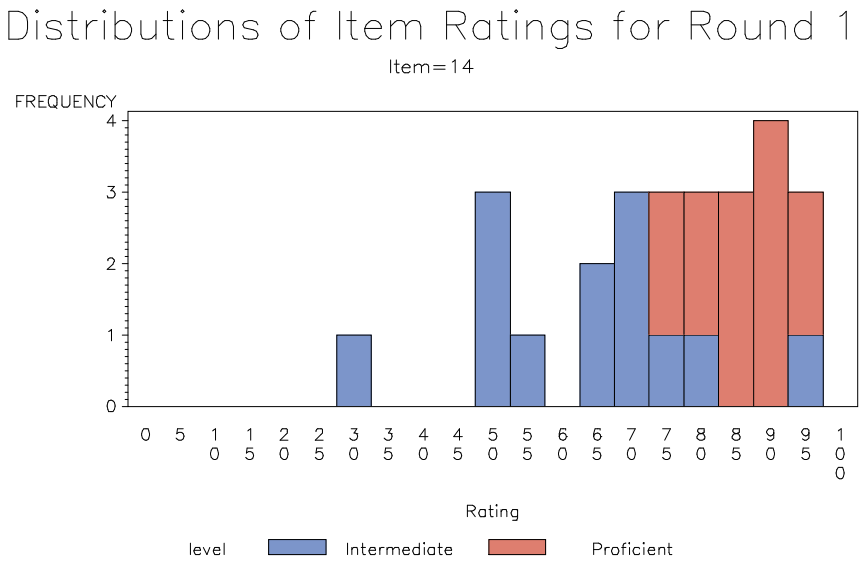

Distributions of Item Ratings for Round 1
Item=9
FREQUENCY
Rating
level
Intermediate
Proficient

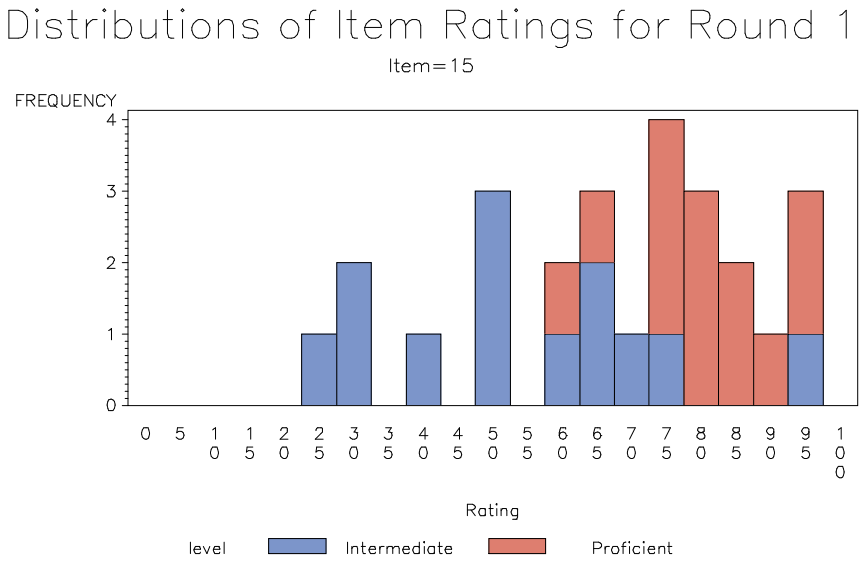

Distributions of Item Ratings for Round 1
Item=10
FREQUENCY
Rating
level
Intermediate
Proficient

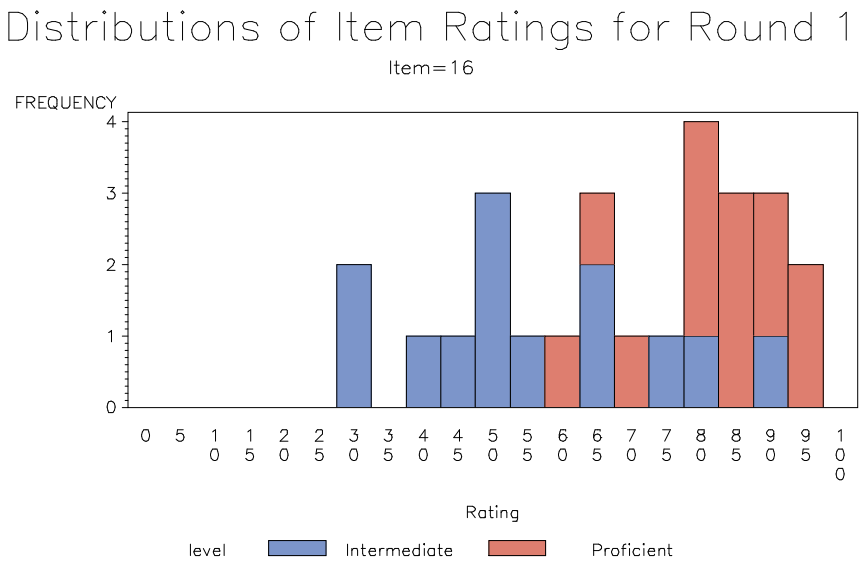

Distributions of Item Ratings for Round 1
Item=11
FREQUENCY
Rating
level
Intermediate
Proficient

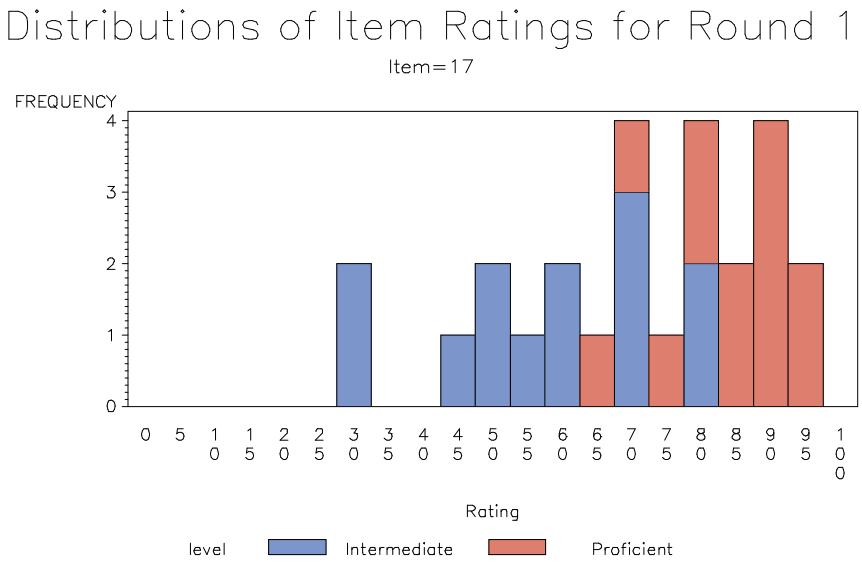

Distributions of Item Ratings for Round 1
Item=12
FREQUENCY
0
1
2
3
4
5
6
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

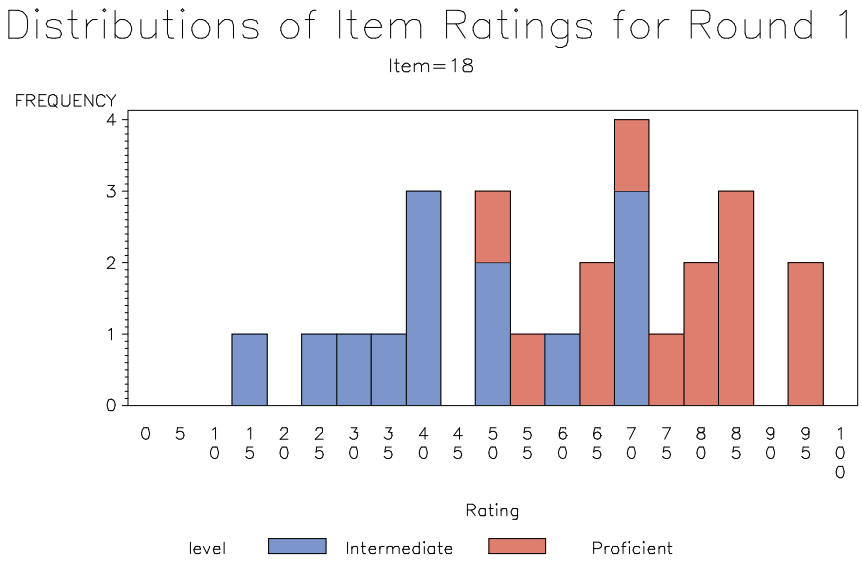

Distributions of Item Ratings for Round 1
Item=13
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

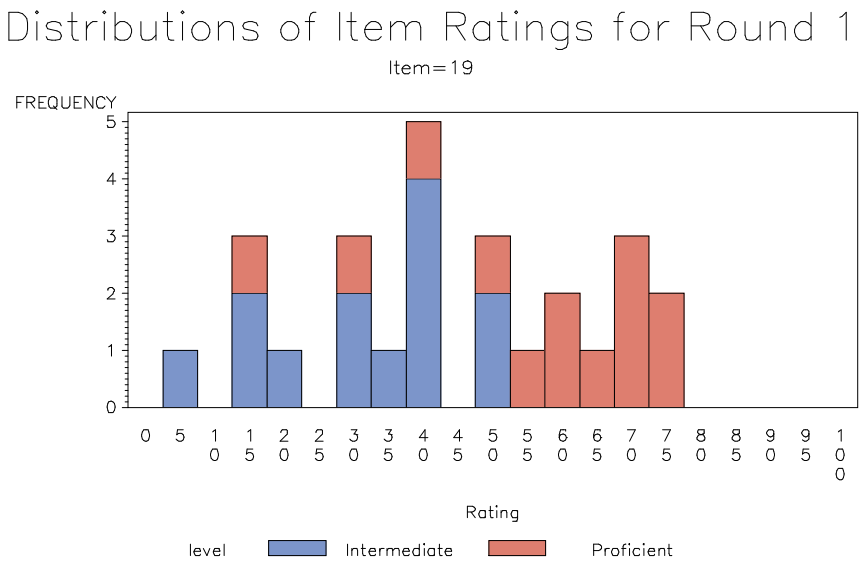

Distributions of Item Ratings for Round 1
Item=14
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

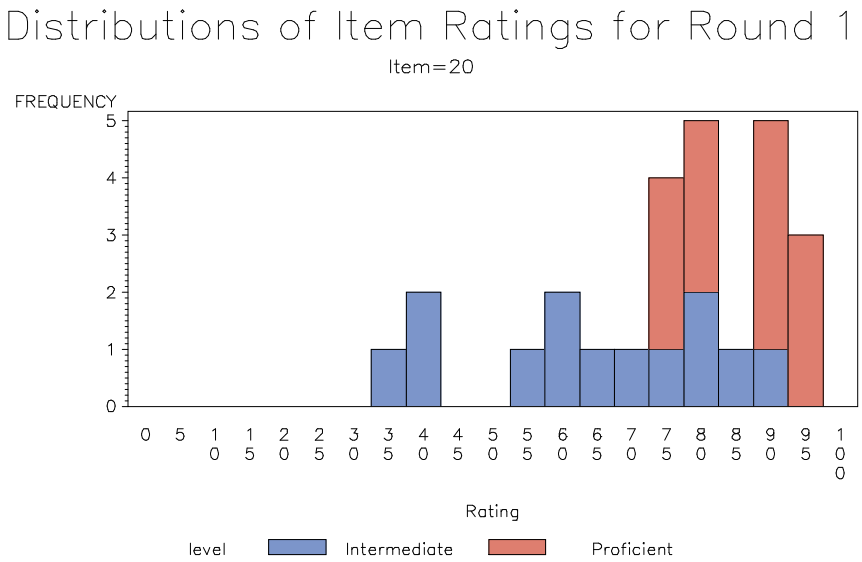

Distributions of Item Ratings for Round 1
Item=15
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

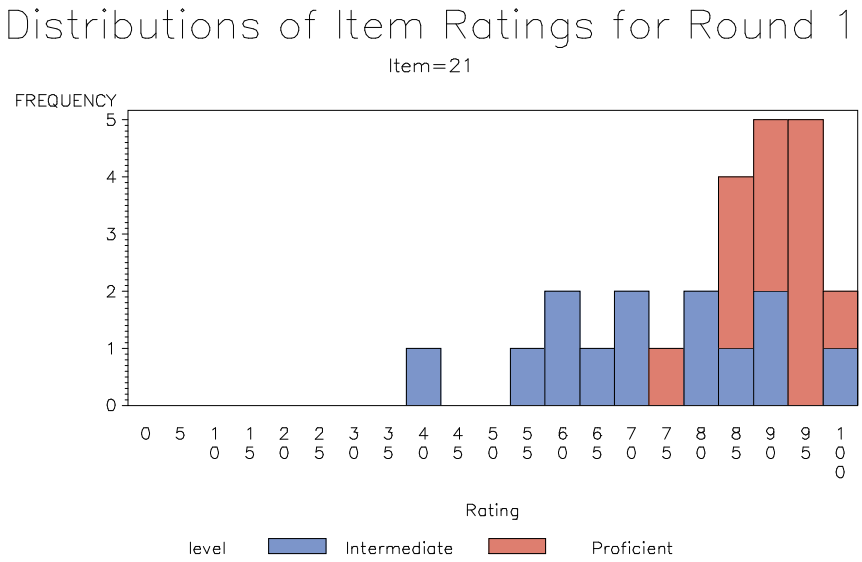

Distributions of Item Ratings for Round 1
Item=16
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

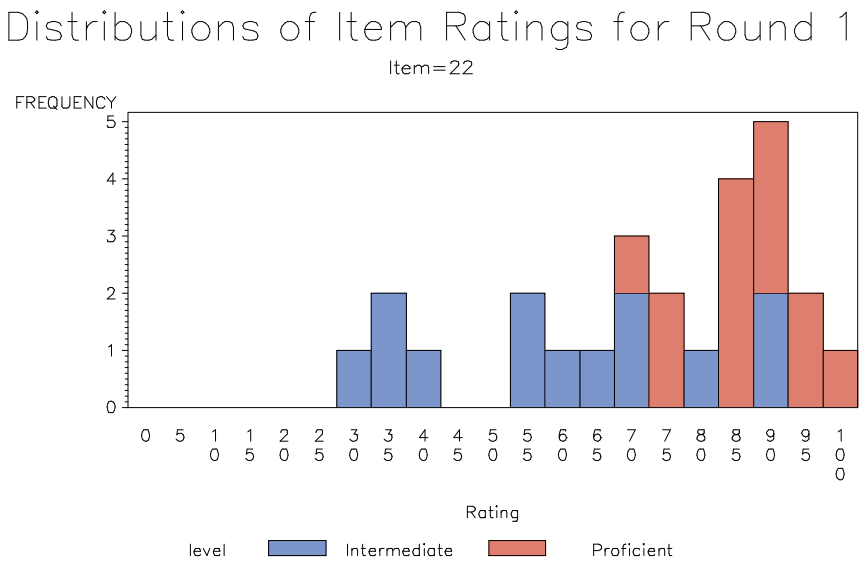

Distributions of Item Ratings for Round 1
Item=17
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

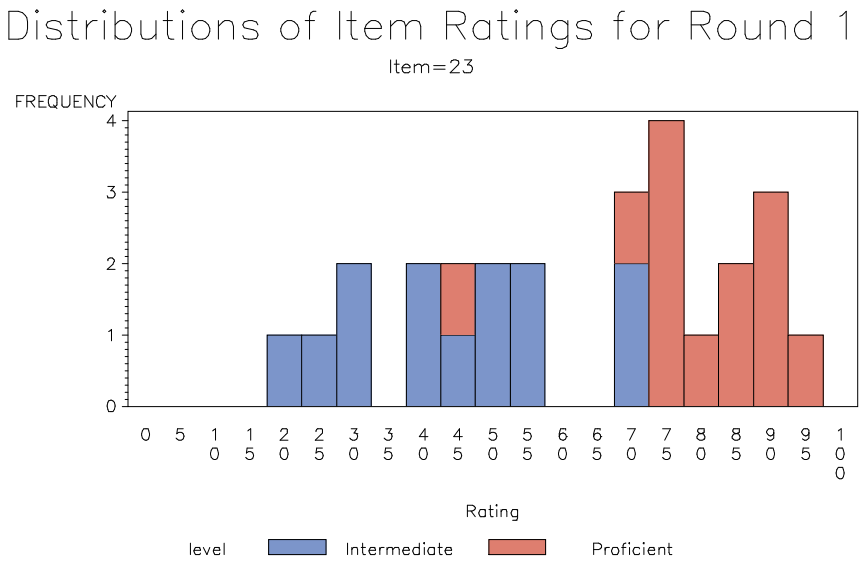

Distributions of Item Ratings for Round 1
Item=18
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

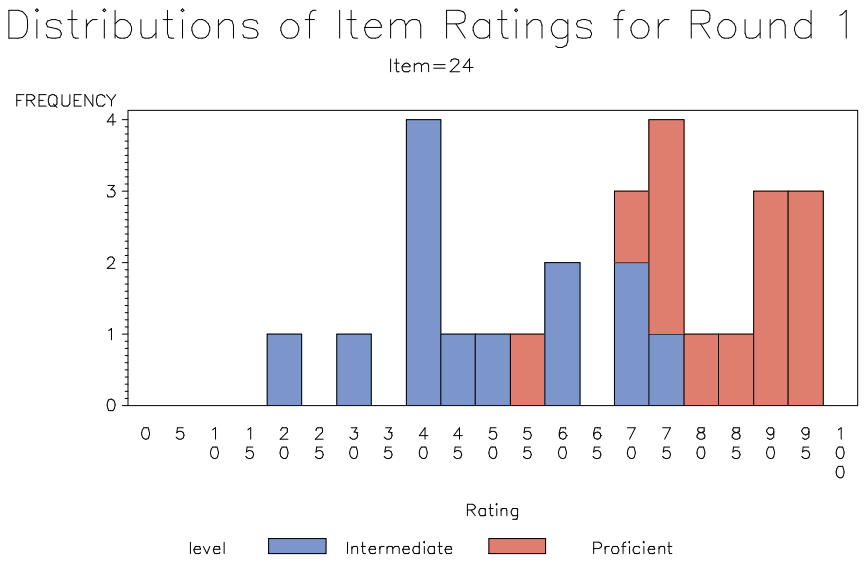

Distributions of Item Ratings for Round 1
Item=19
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

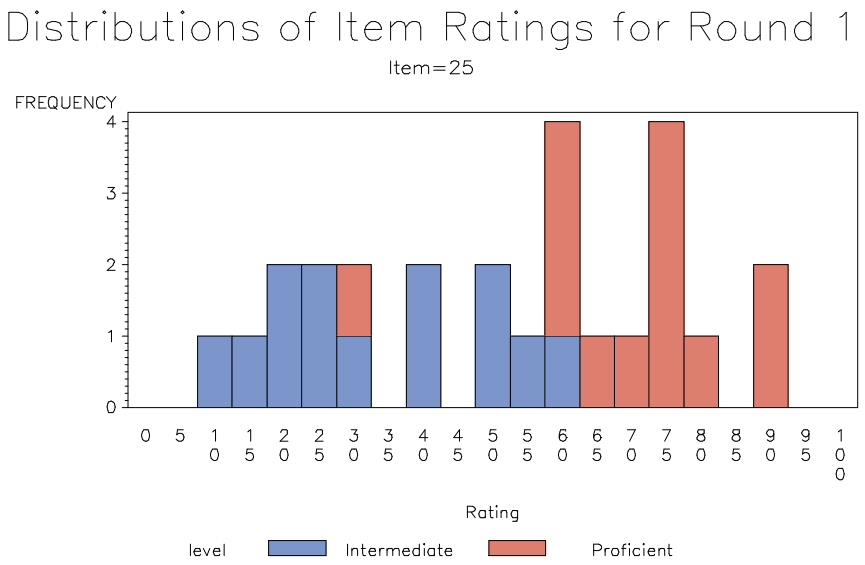

Distributions of Item Ratings for Round 1
Item=20
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

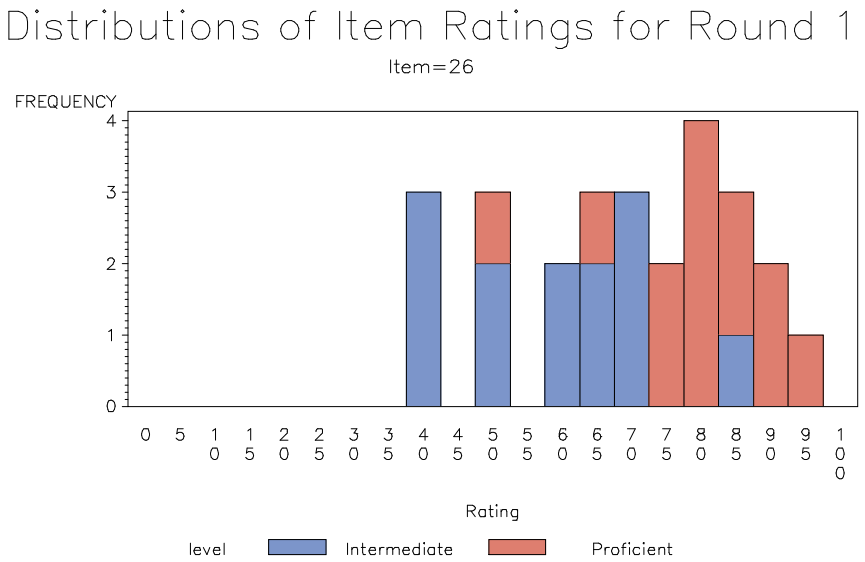

Distributions of Item Ratings for Round 1
Item=21
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

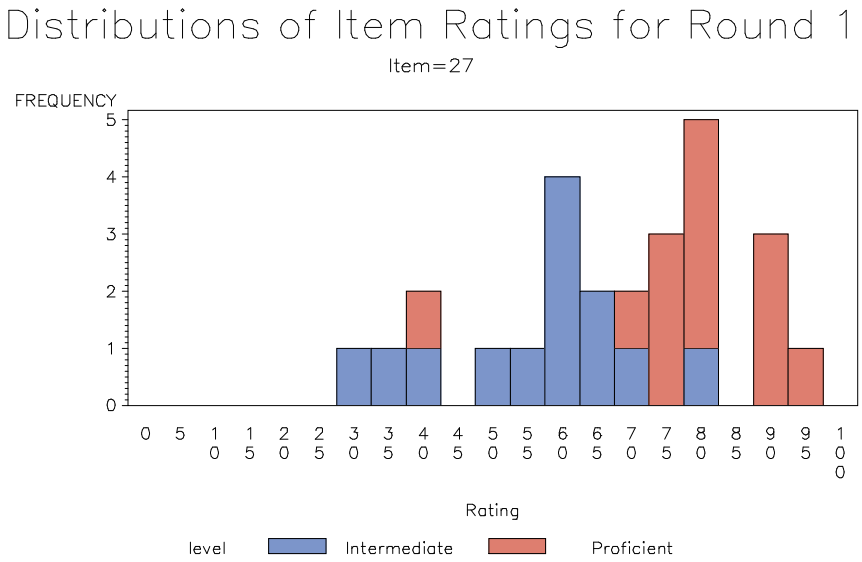

Distributions of Item Ratings for Round 1
Item=22
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

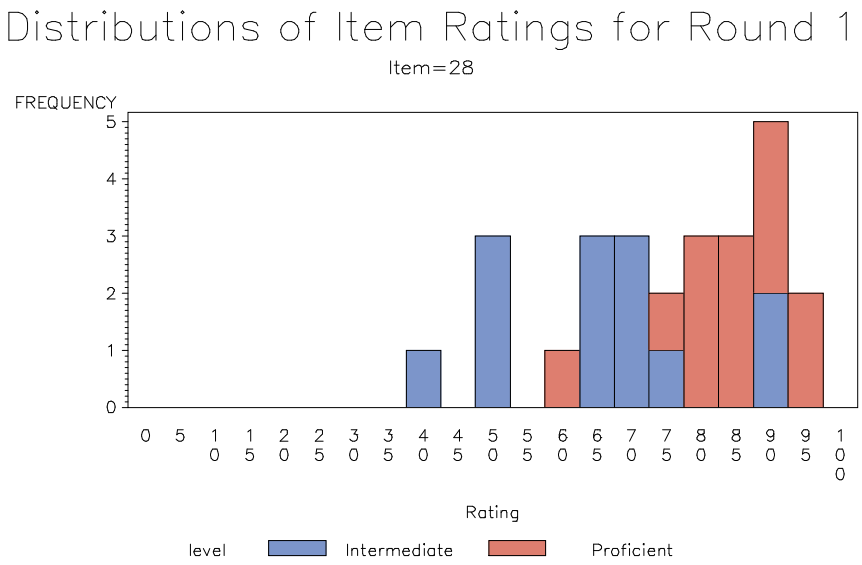

Distributions of Item Ratings for Round 1
Item=23
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

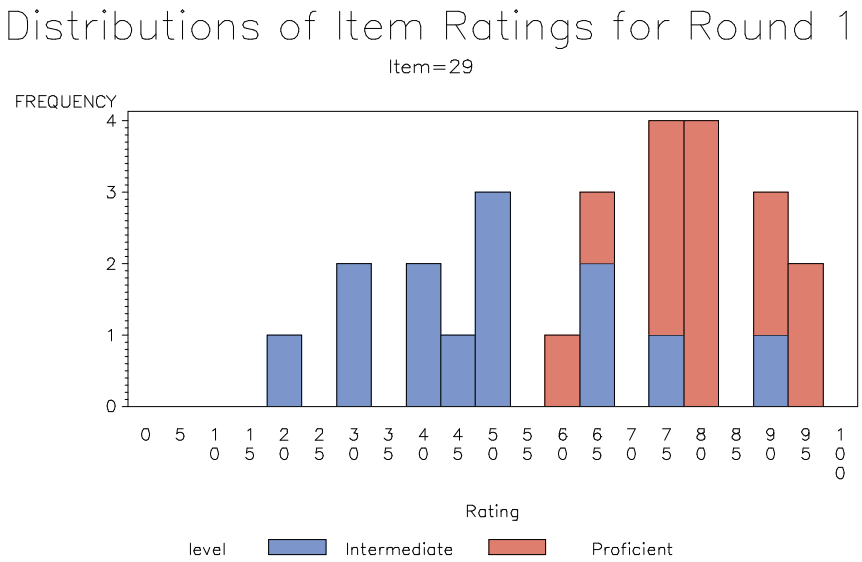

Distributions of Item Ratings for Round 1
Item=24
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

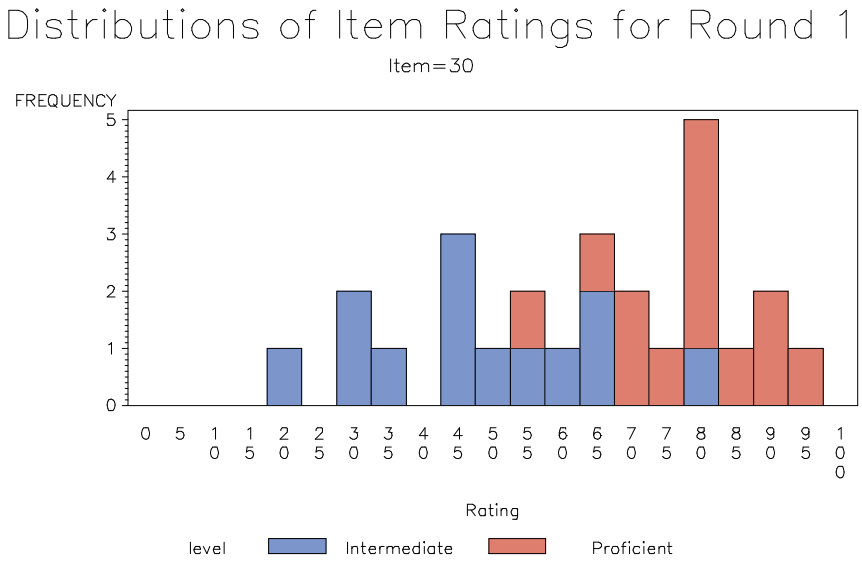

Distributions of Item Ratings for Round 1
Item=25
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

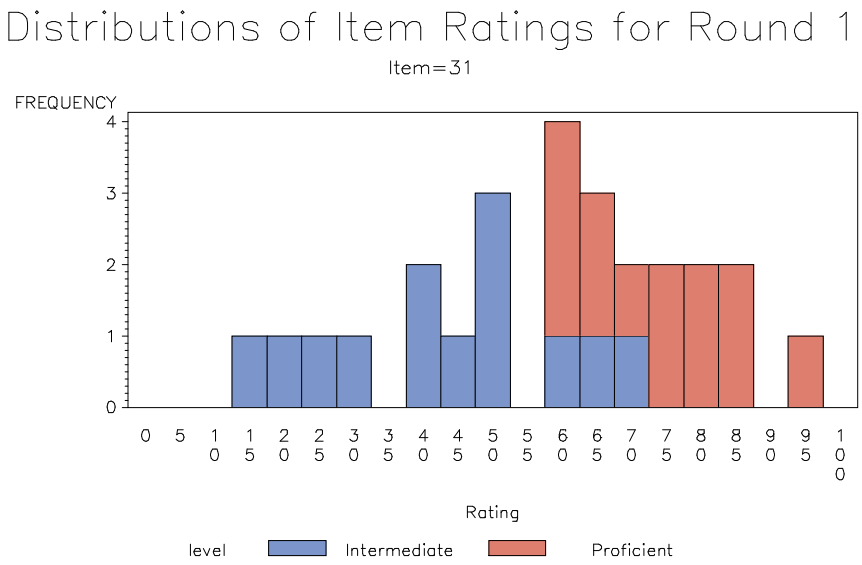

Distributions of Item Ratings for Round 1
Item=26
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

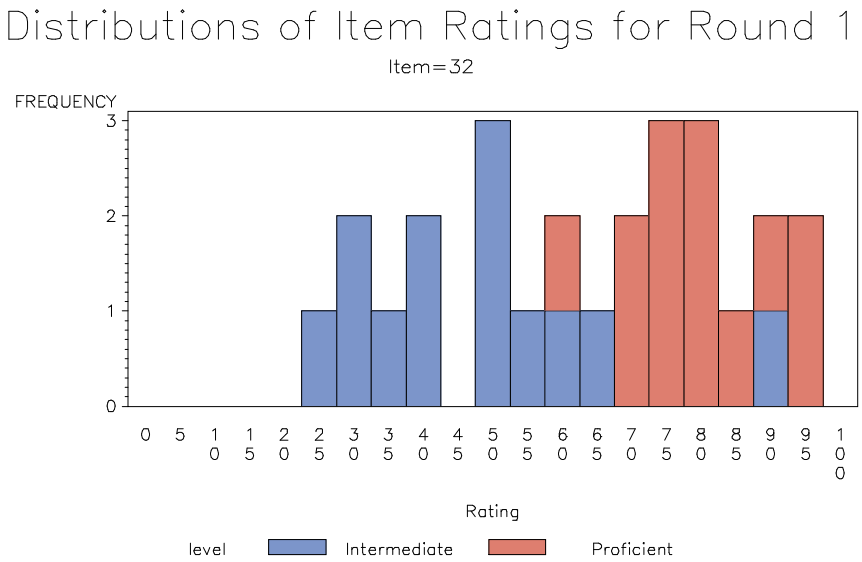

Distributions of Item Ratings for Round 1
Item=27
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

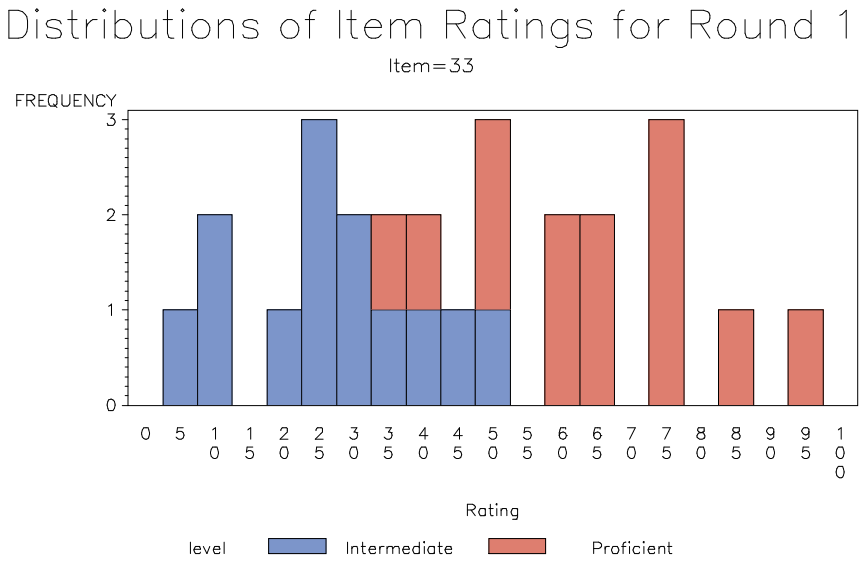

Distributions of Item Ratings for Round 1
Item=28
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

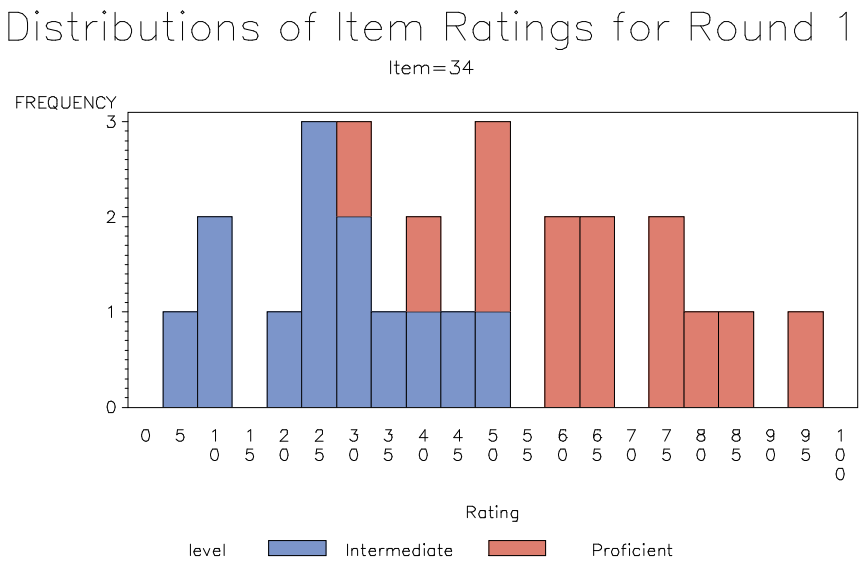

Distributions of Item Ratings for Round 1
Item=29
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

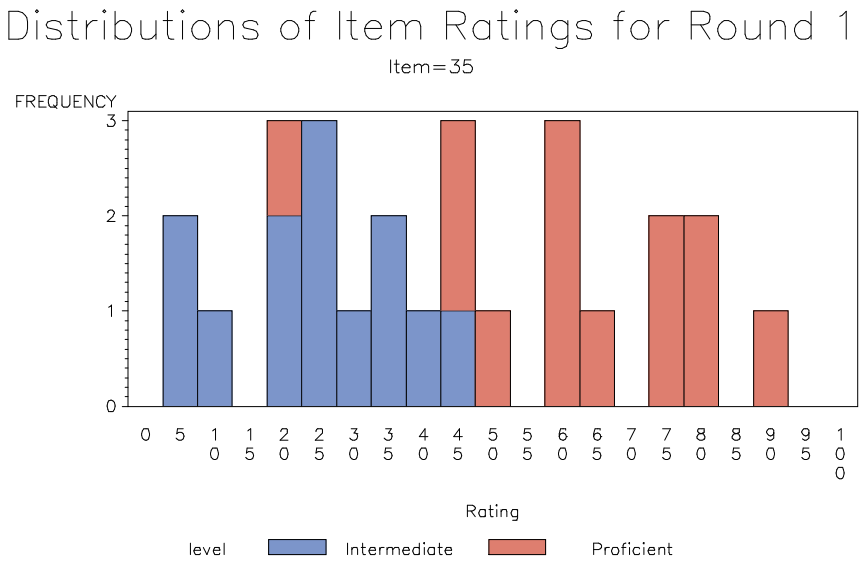

Distributions of Item Ratings for Round 1
Item=30
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

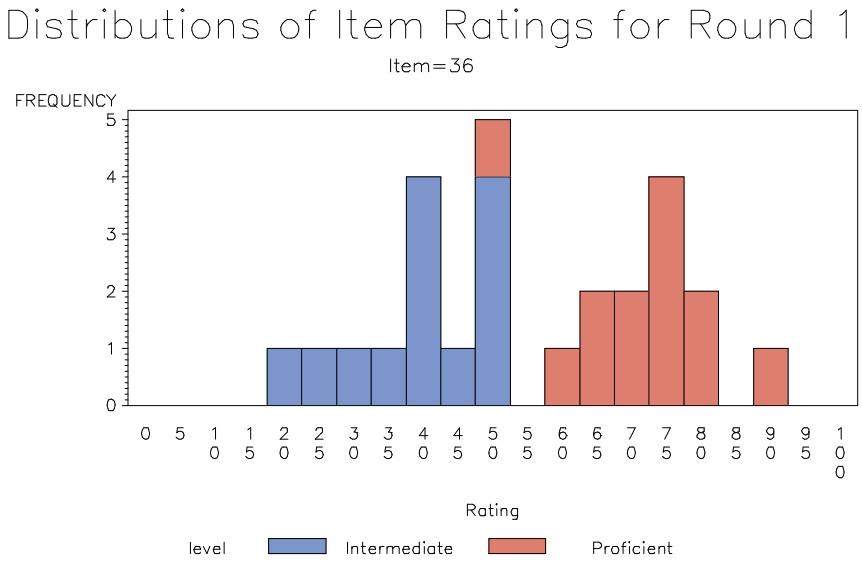

Distributions of Item Ratings for Round 1
Item=31
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

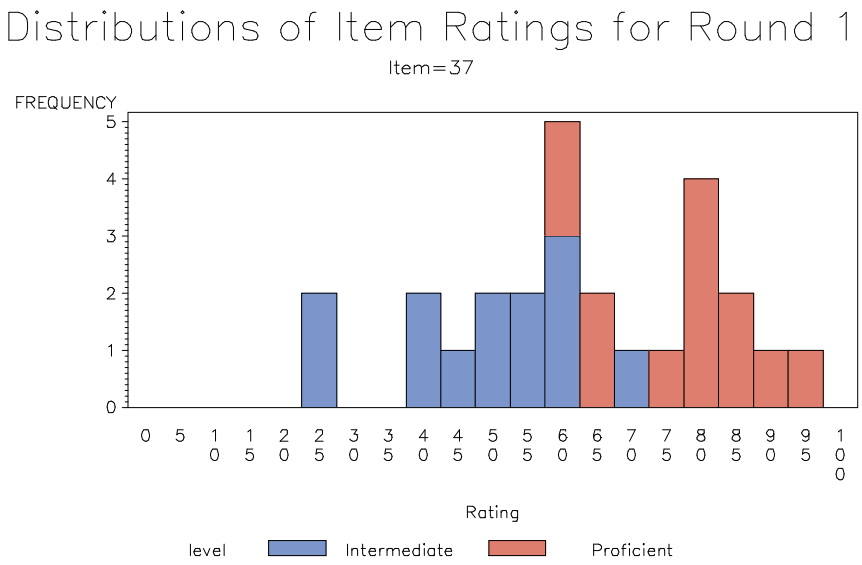

Distributions of Item Ratings for Round 1
Item=32
FREQUENCY
0
1
2
3
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

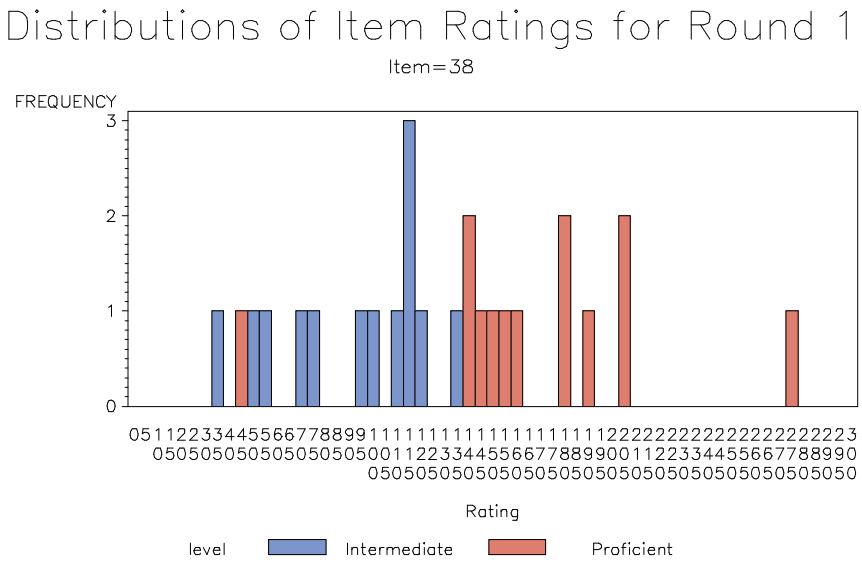

Distributions of Item Ratings for Round 1
Item=33
FREQUENCY
0
1
2
3
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient


Distributions of Item Ratings for Round 1
Item=34
FREQUENCY
0
1
2
3
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

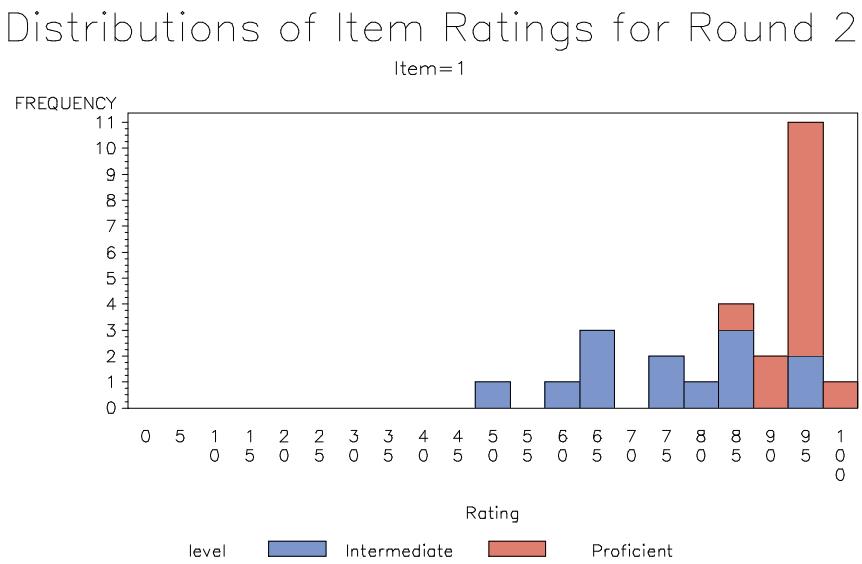

Distributions of Item Ratings for Round 1
Item=35
FREQUENCY
0
1
2
3
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

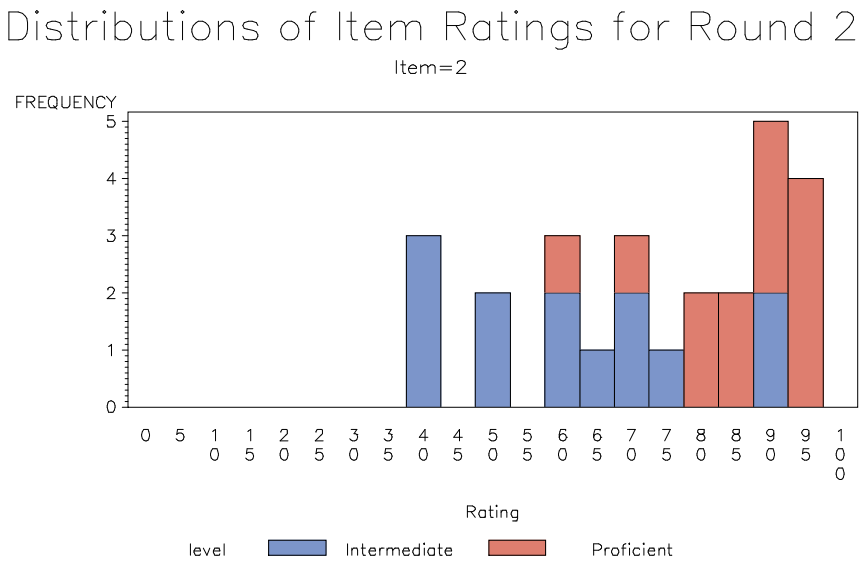

Distributions of Item Ratings for Round 1
Item=36
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

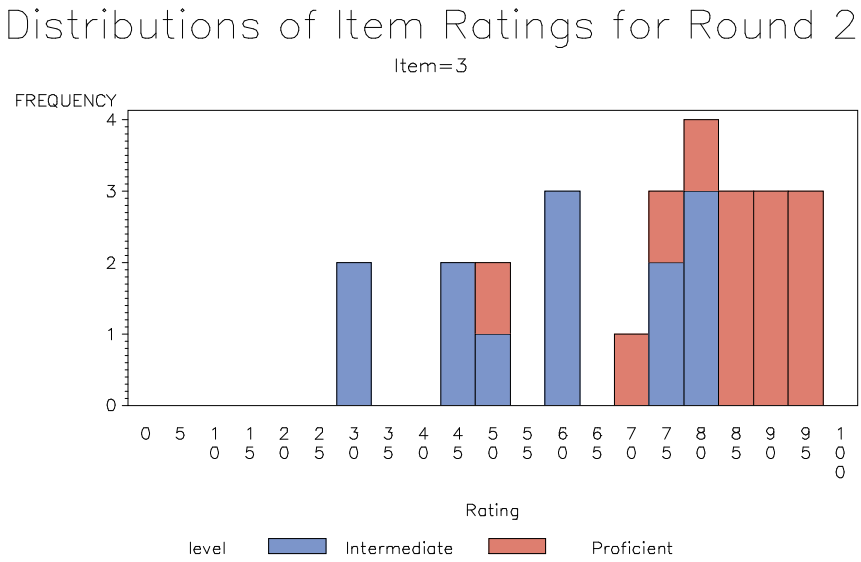

Distributions of Item Ratings for Round 1
Item=37
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

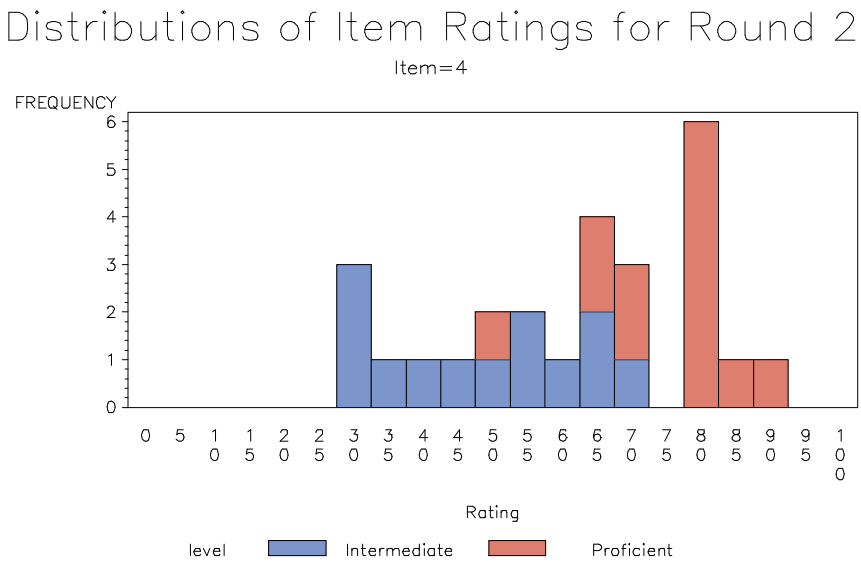

Distributions of Item Ratings for Round 1
Item=38
FREQUENCY
0
1
2
3
Rating
level
Intermediate
Proficient

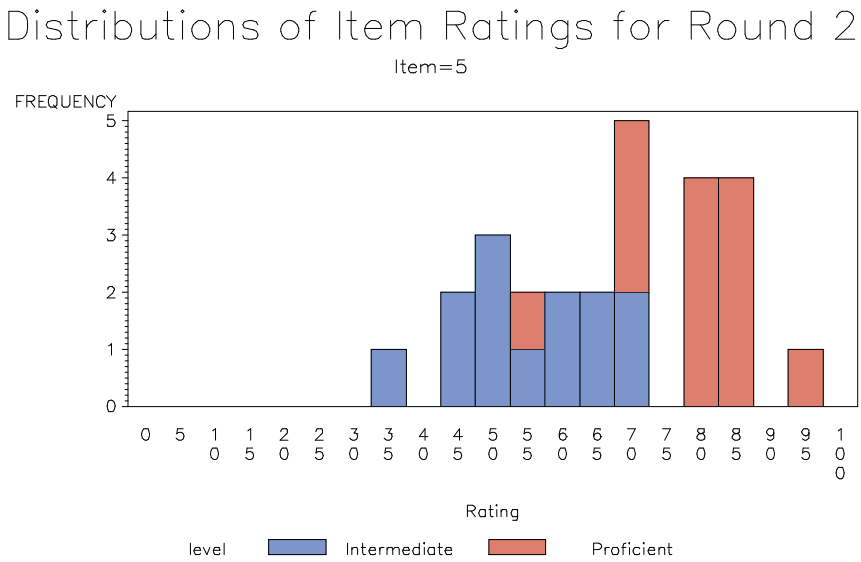
## Appendix L.2: Rating Distribution at Round 2

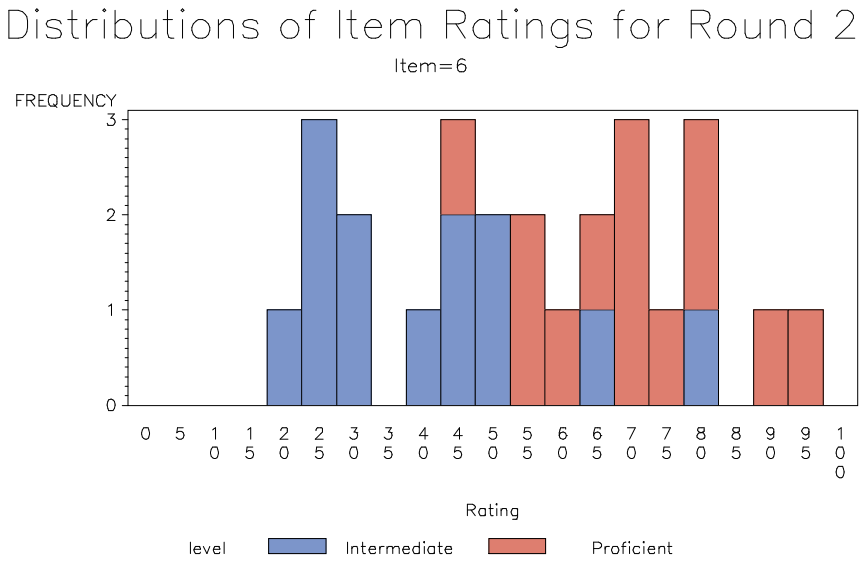


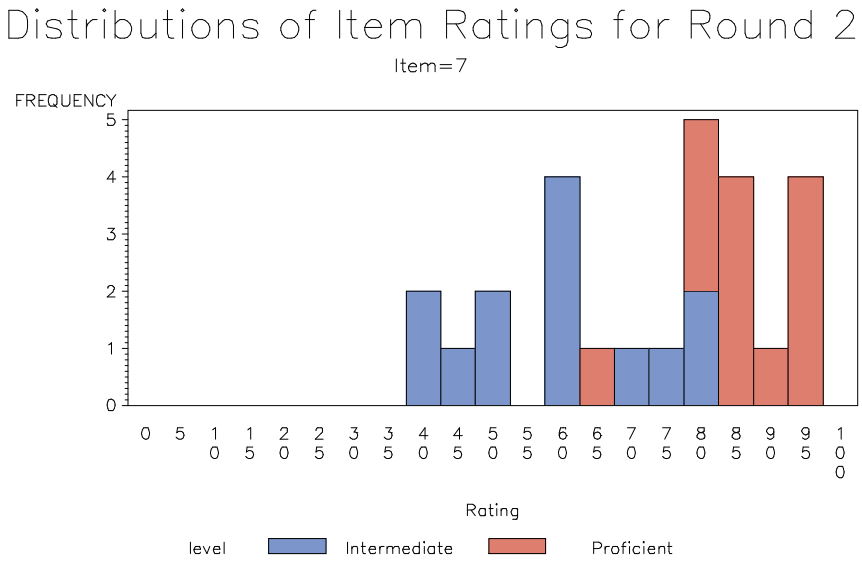



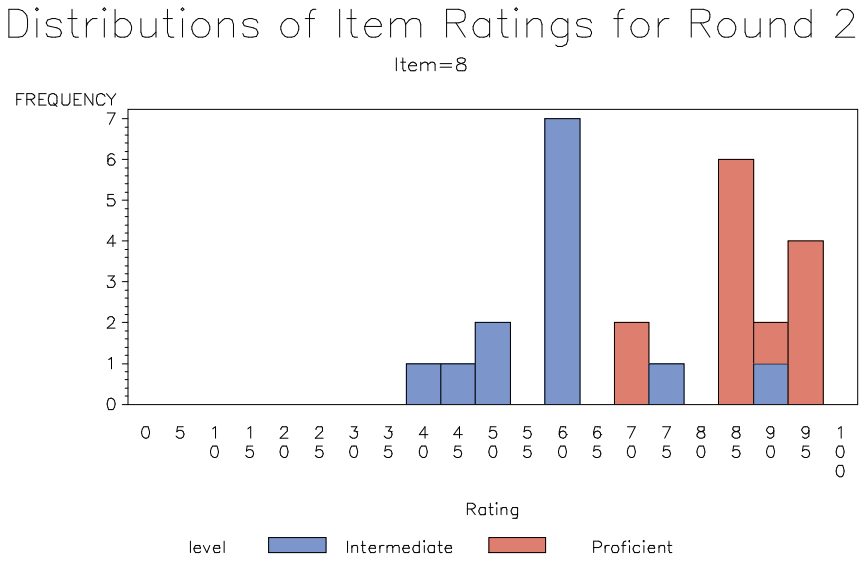

Distributions of Item Ratings for Round 2
Item=3
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

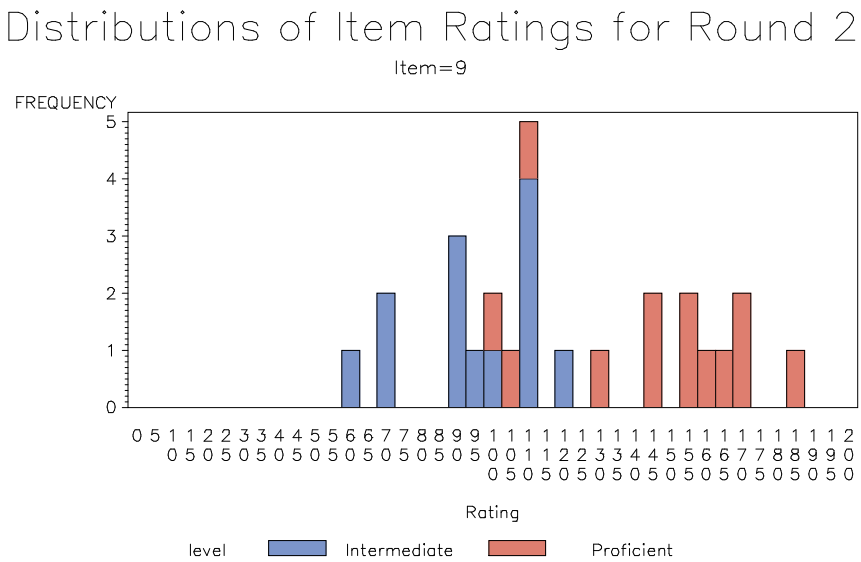

Distributions of Item Ratings for Round 2
Item=4
FREQUENCY
0
1
2
3
4
5
6
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

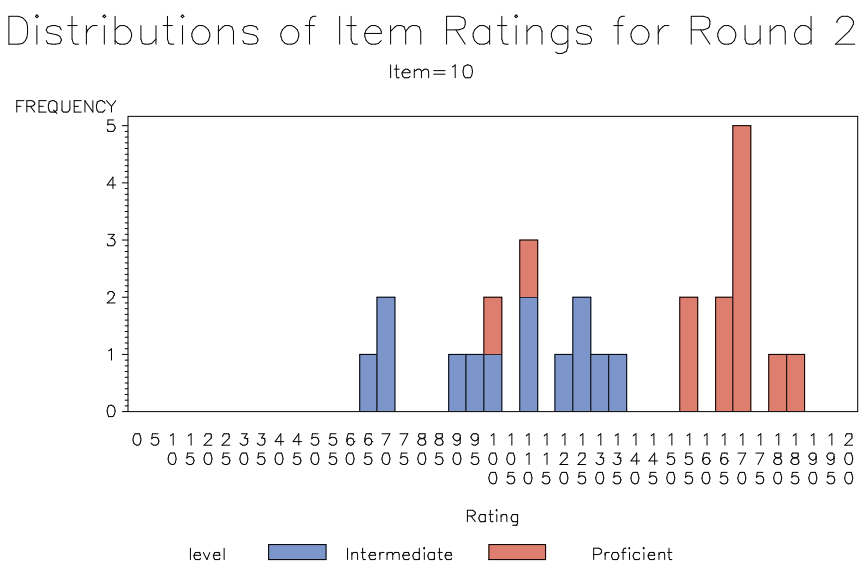

Distributions of Item Ratings for Round 2
Item=5
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

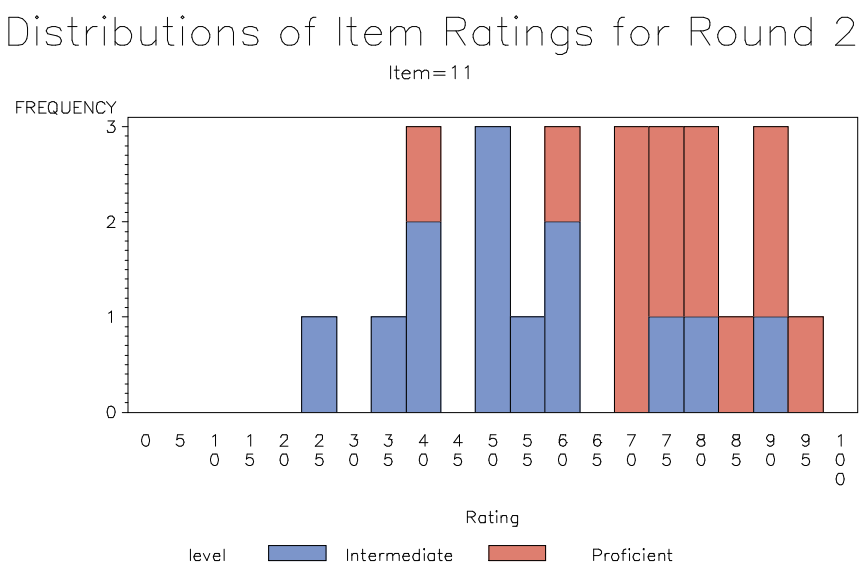

Distributions of Item Ratings for Round 2
Item=6
FREQUENCY
0
1
2
3
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

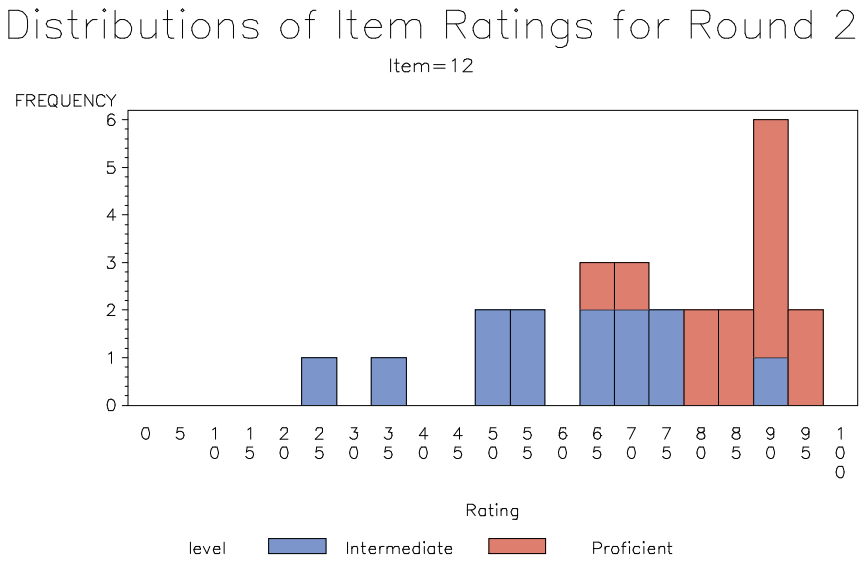

Distributions of Item Ratings for Round 2
Item=7
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

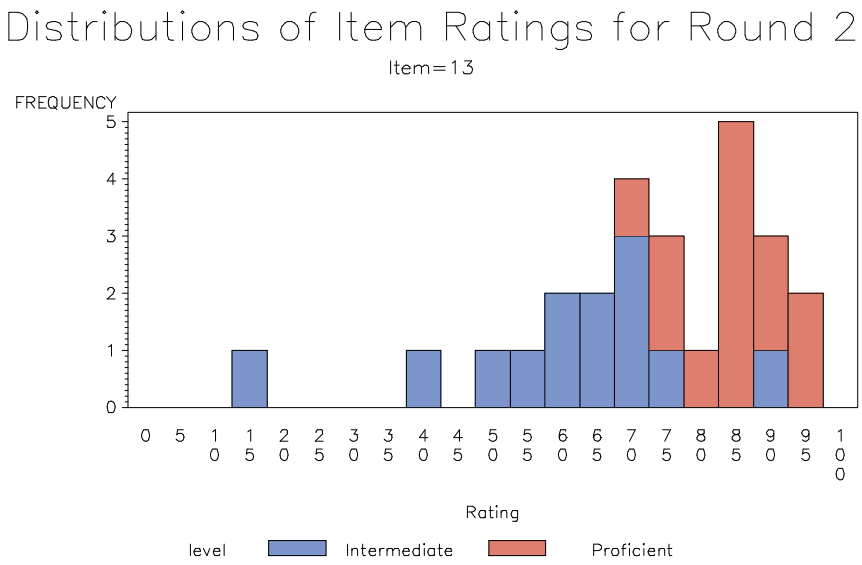

Distributions of Item Ratings for Round 2
Item=8
FREQUENCY
0
1
2
3
4
5
6
7
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

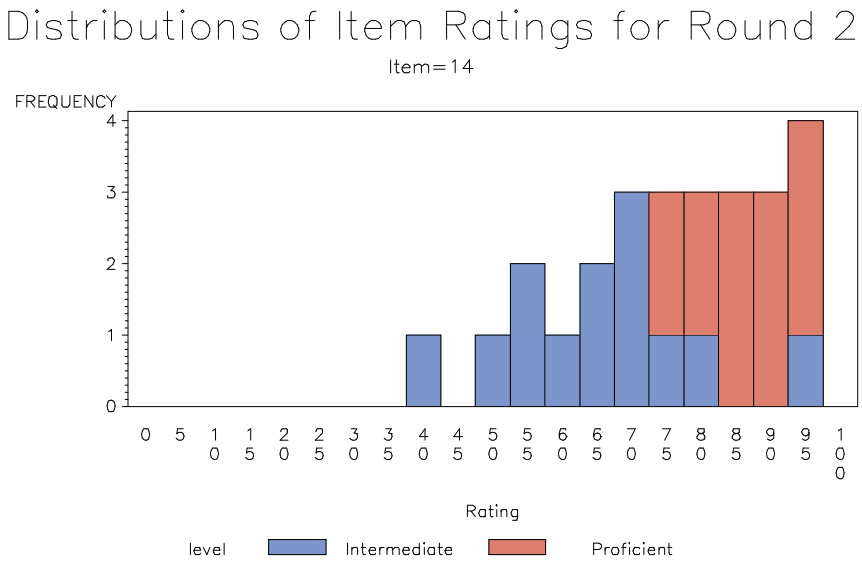

Distributions of Item Ratings for Round 2
Item=9
FREQUENCY
Rating
level
Intermediate
Proficient

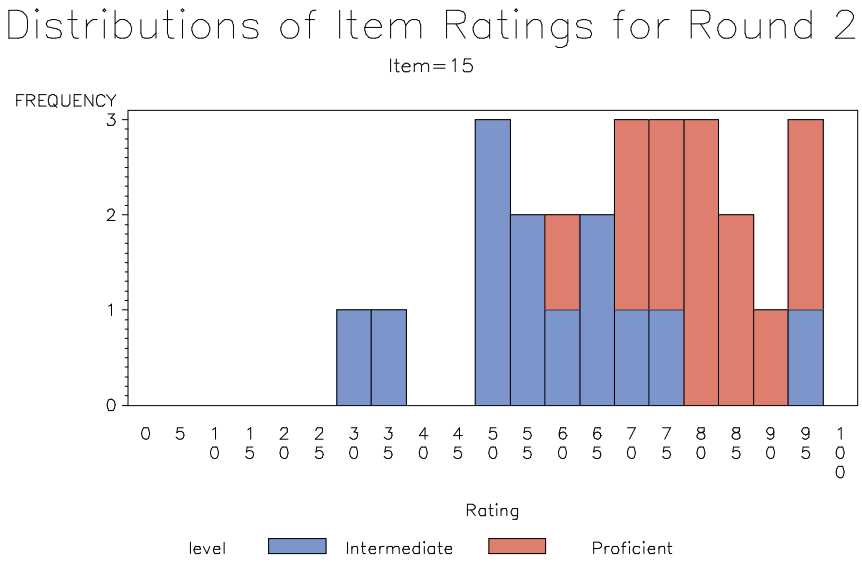

Distributions of Item Ratings for Round 2
Item=10
FREQUENCY
Rating
level
Intermediate
Proficient

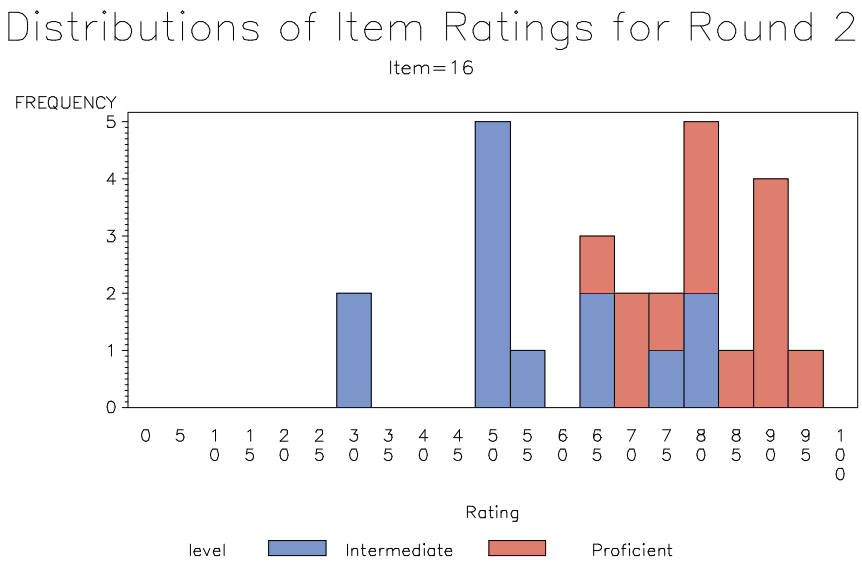

Distributions of Item Ratings for Round 2
Item=11
FREQUENCY
Rating
level
Intermediate
Proficient

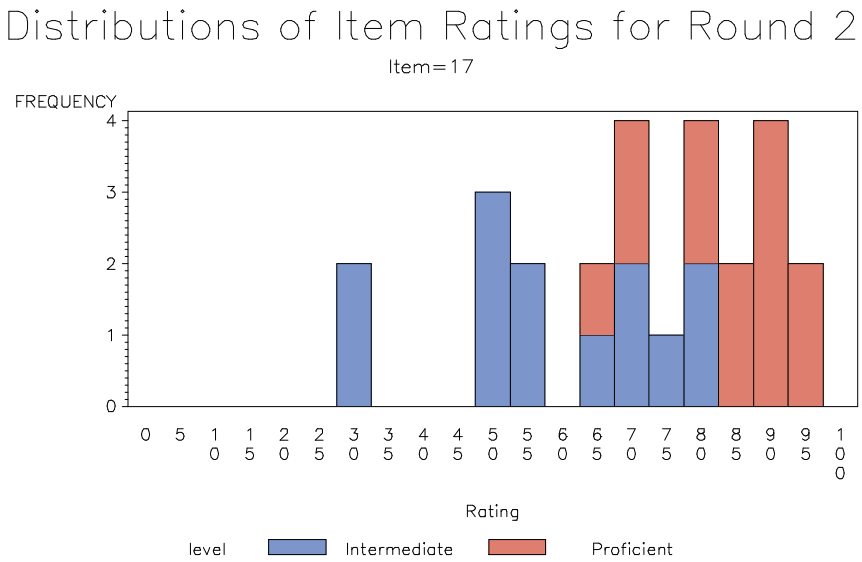

Distributions of Item Ratings for Round 2
Item=12
FREQUENCY
0
1
2
3
4
5
6
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

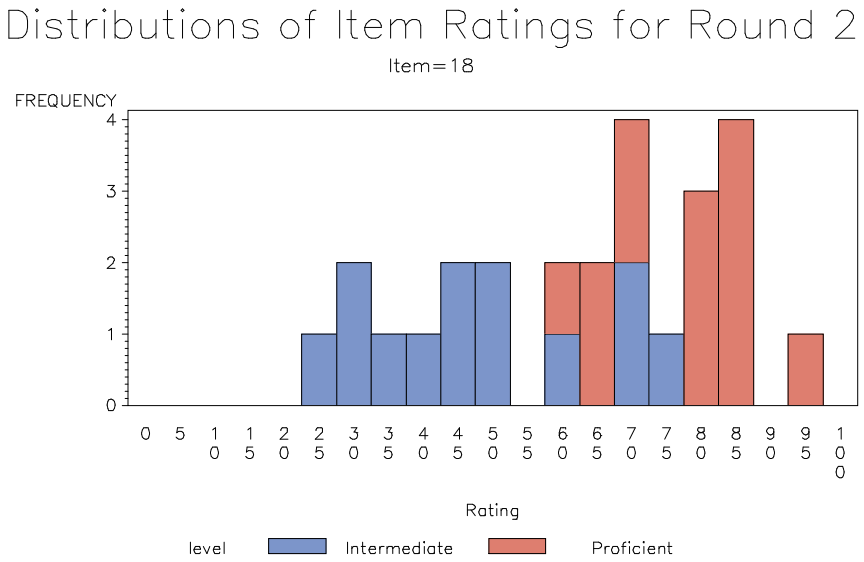

Distributions of Item Ratings for Round 2
Item=13
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

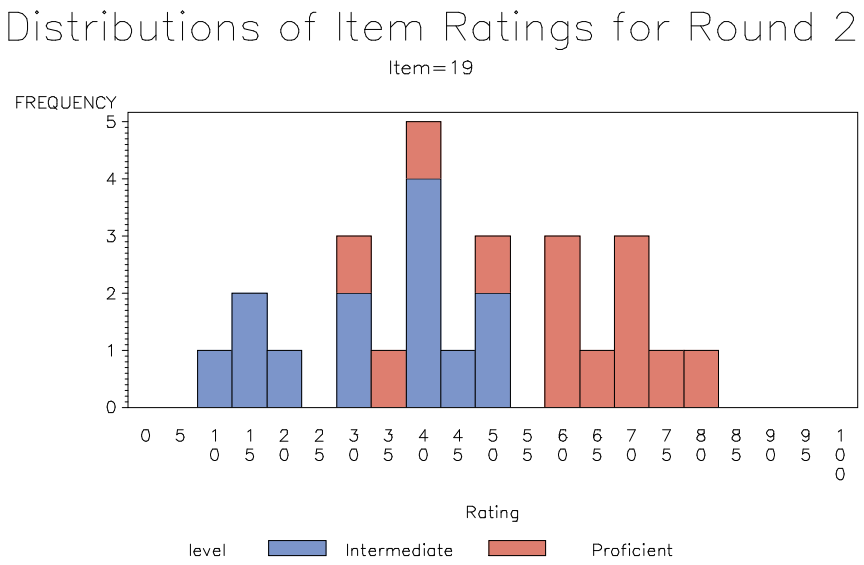

Distributions of Item Ratings for Round 2
Item=14
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

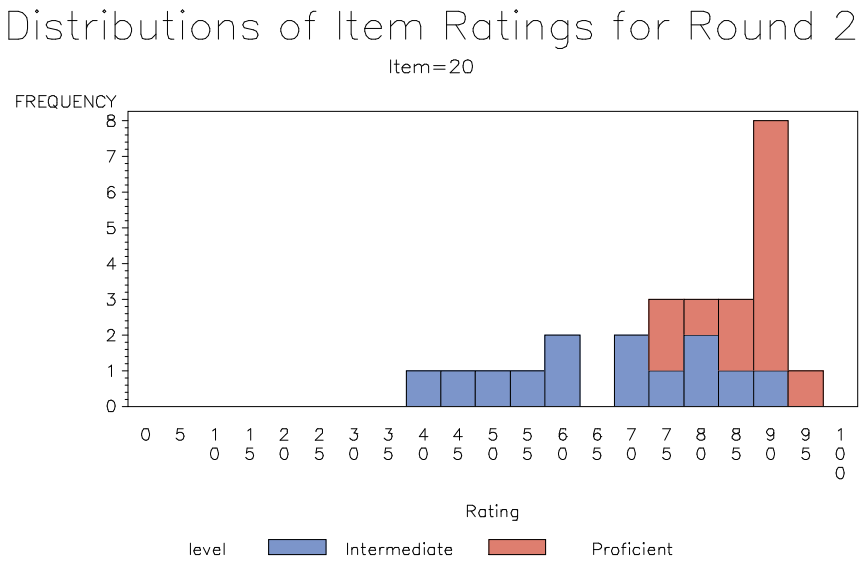

Distributions of Item Ratings for Round 2
Item=15
FREQUENCY
0
1
2
3
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

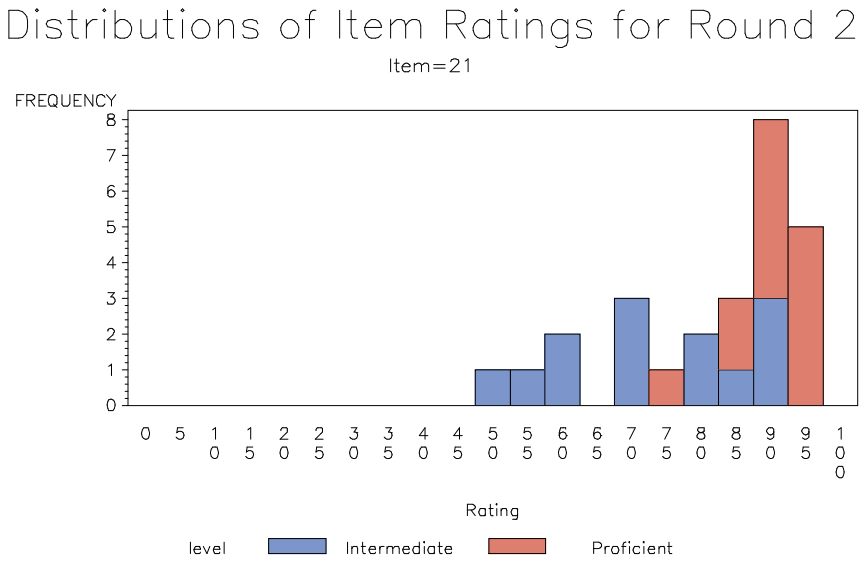

Distributions of Item Ratings for Round 2
Item=16
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

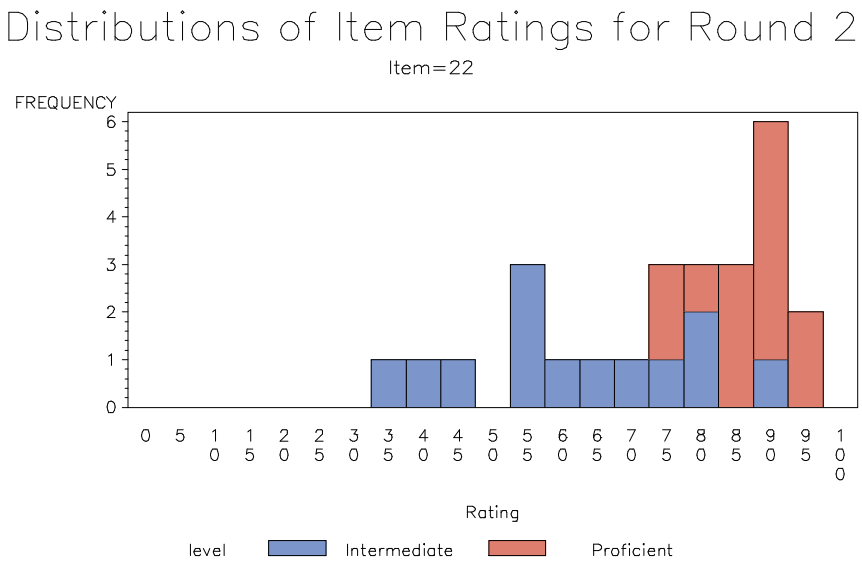

Distributions of Item Ratings for Round 2
Item=17
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

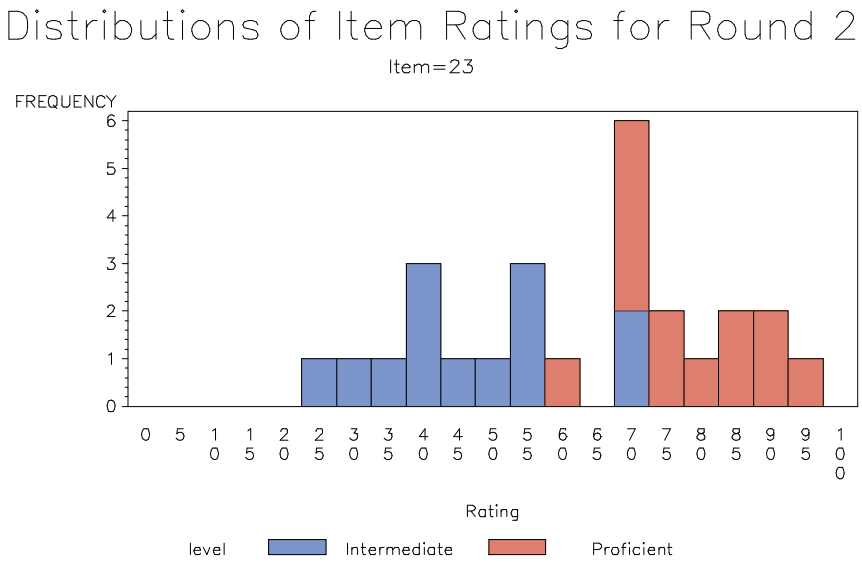

Distributions of Item Ratings for Round 2
Item=18
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

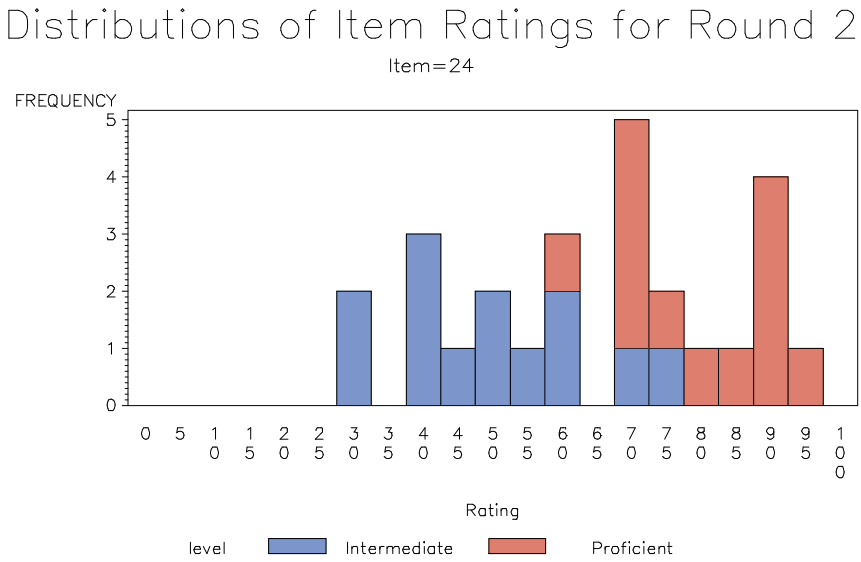

Distributions of Item Ratings for Round 2
Item=19
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

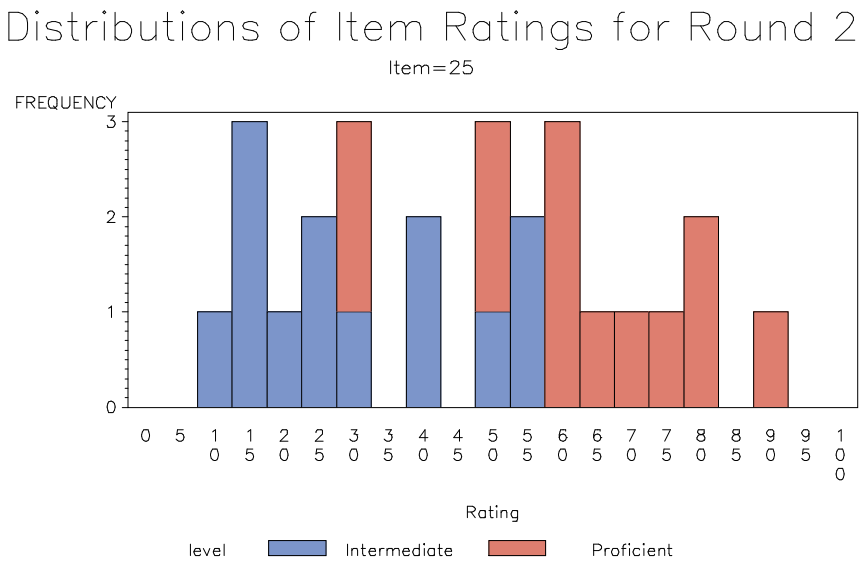

Distributions of Item Ratings for Round 2
Item=20
FREQUENCY
0
1
2
3
4
5
6
7
8
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

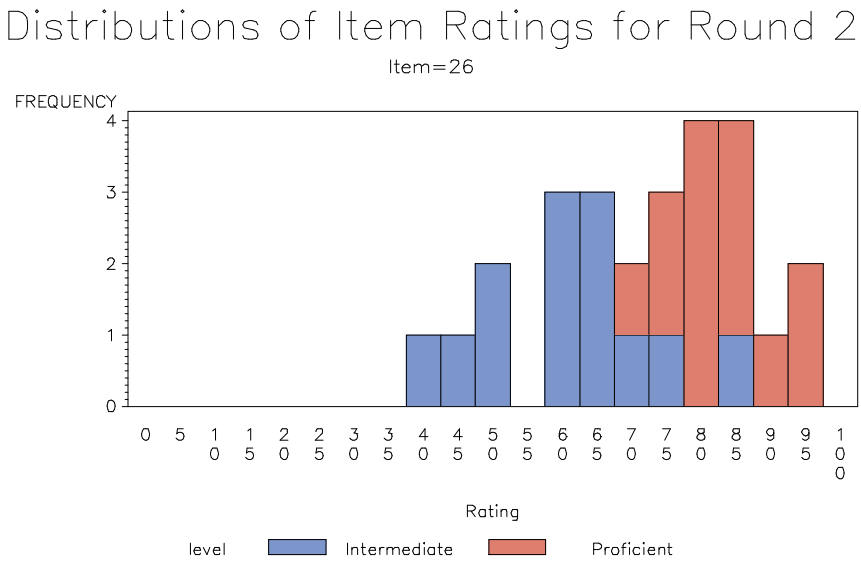

Distributions of Item Ratings for Round 2
Item=21
FREQUENCY
0 1 2 3 4 5 6 7 8
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

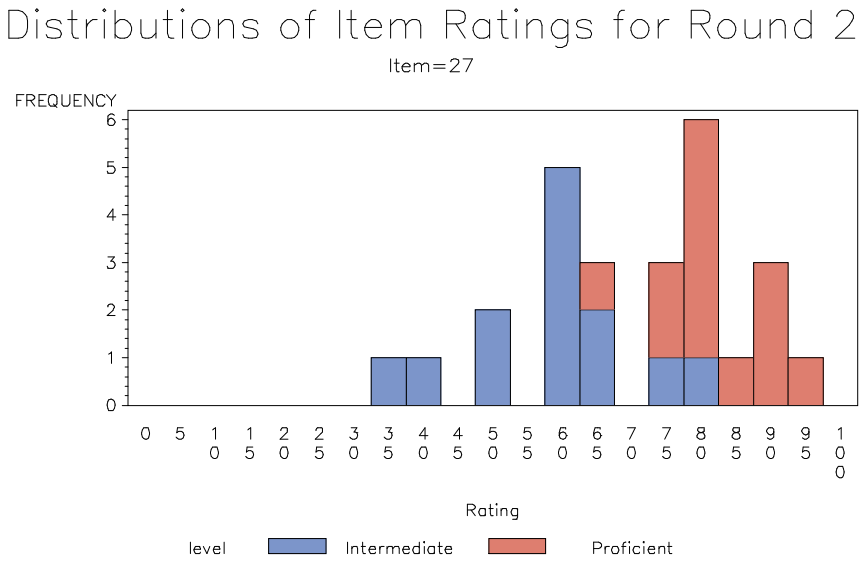

Distributions of Item Ratings for Round 2
Item=22
FREQUENCY
0 1 2 3 4 5 6
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

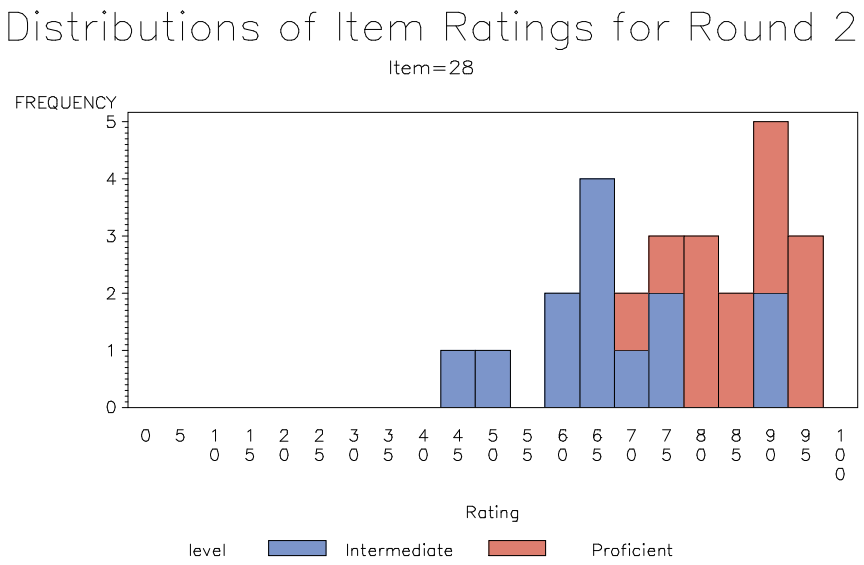

Distributions of Item Ratings for Round 2
Item=23
FREQUENCY
0 1 2 3 4 5 6
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

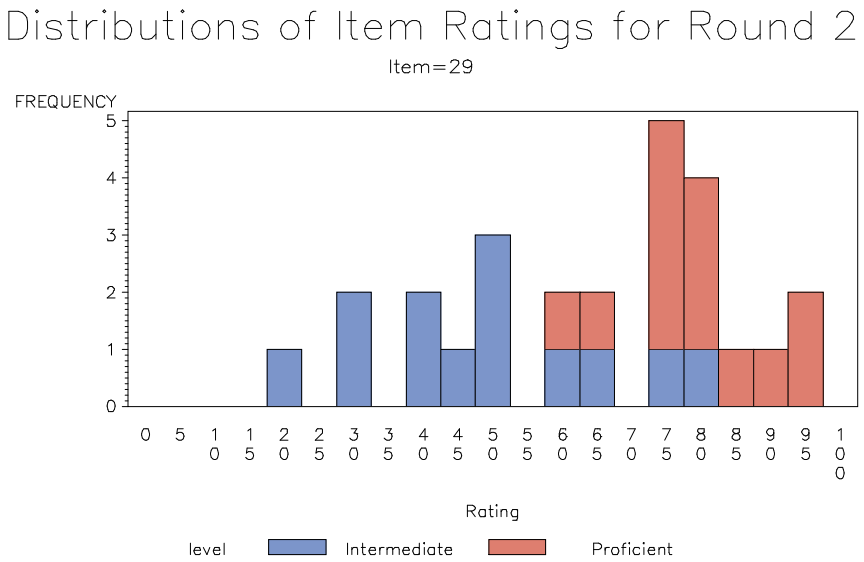

Distributions of Item Ratings for Round 2
Item=24
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

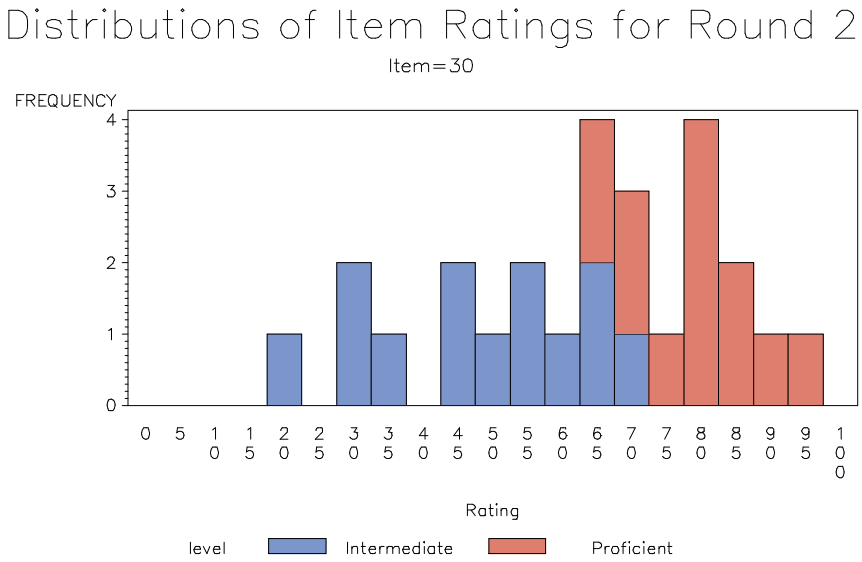

Distributions of Item Ratings for Round 2
Item=25
FREQUENCY
0
1
2
3
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

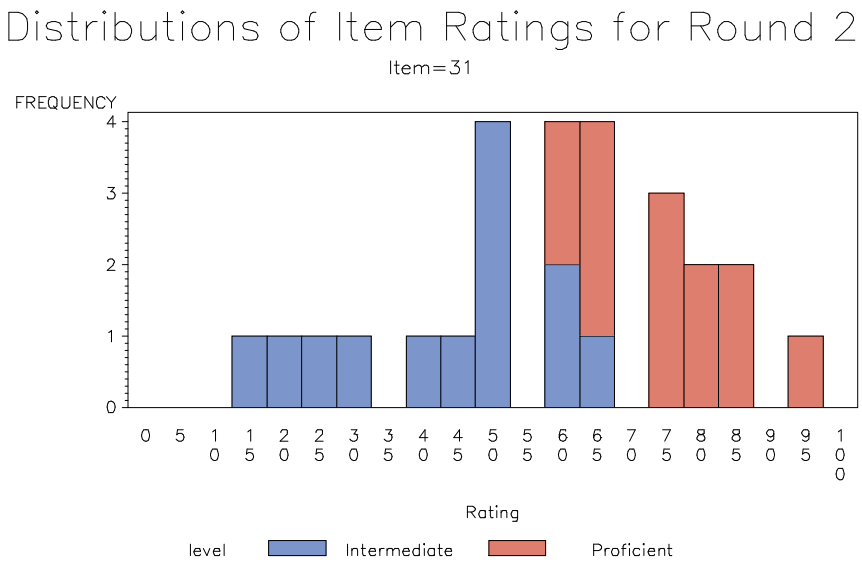

Distributions of Item Ratings for Round 2
Item=26
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

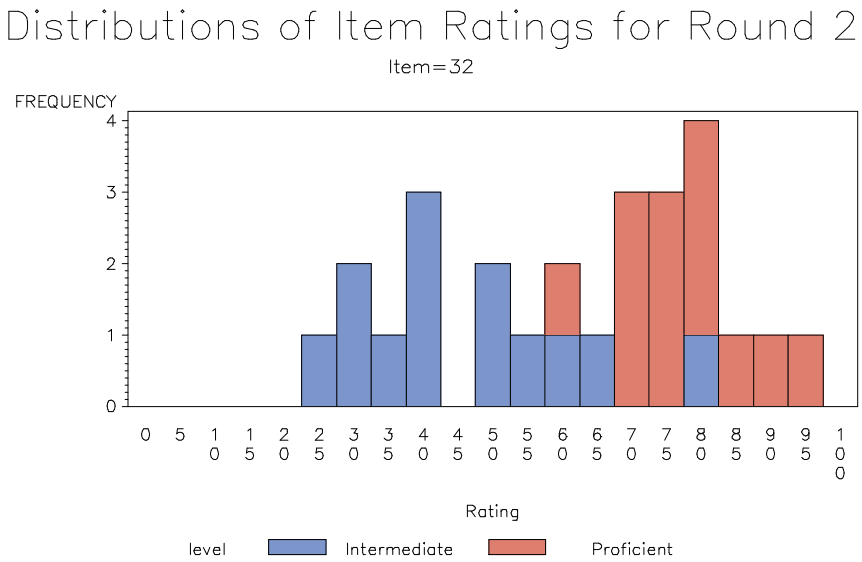

Distributions of Item Ratings for Round 2
Item=27
FREQUENCY
0
1
2
3
4
5
6
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

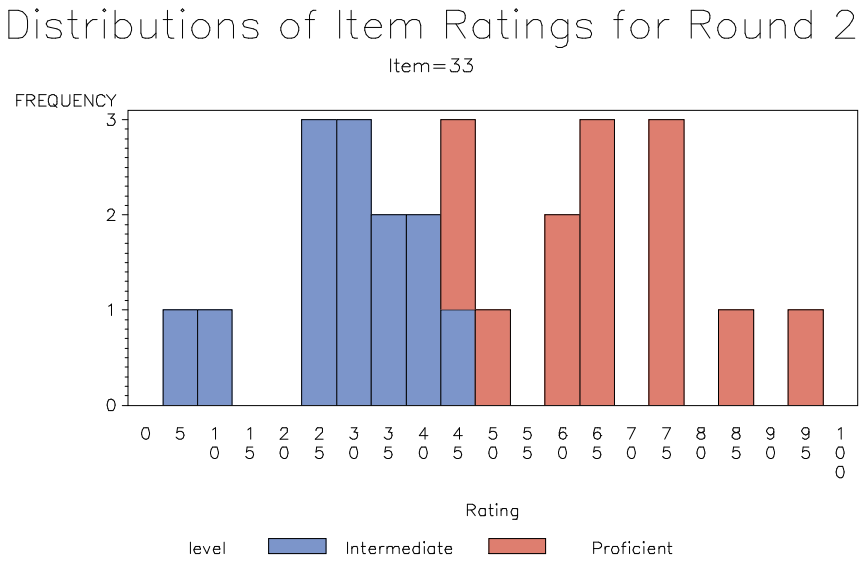

Distributions of Item Ratings for Round 2
Item=28
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

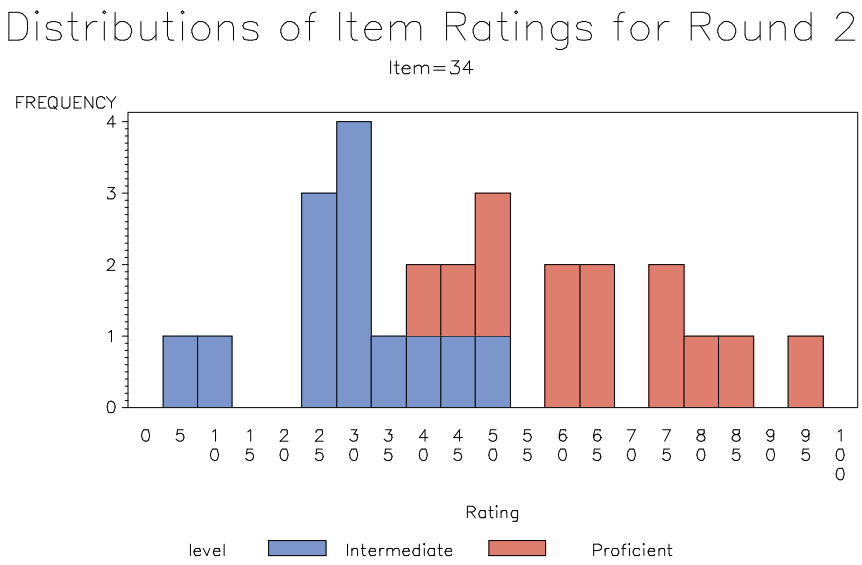

Distributions of Item Ratings for Round 2
Item=29
FREQUENCY
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

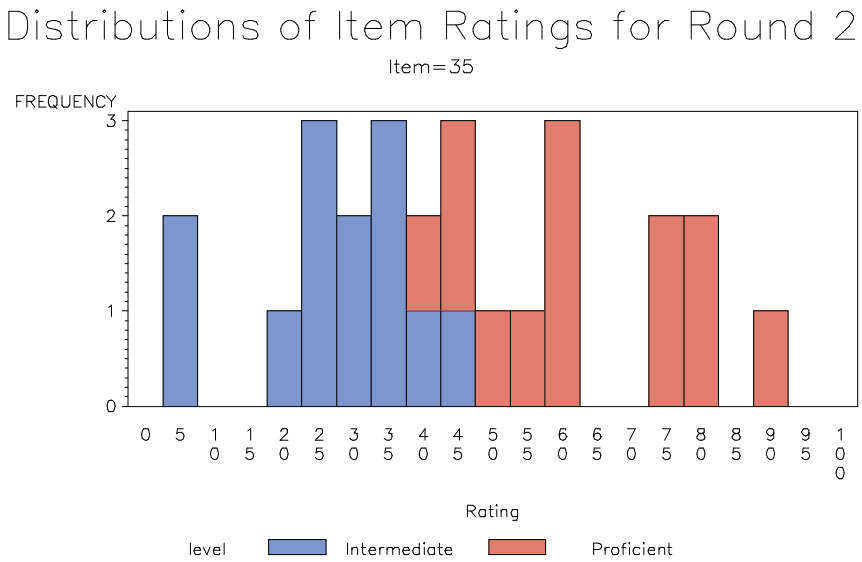

Distributions of Item Ratings for Round 2
Item=30
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

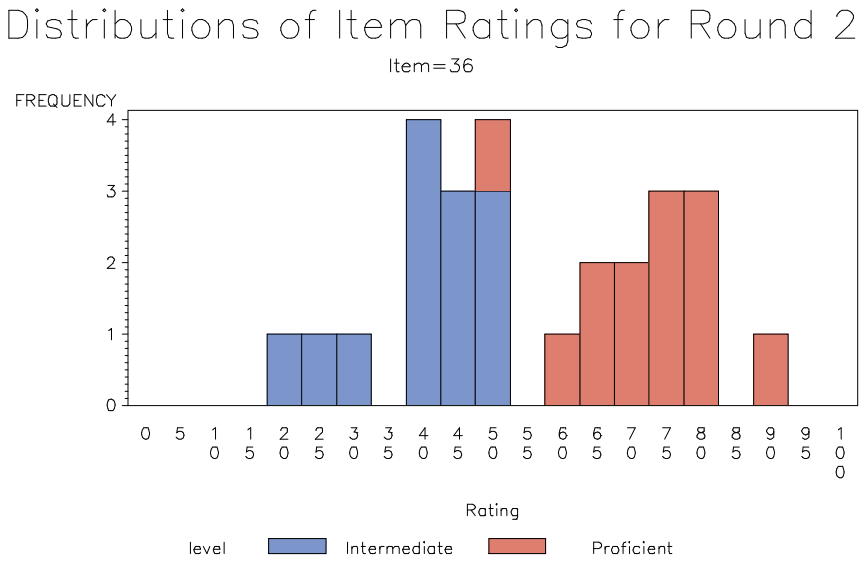

Distributions of Item Ratings for Round 2
Item=31
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

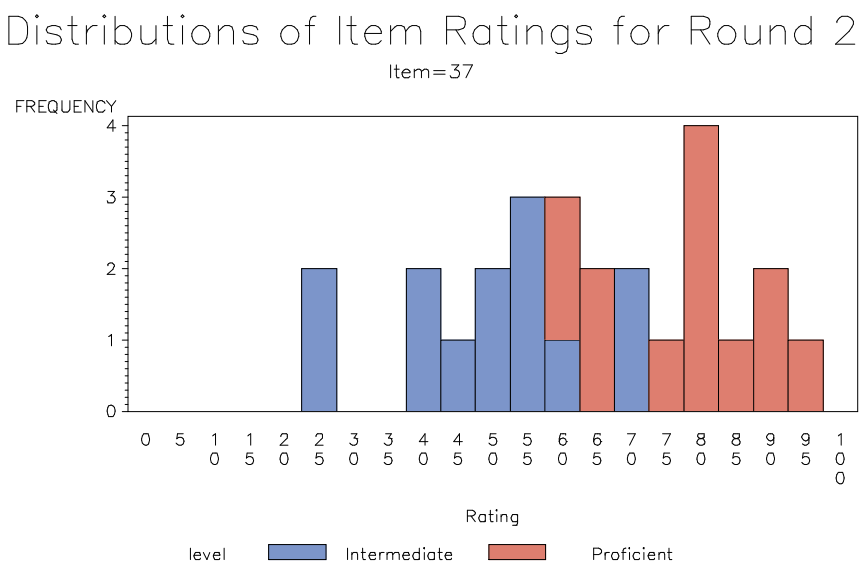

Distributions of Item Ratings for Round 2
Item=32
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient

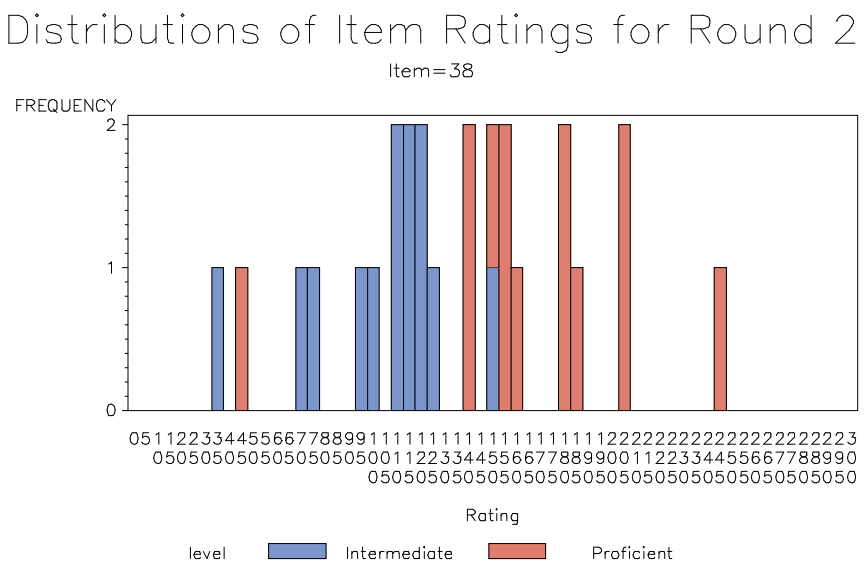

Distributions of Item Ratings for Round 2
Item=33
FREQUENCY
0
1
2
3
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient


Distributions of Item Ratings for Round 2
Item=34
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient


Distributions of Item Ratings for Round 2
Item=35
FREQUENCY
0
1
2
3
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient


Distributions of Item Ratings for Round 2
Item=36
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient


Distributions of Item Ratings for Round 2
Item=37
FREQUENCY
0
1
2
3
4
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Rating
level
Intermediate
Proficient


Distributions of Item Ratings for Round 2
Item=38
FREQUENCY
0
1
2
Rating
level
Intermediate
Proficient

## Appendix L.3: Round by Round Raw Score Cut Summaries

Table L.3.1.

*Minimum, Median, and Maximum Raw Cut Scores by Round*

| | Basic/Intermediate | Proficient |
|---|---|---|
| | *Round 1* | |
| Minimum | 14 | 22 |
| Median | 22 | 32 |
| Maximum | 32 | 38 |
| | *Round 2* | |
| Minimum | 15 | 25 |
| Median | 22 | 32 |
| Maximum | 30 | 38 |
| | *Round 3* | |
| Minimum | 18 | 30 |
| Median | 22 | 32 |
| Maximum | 30 | 38 |

Table L.3.2.

*Median, Minimum, and Maximum Raw Cut Scores by Group after Round 1*

| | Basic/Intermediate | Proficient |
|---|---|---|
| | *Overall* | |
| Minimum | 14 | 22 |
| Median | 22 | 32 |
| Maximum | 32 | 38 |
| | *Group 1* | |
| Minimum | 14 | 30 |
| Median | 22 | 33 |
| Maximum | 23 | 34 |
| | *Group 2* | |
| Minimum | 15 | 29 |
| Median | 22 | 36 |
| Maximum | 32 | 38 |
| | *Group 3* | |
| Minimum | 14 | 22 |
| Median | 20 | 31 |
| Maximum | 22 | 33 |

Table L.3.3.

*Median, Minimum, and Maximum Raw Cut Scores by Group after Round 2*

| | Basic/Intermediate | Proficient |
|---|---|---|
| | *Overall* | |
| Minimum | 15 | 25 |
| Median | 22 | 32 |
| Maximum | 30 | 38 |
| | *Group 1* | |
| Minimum | 15 | 32 |
| Median | 22 | 34 |
| Maximum | 24 | 34 |
| | *Group 2* | |
| Minimum | 17 | 30 |
| Median | 22 | 34 |
| Maximum | 30 | 38 |
| | *Group 3* | |
| Minimum | 18 | 25 |
| Median | 20 | 31 |
| Maximum | 23 | 32 |

Table L.3.4.

*Median, Minimum, and Maximum Raw Cut Scores by Group after Round 3*

| | Basic/Intermediate | Proficient |
|---|---|---|
| | *Overall* | |
| Minimum | 18 | 30 |
| Median | 22 | 32 |
| Maximum | 30 | 38 |
| | *Group 1* | |
| Minimum | 20 | 33 |
| Median | 22 | 34 |
| Maximum | 24 | 34 |
| | *Group 2* | |
| Minimum | 18 | 30 |
| Median | 22 | 32 |
| Maximum | 30 | 38 |
| | *Group 3* | |
| Minimum | 18 | 32 |
| Median | 20 | 32 |
| Maximum | 23 | 32 |