

Oregon

English Language Proficiency Assessments

PRELIMINARY
Technical Report
for
Bookmark Standard Setting
held
November 2007

for

**Reading, Writing,
Listening, and Speaking**
Kindergarten and Grades 1, 2, 5, 7, & 11

Submitted to
Oregon Department of Education
November 2007

 **CTB**
McGraw-Hill

Developed and published under contract with the Oregon Department of Education by CTB/McGraw-Hill LLC, a subsidiary of The McGraw-Hill Companies, Inc., 20 Ryan Ranch Road, Monterey, California 93940-5703. Copyright © 2007* by the Oregon Department of Education. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the Oregon Department of Education and the publisher. This work is based on the Bookmark Standard Setting Procedure, copyright © 2004 by CTB/McGraw-Hill LLC. All rights reserved. Bookmark Standard Setting Procedure is a trademark of The McGraw-Hill Companies, Inc.

Table of Contents

<u>Section</u>	<u>Page</u>
A. Executive Summary	A1-A5
B. Oregon Standard Setting: Day-by-Day Synopsis	B1-B18
C. Master Agenda	C1-C6
D. Training Overheads	D1-D10
E. Detailed Bookmark Judgments	E1-E90
F. Proficiency Level Descriptions	F1-F13
G. Graphical Representations of Participants' Judgments	G1-G24
H. Training Materials	H1-H9
I. Participant Evaluation	I1-I16
J. Charts of Impact Data from Articulation Discussion and Final Cut Scores	J1-J2
K. The Bookmark Standard Setting Procedure: Methodology and Recent Implementations	K1-K20

SECTION A

Executive Summary

DRAFT

Executive Summary

In November 2007, staff members from the Oregon Department of Education (ODE) and CTB/McGraw-Hill worked in collaboration to perform standard setting on the English Language Proficiency Assessments (ELPA). Oregon educators with specialization in English-language development convened to study the ELPA, consider the English-language skills required of students in each proficiency level, and discuss these expectations with their colleagues.

CTB conducted the Oregon ELPA Standard Setting in Salem, Oregon. The purpose of the standard setting was to recommend cut scores on the ELPA to divide students into five proficiency levels: *Beginning*, *Early Intermediate*, *Intermediate*, *Early Advanced* and *Advanced*. The Bookmark Standard Setting Procedure (BSSP) was used to set the proficiency standards for the ELPA.

A committee of educators from across the state of Oregon convened to engage in the standard setting workshop November 5 – 6, 2007. Participants convened to recommend a well-articulated set of proficiency standards at six grades: Kindergarten and Grades 1, 2, 5, 7, and 11. Proficiency standards for the remaining grades were statistically interpolated based on participants' recommendations.

The ODE divided participants into five grade groups, each with approximately 3 participants. Participants were divided into assigned grade groups that were balanced in terms of relevant demographic characteristics (e.g., gender, geographic location). The standard setting consisted of training, orientation, three rounds of judgments, an articulation discussion, and proficiency level description writing.

Following the standard setting, ODE made adjustments to the recommended cut scores. These adjustments were made to accommodate the cut scores to their impact on students, that is, so that a more appropriate distribution of students by proficiency level could be achieved based on 2006-07 performance data. The final recommended cut scores adopted for the ELPA program are shown in Table 1. The impact data associated with these cut scores—the percentage of students classified in each proficiency level—are shown in Table 2.

This report summarizes the results of the Oregon ELPA Standard Setting. A day-by-day synopsis is included in Section B. The master agenda is included in Section C. The overheads presented to participants during training and orientation are in Section D. Section E presents details of the participants' Bookmark judgments for each group. In Section F, estimates are given of the percentages of students in each proficiency level at plus/minus one, two, and three standard errors of the participants' recommended final round cut scores. Section G contains graphical representations of participants' judgments. The training materials given to participants are provided in Section H. Section I contains the results of the participants' evaluation of the workshop. Section J

contains the data and charts from the articulation discussion as well as the final recommendations given to the state. As a reference for the reader, Section K presents *The Bookmark Standard Setting Procedure: Methodology & Recent Implementations* (Lewis, Green, Mitzel, Baum, & Patz, 1998). The proficiency level descriptions are included in Section L.

Table 1. *Final recommended cut scores for ELPA.*

Grade	Beginning	Early Intermediate	Intermediate	Early Advanced	Advanced
K	<482	482	492	498	507
1	<492	492	507	514	523
2	<495	495	508	514	523
3*	<501	501	514	521	529
4*	<497	497	508	514	521
5	<497	497	508	516	523
6*	<497	497	506	515	522
7	<497	497	507	517	524
8*	<499	499	508	518	526
9*	<491	491	501	515	526
10*	<493	493	501	516	527
11	<494	494	501	515	528
12*	<498	498	504	516	530

*Based on interpolated data.

Table 2. *Impact data associated with the final cut scores for the ELPA.*

Grade	Beginning	Early Intermediate	Intermediate	Early Advanced	Advanced
K	15.8%	44.6%	21.3%	13.7%	4.7%
1	14.0%	41.3%	21.4%	17.2%	6.1%
2	13.9%	38.7%	19.9%	20.9%	6.5%
3*	10.9%	31.1%	25.8%	21.9%	10.3%
4*	9.3%	24.4%	23.7%	29.0%	13.5%
5	6.9%	16.3%	28.5%	32.2%	16.2%
6*	6.7%	13.9%	33.1%	32.0%	14.2%
7	7.5%	11.7%	36.1%	31.3%	13.4%
8*	8.1%	10.9%	35.1%	34.1%	11.8%
9*	8.6%	8.9%	33.1%	40.1%	9.3%
10*	8.6%	8.6%	33.0%	40.9%	8.9%
11	8.8%	7.6%	30.0%	46.2%	7.4%
12*	8.8%	6.9%	29.6%	48.1%	6.5%

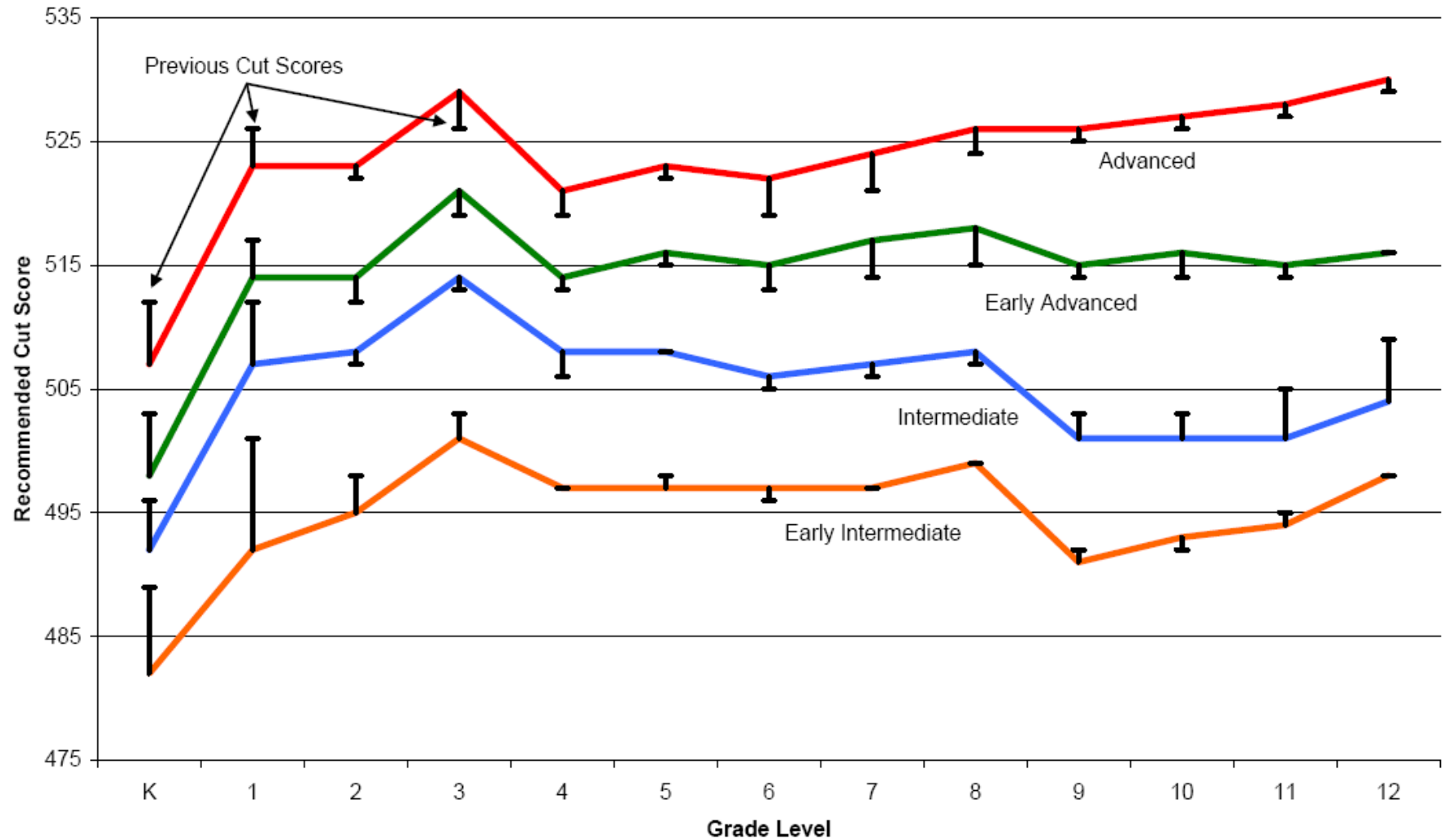
*Based on interpolated data.

Table 3. *Recommended Changes in Cut Scores.*

The recommendations from this session and comparisons to current standards are presented in the table below. Recommended standards are presented in **bold**; current standards are in (parentheses); arrows indicate the direction of the recommended change [↓ lower, ↑ higher, or ⇔ unchanged].

Grade Level	Early Intermediate	Intermediate	Early Advanced	Advanced
K	492 (489) ↓	492 (496) ↓	498 (503) ↓	507 (512) ↓
1	492 (501) ↓	507 (512) ↓	514 (517) ↓	523 (526) ↓
2	495 (498) ↓	508 (507) ↑	514 (512) ↑	523 (522) ↑
3	501 (503) ↓	514 (513) ↑	521 (519) ↑	529 (526) ↑
4	497 (497) ⇔	508 (506) ↑	514 (513) ↑	521 (519) ↑
5	497 (498) ↓	508 (508) ⇔	516 (515) ↑	523 (522) ↑
6	497 (496) ↑	506 (505) ↑	515 (513) ↑	522 (519) ↑
7	497 (497) ⇔	507 (506) ↑	517 (514) ↑	524 (521) ↑
8	499 (499) ⇔	508 (507) ↑	518 (515) ↑	526 (524) ↑
9	491 (492) ↓	501 (503) ↓	515 (514) ↑	526 (525) ↑
10	493 (492) ↑	501 (503) ↓	516 (514) ↑	527 (526) ↑
11	494 (495) ↓	501 (505) ↓	515 (514) ↑	528 (527) ↑
12	498 (498) ⇔	504 (509) ↓	516 (516) ⇔	530 (529) ↑

Recommended ELPA Cut Scores
With Magnitude and Direction of Recommended Change



SECTION B

Oregon Standard Setting: Day-by-Day Overview

DRAFT

Oregon ELPA Standard Setting: Day-by-Day Synopsis

The Oregon Department of Education (ODE) partnered with CTB/McGraw-Hill (CTB) to perform standard setting on the English Language Proficiency Assessments (ELPA). The purpose of the standard setting was to identify cut scores on the ELPA to divide students into five proficiency levels: *Beginning*, *Early Intermediate*, *Intermediate*, *Early Advanced* and *Advanced*.

CTB staff conducted the Oregon ELPA Standard Setting in Salem, Oregon. The Bookmark Standard Setting Procedure (BSSP; Lewis, Mitzel & Green, 1996; Mitzel, Lewis, Patz, & Green, 2001) was used to set the proficiency standards for the ELPA. The standard setting consisted of training, orientation, three rounds of judgments, an articulation discussion, and proficiency level description writing.

This document describes the implementation of the BSSP to establish cut scores at selected grades, the interpolation of cut scores for the off-grades, the articulation discussion, and proficiency level description writing for the Oregon ELPA.

The Oregon ELPA Standard Setting workshop was held in Salem, Oregon, on November 5 – 6, 2007. Oregon educators with specialization in English-language development convened to study the ELPA, consider the English-language skills required of student in each proficiency level, and discuss these expectations with their colleagues. The BSSP was implemented to set standards for six grades: Kindergarten and Grades 1, 2, 5, 7, and 11. CTB interpolated cut scores for the remaining grades—Grades 3, 4, 6, 8, 9, 10, and 12—based on participants' recommended cut scores.

Bookmark Roles

CTB Staff

The CTB Standard Setting Team worked with staff from the ODE to design, organize, and conduct the standard setting activities. The CTB Standard Setting Team was composed of Ricardo Mercado, Research Project Manager; Dr. Christina Schneider, Research Scientist; Dorothy Tele'a, Standard Setting Specialist; and Adele Brandstrom, Standard Setting Specialist.

Prior to the workshop, the CTB Standard Setting Team prepared all materials for the workshop. During the workshop, the team was responsible for facilitating the workshop, training participants, entering participant results into a database, and tracking secure materials. Following the workshop, the team prepared the standard setting technical report.

Sandra Snell, CTB Program Manager, Nadia Greer, CTB Program Office Coordinator, and Agneta Lenberg, CTB National Accounts Manager, attended the standard setting and helped with on-site logistics.

Lorena Houston, CTB Development Manager, and Deborah Busch, CTB Content Editor, attended the standard setting and served as group leaders.

Group Leaders

At the standard setting, the group leaders helped implement the BSSP for each of the five groups. Group leaders were staff members from CTB Development with expertise in English-language development. A description of the group leader's role follows.

Group leader. The group leader served as a facilitator and was in charge of time management, focusing the participants on the task at hand and interacting with the participants. The group leader also facilitated discussions and was in charge of security and data management. The group leader collected the rating forms from participants and communicated with CTB Research and the ODE staff. Group leaders were non-voting members of the workshop staff.

Participants

Participants were recruited from across the state of Oregon. All participants were selected by the ODE such that the committees were composed of a diverse, experienced group of Oregon educators with specialization in English-language development. The standard setting committee comprised 14 participants.

The committee was divided into five groups: Kindergarten and Grades 2, 5, 7, and 11. The Kindergarten group also set standards for Grade 1, allowing participants a richer dialogue about the English-language skills expected of students at these lowest grade levels. Each of the groups comprised approximately three participants. Table 1 shows the number of participants for each grade. Note that the committee included Oregon educators, administrators, and stakeholders to add a diversity of perspectives to the discussions held at the workshop.

Table 1. Number of participants for the standard setting workshop by grade.

Grade	Number of Participants
K & 1	3
2	3
5	2
7	3
11	3
Total	14

Configuration of the Grade Panels

The ODE assigned participants such that each table was as representative and balanced as possible in regard to the relevant demographic characteristics (e.g., gender, geographic location). In addition, each group was asked by ODE to select their own table leader. A description of the table leaders' role follows.

Table leaders. Table leaders were experienced educators within the English-language development community and were chosen from among the participants. Some table leaders had a previous role with the assessment, such as serving as item-writers. The primary role of the table leader was to monitor the group discourse, keep the group focused on the task at hand, and keep time for the group. As needed, table leaders found a diplomatic middle ground between participants or requested assistance from CTB and the ODE. Table leaders were voting members of their panels.

Committee Demographics

Following the workshop, all 14 participants completed written evaluations from which CTB collected self-reported demographic information. This information about the participants has been summarized. Table 2 shows the educational background of the participants at each workshop, and Table 3 shows their primary role. At the standard setting, 100% of the participants described their role at the workshop as educators. Tables 4 and 5 show the occupation and work experience of the participants. Approximately 57% of the participants were teachers or administrators. The remainder of participants in the workshop listed their occupation as “Other,” and included English-language development specialists and content experts.

Table 6 shows participants’ experience teaching English-language learners and students with disabilities. At the standard setting, approximately 29% of participants had experience with Special Education, 100% with ESL/ELD, 21% with Vocational Education, 29% with Alternative Education, and 64% with Adult Education. Section I contains the complete results of the participant evaluation from the workshop.

Table 2. Educational background of participants by grade.

Grade	N	HSD or GED	Bachelor's	Master's	Doctorate
Overall	14	0.0%	7.1%	85.7%	7.1%
K & 1	3	0.0%	0.0%	100.0%	0.0%
2	3	0.0%	0.0%	100.0%	0.0%
5	2	0.0%	50.0%	50.0%	0.0%
7	3	0.0%	0.0%	100.0%	0.0%
11	3	0.0%	0.0%	66.7%	33.3%

Table 3. Primary role of participants by grade.

Grade	N	Educator	Parent	Community Member	Business Member
Overall	14	100.0%	0.0%	0.0%	0.0%
K & 1	3	100.0%	0.0%	0.0%	0.0%
2	3	100.0%	0.0%	0.0%	0.0%
5	2	100.0%	0.0%	0.0%	0.0%
7	3	100.0%	0.0%	0.0%	0.0%
11	3	100.0%	0.0%	0.0%	0.0%

Table 4. Occupation of participants by grade.

Grade	N	Teacher	Administrator	Instructional Assistant	Other
Overall	14	50.0%	7.1%	0.0%	42.9%
K & 1	3	66.7%	0.0%	0.0%	33.3%
2	3	33.3%	33.3%	0.0%	33.3%
5	2	50.0%	0.0%	0.0%	50.0%
7	3	100.0%	0.0%	0.0%	0.0%
11	3	0.0%	0.0%	0.0%	100.0%

Table 5. Work experience in years of participants by grade.

Grade	N	1-5	6-10	11-15	16-20	21+
Overall	13	15.4%	15.4%	30.8%	7.7%	30.8%
K & 1	3	0.0%	0.0%	66.7%	0.0%	33.3%
2	2	50.0%	50.0%	0.0%	0.0%	0.0%
5	2	0.0%	0.0%	100.0%	0.0%	0.0%
7	3	0.0%	33.3%	0.0%	0.0%	66.7%
11	3	33.3%	0.0%	0.0%	33.3%	33.3%

Table 6. Experience of participants by grade, teaching English-language learners, students with disabilities, and other special groups.

Grade	N	Special Ed.	N	ESL/ ELD	N	Voca-tional Ed.	N	Alternative Ed.	N	Adult Ed.
Overall	14	28.6%	14	100.0%	14	21.4%	14	28.6%	14	64.3%
K & 1	3	66.7%	3	100.0%	3	0.0%	3	33.3%	3	66.7%
2	3	0.0%	3	100.0%	3	33.3%	3	0.0%	3	33.3%
5	2	0.0%	2	100.0%	2	0.0%	2	0.0%	2	50.0%
7	3	66.7%	3	100.0%	3	66.7%	3	66.7%	3	66.7%
11	3	0.0%	3	100.0%	3	0.0%	3	33.3%	3	100.0%

Bookmark Materials

Ordered Item Booklets

The Ordered Item Booklets (OIB) for each grade was made up of multiple-choice (MC) items, other dichotomously scored items, and polytomously scored constructed-response (CR) items. The ODE selected items for the OIBs with a broad range of difficulty levels—easy to hard—and from each of the four skill areas: Listening, Speaking, Reading, and Writing. More items were selected for the OIBs than would be administered to a single student, as shown in Table 8.

The items for each grade were ordered according to their scale location using a response probability criterion of 0.67. With this criterion, each MC or other dichotomously scored item was located at the ability level that students would need in order to have a 0.67 probability of answering the item correctly. Each non-zero score level associated with a CR item was located at the ability level that students would need in order to have a 0.67 probability of attaining that score level. The Rasch model was used to scale the MC and other dichotomously scored items. The Rating Scale model (Wright & Masters, 1982) was used to scale the polytomously scored CR items. For more information about the construction of the OIBs, see Lewis, Green, Mitzel, Baum, & Patz (1998), which is included in Section K. Additionally, Beretvas (2004) includes a discussion of the calculation of response probability-adjusted locations for items scaled with the Rasch model.

Table 7 shows the percentage of items in each OIB that measure each of the ELPA skill areas: Listening, Reading, Speaking, and Writing. Table 8 shows the total number of score points—the MC and other dichotomously scored items plus non-zero CR score levels—in each OIB for each skill area. It should be noted that CR items comprised only a fraction of the total score points in the OIBs. Note that the Listening and Reading portions of the tests were comprised exclusively of dichotomously-scored items, and that the K–1 test had no CR items.

Table 7. Percent of items in the OIB that measure each ELPA skill area by grade.

Grade	Listening	Reading	Speaking	Writing	Total
K-1	27%	33%	14%	26%	100%
2-3	23%	27%	22%	28%	100%
4-5	32%	27%	13%	28%	100%
6-8	28%	25%	16%	31%	100%
9-12	29%	29%	14%	29%	100%

Table 8. Total number of score points in each OIB, by skill area and grade.

Grade	Listening	Reading	Speaking	Writing	Total
K	23	28	12	22	85
2	22	25	27	30	104
5	25	21	13	26	85
7	22	20	16	30	88
11	23	23	13	27	86

Item Maps

The item map for each grade included the order of difficulty, location, domain, item type, assessment point (illocutionary competence or grammatical competence), language function, and score key (correct response or number of score points). Participants filled in the final two columns as they studied the items in the OIB. The first of these columns asks, “What does this item measure? That is, what do you know about a student who can respond successfully to this item/score point?” The second of these columns asks “Why is this item more difficult than the preceding items?”

Workshop Synopsis

This section presents a chronological description of the events at the standard setting workshop. The ODE and CTB conducted the opening session and training on the first morning of the workshop; the remaining time was used for standard setting activities and proficiency level description writing.

Standard Setting: Day 1

The implementation of the BSSP consisted of training, orientation, and three rounds of judgments. This was followed by proficiency level description writing and an articulation discussion.

Opening Session

Tony Alpert, ODE’s Director of Assessment and Accountability, gave the welcoming address and described the purpose of the standard setting. The ODE described the expectations for the type of cut scores that the state anticipated from the process.

Training

Following the presentation by the ODE, Ricardo Mercado, a member of the CTB Standard Setting Team, provided an overview of the purpose of the standard setting and described the implementation of the BSSP. Participants were introduced to key concepts and materials of the BSSP, including the OIB and the item map. During this training, it was explained that table leaders would facilitate discussion at their tables and help participants in completing tasks in a timely manner. Participants were given a synopsis of each day’s activities. The Master Agenda is included in Section C, and the training slides are included in Section D.

Participants then engaged in a brief, mock standard setting using sample items from a CTB-published test of English-language proficiency. During the mock standard setting, participants reviewed the tools of the BSSP, including a sample OIB and item map.

Following the mock standard setting, participants were directed to their tables. Each grade was represented by one table, except for Kindergarten and Grade 1, which were represented by a single committee. All participants met in a single, large meeting room.

Target Student Descriptions

Participants were presented the proficiency level descriptions. Participants were instructed to familiarize themselves with the descriptions, and to discuss the English-language skills of each target student.

A target student is defined as a student whose performance minimally meets the criteria for entry into a particular proficiency level, for example, the “just” *Early Advanced* student. For each grade there were four target student descriptions, one for each cut score (*Early Intermediate, Intermediate, Early Advanced, and Advanced*). Participants were encouraged to take notes during the target student discussion and were referred to the target student descriptions throughout the standard setting.

Examine the Test

Participants examined items in the OIB for their grade to familiarize themselves with the items.

Study the Scoring Guides (Rubrics)

The group leader oriented participants to the scoring guides for the CR items, for Writing and Speaking. Participants were directed to study the scoring guides at their tables and to discuss the differences between responses for each score level.

Study Items in the Ordered Item Booklet

Participants at each table studied each of the items in the OIB in terms of what each item measured and why it was more difficult than the items preceding it. Participants recorded their notes about the items on the item maps. At each table, one participant, denoted as the scribe, recorded the group’s comments about each item.

Review Bookmark Placement

Prior to setting their Round 1 bookmarks, Dr. Christina Schneider, a member of the CTB Standard Setting Team, presented a refresher of bookmark placement. Participants were instructed to use four tools when placing their bookmarks: the Oregon ELPA standards, the target student descriptions, the proficiency level descriptions, and the English-language skills represented by the items.

Participants were given training materials and three explanations of bookmark placement. The training materials titled “Bookmark Placement” and “Frequently Asked Questions about Bookmark Placement” were summarized orally to all participants. The first explanation of bookmark placement demonstrated the mechanics: participants were instructed that all items preceding the bookmark define the knowledge, skills, and abilities that a “just” *Early Advanced* student, for example, is expected to know. The second explanation of bookmark placement was more conceptual in that participants were instructed to examine each item in terms of its skills and to make a judgment about the type of skills that a student would need to know in order to be considered, for example, “just” *Early Advanced*. The final explanation discussed the relationship between the bookmarks and the scale scores, as described in the training material titled “Mastery.” The bookmark training materials are included in Section H.

The participants were tested on their understanding of bookmark placement with a short check set. The check set questions and the results are presented in Tables 9 and 10, respectively. After participants took the check set, the correct answers were provided and the rationales for the correct answers were discussed. The responses to the check set indicated that participants understood how to place their bookmarks. The check set (and its graphic) is included in Section H.

Table 9. Questions in the check set that followed bookmark training.

Question	
1	Which items does a student need to master to just make it into the <i>Early Advanced</i> level?
2	If a student mastered only items 1 through 2, in which performance level would this student be?
3	Suppose a student mastered items 1 through 13. Which performance level is this student in?
4	For students who are classified as <i>Early Advanced</i> , with at least what likelihood will they be able to answer item 10?
5	Will the items BEFORE the <i>Early Advanced</i> bookmark be more or less difficult to answer than the items AFTER the bookmark or about the same?

Table 10. Number and percent of participants responding correctly to each question on the check set ($N = 12$).

Question	# Correct	% Correct
1	12	100%
2	12	100%
3	11	92%
4	12	100%
5	12	100%

Round 1 Bookmark Placement

Once participants demonstrated that they understood how to place their bookmarks through the check set, they placed bookmarks in the following order: *Early Intermediate*, *Intermediate*, *Early Advanced*, and *Advanced*. The training materials indicated that the bookmarks should be placed starting with *Early Advanced*, but participants felt it was cognitively simpler to place the bookmarks in order from the lowest to the highest proficiency level. Participants were instructed that bookmark placement is always an individual activity.

Participants placed their Round 1 bookmarks for *Early Intermediate*, *Intermediate*, *Early Advanced*, and *Advanced*, while keeping in mind the Oregon ELPA standards, the target student descriptions, the proficiency level descriptions, and the English-language skills measured by the items on the test. Participants in the Kindergarten and Grade 1 group set their bookmarks for both grades at the same time.

As suggested in Section K, data from the second round of Bookmark judgments are typically used in the calculation of the standard error of the cut score (SE_{bk}). This round is selected because it is the final round where participants work independently within their groups. However, when only one group of participants is used to recommend proficiency standards—as was the case at this standard setting—it may be useful to examine the conventional standard error of participants' judgments in each round of the Bookmark Procedure. These results are shown in Table 11.

Table 11. Standard error (conventional calculation) of participants' cut score recommendations, by grade and round.

Grade	Round	<i>Early Intermediate</i>	<i>Intermediate</i>	<i>Early Advanced</i>	<i>Advanced</i>
K	1	1.77	2.00	1.20	1.00
	2	1.67	2.19	0.33	0.58
	3	0.00	0.66	0.33	0.33
1	1	0.33	0.33	0.33	0.88
	2	0.00	0.33	0.00	1.00
	3	1.00	0.58	0.00	1.33
2	1	0.58	3.38	2.85	4.34
	2	0.66	3.00	3.00	1.85
	3	0.66	1.00	3.00	1.33
5	1	1.00	1.00	1.00	3.50
	2	0.50	0.50	0.00	0.00
	3	0.00	0.00	0.00	0.00
7	1	2.19	0.88	1.15	2.31
	2	0.00	0.33	0.66	0.66
	3	0.00	0.00	0.33	0.33
11	1	5.33	3.00	1.77	0.66
	2	1.00	0.33	0.33	1.53
	3	0.00	2.19	0.00	0.00

As shown in Table 11, the standard error of participants' cut score recommendations tended to decrease with each round. Note that a standard error of zero does not necessarily indicate that the cut scores would not change if another group of participants were selected for the workshop from the same pool of qualified participants: rather, it indicates that the standard setting committee reached consensus on the recommendation for a given cut score.

Standard Setting: Day 2

Round 2 Bookmark Placement

In each grade, the table leader facilitated a discussion of all the bookmark placements for the table. Participants were encouraged to focus on the differences among their bookmarks by discussing the items between the lowest and highest bookmarks at their table.

Participants were then directed back to their OIBs and item maps to continue their discussions of the English-language skills expected of students in each proficiency level. After discussion, participants were reminded to place their bookmarks independently.

Round 3 Bookmark Placement

Participants received feedback based on their Round 2 bookmark placements from a member of the CTB Standard Setting Team in collaboration with an ODE representative. Participants were shown the median bookmark placement for each proficiency level for their grade. In addition, participants were shown impact data based on the median Round 2 bookmarks. CTB staff answered process-related questions, and the ODE staff answered all policy-related questions concerning the impact data. It was emphasized to the participants that the impact data were being presented as a "reality check."

After the presentation of Round 2 results, participants discussed the rationale of their bookmark placements within their grade. The group leader facilitated the discussion among all participants. After the discussion, participants were instructed to place their bookmarks independently for the final time.

Round 3 Results

Participants received feedback based on their final bookmark placements from a member of the CTB Standard Setting Team in collaboration with an ODE representative. On an overhead slide, participants were shown the median bookmarks for each table as well as the medians for their grade and the impact data based on the median final bookmarks. In addition, participants were shown the impact data for all grades as an introduction to the articulation discussion. The impact data are from the Oregon ELPA Spring 2007 administration.

Table 12 shows the participant-recommended cut scores and associated impact data based on Round 3. The impact data in Table 12 were shown to the participants at the workshop.

Table 12. Participant-recommended cut scores and associated impact data, based on the final round of bookmark placements.

Grade	Cut Scores				Impact Data				
	<i>Early Inter.</i>	<i>Inter.</i>	<i>Early Adv.</i>	<i>Adv.</i>	<i>Beg.</i>	<i>Early Inter.</i>	<i>Inter.</i>	<i>Early Adv.</i>	<i>Adv.</i>
Kindergarten	479	489	495	500	6.2%	39.9%	27.3%	12.9%	13.7%
1	492	499	506	523	14.0%	17.1%	21.3%	41.5%	6.1%
2	495	508	514	523	13.9%	38.7%	19.9%	20.9%	6.5%
5	497	508	516	523	6.9%	16.3%	28.5%	32.2%	16.2%
7	497	507	518	524	7.5%	11.7%	40.3%	27.1%	13.4%
11	494	501	515	528	8.8%	7.6%	30.0%	46.2%	7.4%

Section E presents details of the participants' Bookmark judgments for each grade. In Section F, the proficiency level descriptions are included. Section G contains graphical representations of participants' judgments and standard errors.

Evaluations

Following the presentation of final results, participants were asked to complete an evaluation of the standard setting. Some results of the evaluation have been presented in earlier sections, and the results of one statement are shown in Table 13. More results from the evaluation are presented in Tables 17-21. Complete results of the evaluation are included in Section I.

Participants were asked to respond to the statement, "Overall, I am satisfied with my group's final bookmarks." The majority of participants agreed or strongly agreed as shown in Table 13.

Table 13. Participants' agreement/disagreement with the statement, "Overall, I am satisfied with my group's final bookmarks."

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	7.1%	42.9%	50.0%	92.9%
K & 1	3	0.0%	0.0%	33.3%	66.7%	0.0%	66.7%
2	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
5	2	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Interpolation

After all grade panels completed Round 3, CTB interpolated the cut scores for the intervening grades—Grades 3, 4, 6, 8, 9, 10, and 12—using piece-wise linear interpolation. This interpolation method was specified *a priori* by the ODE.

To calculate cut scores for the intervening grades, the impact data associated with the cut scores for each standard setting grade were collected. Next, CTB calculated the impact data points for the intervening grades through piece-wise linear interpolation. Lastly, the cut scores that yielded these impact data points were calculated.

An example of this calculation follows. The *Intermediate* cut score for Grade 5 is 508, and 23.2% of Grade 5 students fall below it. In Grade 7, 19.2% of students fall below the *Intermediate* cut score of 507. For Grade 6, the cut score that would permit the average of these impact data points—approximately 21.2%—was identified. The cut score that yields impact data closest to this value, 20.6%, was then identified as 506. The same procedure was performed for the other cut scores and grades. Table 14 shows the interpolated cut scores and impact data for Grades 3, 4, 6, 8, 9, 10, and 12.

Table 14. Interpolated cut scores and impact data for Grades 3, 4, 6, 8, 9, 10, and 12, as based on the participant-recommended cut scores and associated impact data from the final round of bookmark placements.

Grade	Interpolated Cut Scores				Impact Data				
	Early Inter.	Inter.	Early Adv.	Adv.	Beg.	Early Inter.	Inter.	Early Adv.	Adv.
3	501	514	521	529	10.9%	31.1%	25.8%	21.9%	10.3%
4	497	508	514	521	9.3%	24.4%	23.7%	29.0%	13.5%
6	497	506	515	522	6.7%	13.9%	33.1%	32.0%	14.2%
8	499	508	518	526	8.1%	10.9%	35.1%	34.1%	11.8%
9	491	501	516	526	8.6%	8.9%	37.1%	36.1%	9.3%
10	493	501	516	527	8.6%	8.6%	33.0%	40.9%	8.9%
12	498	504	515	530	8.8%	6.9%	26.7%	51.1%	6.5%

Description Writing

The Group Leader introduced the process for description writing. Participants recommended changes to the existing proficiency level descriptions that detailed the English-language skills needed to be classified in each proficiency level. CTB Development incorporated the changes recommended by the participants. Section F contains the final proficiency level descriptions from the workshop.

Articulation (Smoothing) Discussion

Following description writing, all participants engaged in an articulation (smoothing) discussion. The purpose of this discussion was to establish a system of cut scores that was well-articulated and, at the same time, respectful of the committee's original recommendations. A representative from the ODE was present during these discussions to answer policy-related questions.

Participants recommended no changes to the cut scores for Grades 2, 3, 4, 5, 6, 8, 10, and 11. Participants for Kindergarten recommended a five-point increase for the *Advanced* cut score to bring the percentage of students classified as *Early Advanced* and *Advanced* more in line with the other grades. In addition, participants for Kindergarten also recommended changes for Grade 1. Participants recommended a three-point decrease for the *Early Intermediate* cut score and a four-point increase for the *Early Advanced* cut score to bring the percentage of students in classified in *Beginning*, *Early Intermediate*, *Intermediate*, and *Early Advanced* more in line with the other grades. Using similar reasoning, participants recommended a one-point decrease for the *Early Advanced* interpolated cut score for Grade 7. Participants for Grade 11 recommended a one-point decrease for the *Early Advanced* interpolated cut score for Grade 9 and a one-point increase for the *Early Advanced* interpolated cut score for Grade 12 to promote better articulation with surrounding grades.

At the conclusion of the articulation discussion, all participants were asked to review their recommended cut scores in their OIBs and item maps. Specifically, participants were asked to verify that the changes they recommended during the articulation discussion were reasonable when compared to the skills of the assessments. All participants reported that their recommended cut scores were reasonable when compared to the skills of the assessments.

Table 15 shows the cut scores developed during the articulation discussion, as well as the associated impact data. Data points italicized in Table 15 represent cut scores and impact data changed by participants during the articulation discussion. Section J contains a graphical representation of the impact data from the articulation discussion.

Following the standard setting, ODE and CTB made further adjustments to the cut scores for Kindergarten and Grade 1. Table 16 shows the final cut scores, as well as the associated impact data. Section J contains a graphical representation of the final impact data.

Table 15. Cut scores and associated impact data, after the articulation discussion.

Grade	Cut Scores				Impact Data				
	Early Inter.	Inter.	Early Adv.	Adv.	Beg.	Early Inter.	Inter.	Early Adv.	Adv.
Kindergarten	479	489	495	505	6.2%	39.9%	27.3%	20.4%	6.3%
1	489	499	510	523	8.8%	22.3%	33.3%	29.5%	6.1%
2	495	508	514	523	13.9%	38.7%	19.9%	20.9%	6.5%
3*	501	514	521	529	10.9%	31.1%	25.8%	21.9%	10.3%
4*	497	508	514	521	9.3%	24.4%	23.7%	29.0%	13.5%
5	497	508	516	523	6.9%	16.3%	28.5%	32.2%	16.2%
6*	497	506	515	522	6.7%	13.9%	33.1%	32.0%	14.2%
7	497	507	517	524	7.5%	11.7%	36.1%	31.3%	13.4%
8*	499	508	518	526	8.1%	10.9%	35.1%	34.1%	11.8%
9*	491	501	515	526	8.6%	8.9%	33.1%	40.1%	9.3%
10*	493	501	516	527	8.6%	8.6%	33.0%	40.9%	8.9%
11	494	501	515	528	8.8%	7.6%	30.0%	46.2%	7.4%
12*	498	504	516	530	8.8%	6.9%	29.6%	48.1%	6.5%

* Based on interpolated data.

Table 16. Final cut scores and associated impact data.

Grade	Cut Scores				Impact Data				
	Early Inter.	Inter.	Early Adv.	Adv.	Beg.	Early Inter.	Inter.	Early Adv.	Adv.
Kindergarten	482	492	498	507	15.8%	44.6%	21.3%	13.7%	4.7%
1	492	507	514	523	14.0%	41.3%	21.4%	17.2%	6.1%
2	495	508	514	523	13.9%	38.7%	19.9%	20.9%	6.5%
3*	501	514	521	529	10.9%	31.1%	25.8%	21.9%	10.3%
4*	497	508	514	521	9.3%	24.4%	23.7%	29.0%	13.5%
5	497	508	516	523	6.9%	16.3%	28.5%	32.2%	16.2%
6*	497	506	515	522	6.7%	13.9%	33.1%	32.0%	14.2%
7	497	507	517	524	7.5%	11.7%	36.1%	31.3%	13.4%
8*	499	508	518	526	8.1%	10.9%	35.1%	34.1%	11.8%
9*	491	501	515	526	8.6%	8.9%	33.1%	40.1%	9.3%
10*	493	501	516	527	8.6%	8.6%	33.0%	40.9%	8.9%
11	494	501	515	528	8.8%	7.6%	30.0%	46.2%	7.4%
12*	498	504	516	530	8.8%	6.9%	29.6%	48.1%	6.5%

* Based on interpolated data.

Evaluation of Training

An indication of the effectiveness of training may be found in the participants' answers to statements and questions on the evaluations. Table 17 shows that all participants agreed or strongly agreed that they understood how to place their bookmarks. Table 18 summarizes that most participants agreed or strongly agreed that the task of bookmark placement was clear, with the exception of the Kindergarten and Grade 1 group (33.3%).

Table 19 shows that most participants agreed or strongly agreed that the training materials were helpful except for the Kindergarten and Grade 1 group (66.6%). Table 20 indicates that all participants agreed or strongly agreed that the Bookmark Procedure was well described. As Table 21 demonstrates, participants in Grades 5 and 7 agreed or strongly agreed that the goals of the process were clear.

Table 17. Participants' agreement/disagreement with the statement, "I understood how to place my bookmarks."

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
K & 1	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
2	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
5	2	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
7	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Table 18. Participants' agreement/disagreement with the statement, "The training on Bookmark placement made the task clear to me."

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	7.1%	7.1%	35.7%	50.0%	85.7%
K & 1	3	0.0%	33.3%	33.3%	0.0%	33.3%	33.3%
2	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
5	2	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
7	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
11	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%

Table 19. Participants' agreement/disagreement with the statement, “The training materials were helpful.”

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	7.1%	57.1%	35.7%	92.8%
K & 1	3	0.0%	0.0%	33.3%	33.3%	33.3%	66.6%
2	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%

Table 20. Participants' agreement/disagreement with the statement, “The Bookmark Procedure was well described.”

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	0.0%	42.9%	57.1%	100.0%
K & 1	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
2	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
5	2	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%

Table 21. Participants' agreement/disagreement with the statement, “The goals for the Bookmark Procedure were clear.”

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	21.4%	28.6%	50.0%	78.6%
K & 1	3	0.0%	0.0%	33.3%	0.0%	66.7%	66.7%
2	3	0.0%	0.0%	33.3%	33.3%	33.3%	66.6%
5	2	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
7	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
11	3	0.0%	0.0%	33.3%	33.3%	33.3%	66.6%

Quality Control Procedures

The CTB Standard Setting Team adhered to many quality control procedures to foster the accuracy of the materials used and the results presented during workshop. Prior to the workshop, the CTB Standard Setting Team cross-checked the ordering of items in the ordered item booklets, the accuracy of the information in the item maps, and the accuracy of the Microsoft Excel macros and Bookmark Pro software used to generate results and impact data. All data were scanned on-site at the workshop. The CTB Standard Setting Team checked the reasonableness of the data presented to participants.

References

Beretvas, S.N. (2004). Comparison of Bookmark difficulty locations under different item response models. *Applied Psychological Measurement*, 28, 25-47.

Lewis, D.M., Mitzel, H.C., & Green, D.R. (1996). Standard Setting: A bookmark approach. In D.R. Green (Chair), *IRT-based standard-setting procedures utilizing behavioral anchoring*. Symposium conducted at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.

Lewis, D. M., Green, D. R., Mitzel, H. C., Baum, K., & Patz, R. J., (1998, April). The Bookmark Standard Setting Procedure: Methodology and Recent Implementations. Paper presented at the 1998 annual meeting of the National Council of Measurement in Education annual meeting, San Diego, CA.

Mitzel, H.C., Lewis, D.M., Patz, R.J., & Green, D.R. (2001). The Bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives*. (pp. 249–281). Mahwah, NJ: Lawrence Erlbaum Associates.

Wright, B.D., & Masters, G.N. (1982). *Rating Scale Analysis*. Chicago, IL: MESA Press.

SECTION C

Master Agenda

DRAFT



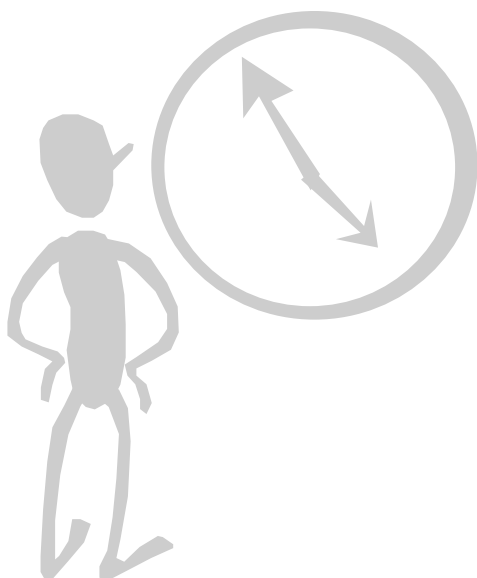
Master Agenda

Oregon Department of Education

Kindergarten and Grades 2, 5, 7, and 11
English Language Proficiency Assessments

Bookmark Standard Setting Workshop
November 5 - 6, 2007
Salem, Oregon

**Mc
Graw
Hill** **CTB
McGraw-Hill**



Welcome to the Standard Setting Workshop for the Oregon ELPA!

The Oregon Department of Education and CTB/McGraw-Hill thank you for your time and expertise during this important process.

Please use this agenda to orient yourself during the workshop. If you have any questions or concerns, please do not hesitate to contact a member of the CTB Standard Setting Team.

Monday, November 5
Welcome!

7:30 AM Participant Registration and Continental Breakfast

Please check in at the reception area to sign the nondisclosure agreement, get your nametag, and collect any other necessary information.

8:30 AM Opening Session and Bookmark Overview

The Oregon Department of Education welcomes participants to the standard setting and overviews the testing program. CTB introduces the Bookmark Standard Setting Procedure and discusses your role and responsibilities during the workshop.

9:30 AM Adjournment to Preassigned Tables

The Group Leader welcomes participants. After brief introductions, the Group Leader distributes secure materials. Secure materials are printed on colored paper.

- Table Leaders ensure that all participants at their tables write their names on each of their secure materials.

10:00 AM Target Student Discussion

Participants engage in structured discussions about the knowledge, skills, and abilities they expect to be demonstrated by students just entering each performance level.

11:00 AM Examine the Test Items

Participants examine the test items to see what students experience.

- Although some discussion about individual test items is normal, Table Leaders focus their participants away from prolonged debate and toward taking the test.
- Table Leaders encourage participants to use provided index cards to record comments about the test items.

12:00 PM Lunch

Monday, November 5 (cont.)

Study Items in Ordered Item Booklets

1:00 PM Discuss Each Item in the Ordered Item Booklet (OIB)

The Group Leader introduces this task by instructing participants to find the item map and OIB in their secure materials. The Group Leader leads the group in a review of each column on the item map and in an examination of several items in the OIB.

- Table Leaders facilitate a discussion among everyone at their tables about each of the items in the OIB. Start with the first item, and discuss each item in turn, focusing on what each item measures and what makes it harder than the previous items. All participants record these details on their item maps.
- Table Leaders assign a scribe to take a master set of notes for their table.
- Table Leaders remind participants to use the index cards, as necessary.
- Table Leaders ensure that each participant at their tables has a chance to speak.

4:15 PM Review Bookmark Placement and Round 1 Ratings

A member of the CTB Standard Setting Team reviews bookmark placement, explaining how bookmarks are placed and what bookmarks mean. After this brief presentation, a short check set is given and discussed. The Group Leader then directs all participants to place their Round 1 bookmarks. The Group Leader reminds participants that bookmark placement is an individual activity.

- See the handouts on “Bookmark Placement,” “Frequently Asked Questions,” and “Mastery” for more information.
- Table Leaders collect their participants’ rating forms as they complete them, ensuring that each participant has made a single, unambiguous rating for each bookmark.
- Table Leaders give their participants’ rating forms to the Group Leader.

4:45 PM Secure Materials Collection

The Group Leader facilitates collection of the secure materials from all participants. A listing of secure materials to be collected is displayed in the room.

- Table Leaders supervise the collection of secure materials at their tables. See the last page of this agenda, “Secure Materials Collection,” for more information.

4:55 PM Secure Materials Audit

The Group Leader directs the Table Leaders to audit the secure materials at one other table.

- Verify that each packet contains all the secure materials.
- Order materials numerically by packet number within each table.
- Verify that all signed-out packets are present.
- Stack materials at each table neatly into one pile with the table tent on top, under the top packet’s rubber band.
- Place the separate stacks on one table. Do not combine tables’ stacks.

5:00 PM Table Leader Debriefing

Table leaders discuss the events of the day and plans for the next day.

5:15 PM Table Leader Dismissal

Tuesday, November 6
Discussion & Bookmark Ratings

7:30 AM Continental Breakfast

8:30 AM Discussion of Round 1 as a Table

Table Leaders lead a discussion of the ratings made at their tables. Impact data are presented. Participants discuss the items between the lowest and highest ratings, explaining the rationale behind their ratings.

9:30 AM Round 2 Ratings

The Group Leader directs all participants to place their Round 2 bookmarks. The Group Leader reminds participants that bookmark placement is an individual activity.

- Table Leaders collect their participants' rating forms as they complete them.
- Table Leaders give their participants' rating forms to the Group Leader.

10:30 AM Discussion of Round 2 as a Large Group

A member of the CTB Standard Setting Team presents a summary of the voting from each table to the entire group. Impact data for each table are presented. Then, the Group Leader leads a discussion with the entire group about the performance standards in each grade.

11:30 AM Round 3 Ratings

The Group Leader directs all participants to place their Round 3 bookmarks. The Group Leader reminds participants that bookmark placement is an individual activity.

- Table Leaders collect their participants' rating forms as they complete them.
- Table Leaders give their participants' rating forms to the Group Leader.

12:00 PM Lunch

1:00 PM Articulation Discussion

A member of the CTB Standard Setting Team presents the group with a summary of the Round 3 recommendations. Participants are asked to review the cut scores for all grades, including the off-grades, and to evaluate how reasonable and consistent the performance standards are across grades.

2:00 PM Performance level description writing training

The Group Leader presents instructions for writing a first draft of the long performance level descriptions.

2:30 PM Performance level descriptions, first draft

- Your group will receive a listing of the items you will work with.
- Your group's descriptions should synthesize the knowledge, skills, and abilities necessary to respond successfully to each of the items assigned to each performance level.

3:30 PM Performance level descriptions, second draft

Each group presents its draft to the entire group and receives comments.

4:45 PM Secure materials collection

The Group Leader facilitates collection of the secure materials from all participants. A listing of secure materials to be collected is displayed in the room.

- Table Leaders supervise the collection of secure materials at their tables. See the last page of this agenda, “Secure Materials Collection,” for more information.

4:50 PM Secure Materials Audit

The Group Leader directs the Table Leaders to audit the secure materials at one other table.

- Verify that each packet contains all the secure materials.
- Order materials numerically by packet number within each table.
- Verify that all signed-out packets are present.
- Stack materials at each table neatly into one pile with the table tent on top, under the top packet’s rubber band.
- Place the separate stacks on one table. Do not combine tables’ stacks.

4:55 PM Participant Evaluation

Each participant completes a written evaluation of the ELPA standard setting.

5:00 PM Dismissal

Why do we do Secure Materials Collection?

A thorough collection of secure test materials protects both the reliability of the testing program and the substantial monetary investment in the assessment. A structured method of collection has been established to gather effectively all of the secure material at the workshop. Each day as you facilitate secure materials collection at your table, refer to this guide for instructions and suggestions.

During the collection, participants should place each secure item, one at a time, in a pile on the table in front of them. After the process, each participant will have a single stack of materials, each stacked in the same way as everyone else in the room. Please follow these steps to facilitate the process.

How do I do Secure Materials Collection?

1. Get the attention of all the participants at your table. Discourage any side conversations or inattention.
2. Using the list provided, call out each item, one at a time, and watch participants place that item on their stack. Discourage participants from moving ahead. Ensure that each participant has placed the item in their stack before moving on.
3. Proceed through the list until each piece of secure material has been collected. Direct participants to place a rubber band around their stack when completed.
4. If any participants wish to leave additional items with their materials, encourage them to place it beneath their stack, inside the rubber band.
5. Table Leaders will audit the secure materials at one other table.
6. Once you have supervised the collection of secure materials and are satisfied that all items have been collected, inform the Group Leader.
7. The collected materials are stored overnight and will be available in the morning.


What should I expect from Secure Materials Collection?

Generally, secure materials collection goes smoothly. If you have any questions about the collection process, or if you have a concern about test security at the standard setting workshop, please contact your Group Leader or a member of the CTB Standard Setting Team.

SECTION D

Training Overheads

DRAFT



Setting the Standard Oregon ELPA

Bookmark Standard Setting
Training
November 5, 2007

CTB/McGraw-Hill | QUALITY ASSESSMENT SINCE 1926

CTB Standard Setting Team

- Rick Mercado
- Christy Schneider
- Dorothy Tele'a
- Adele Brandstrom
- Lorena Houston
- Deborah Busch
- Sandra Snell
- Nadia Greer
- Agneta Lenberg

CTB/McGraw-Hill

What is standard setting?

- A process that lets experts make judgments about the English-language skills that the *Intermediate* student should know.
 - Also the *Beginning*, *Early Intermediate*, *Early Advanced*, and *Advanced* students.

CTB/McGraw-Hill

What is standard setting?

- How much does a student need to know to be classified in a given performance level for the ELPA?

CTB/McGraw-Hill

Why establish cut scores?

- English-language Proficiency (ELP) standards define what students are tested on.
 - These are things students should be able to do.
 - Oregon has ELP standards designed to supplement the ELA standards.
- Performance standards define what students should be able to do in each performance level.
 - You will actively discuss your expectations of the target student for each performance level of the ELPA.

CTB/McGraw-Hill

Performance levels

- Specify the English-language skills a student needs to know in order to be classified as *Beginning*, *Early Intermediate*, *Intermediate*, *Early Advanced*, and *Advanced* in relation to the ELP standards.

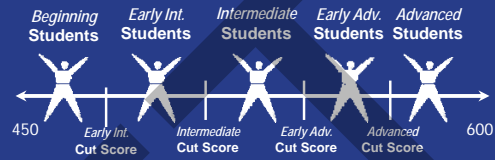
CTB/McGraw-Hill

How We Set Our Standards

- ~~Percentages~~
 - Arbitrary
 - Test-specific
 - Do not consider skills
- English-language Skills
 - Uses pre-established ELP standards
 - Considers educational objectives
- Bookmark Standard Setting Procedure

Purpose of the Standard Setting

- Allows cut scores to be set on the test scale
- The test scale represents the ability of students



Purpose of the Standard Setting

- You will set four cut scores on the test.
- You will set cut scores that reflect the English-language skills of students in each performance level.
- Decisions will be based on the Oregon ELP standards.
- One cut score for each
 - Early Intermediate
 - Intermediate
 - Early Advanced
 - Advanced

Bookmark Standard Setting

- Item-centered method
- Content-based decisions

Committee Roles

- Group Leaders
- Table Leaders
- Participants
- ODE
- CTB



Committee Roles

- Group Leaders
 - Facilitators
 - Participants stay focused on task
 - Participants interact with their own group
 - Participants finish in a timely manner
 - Facilitates discussion
 - Materials collection
 - Secure materials

Standard Setting Committee



Committee Roles

- Table Leaders
 - Lead discussion at the table
 - Standard setters
- Participants
 - Standard setters

Standard Setting Committee



CTB McGraw-Hill

Workshop Overview

- Round 1
 - Review test items
 - Make ratings without discussion
- Round 2
 - Discuss ratings in a small group
- Round 3
 - Discuss ratings in a large group
- Articulation Discussion
- Review Performance Level Descriptors

CTB McGraw-Hill

Ordered Item Booklets

- One item per page
- Easiest item first, hardest item last
- Items ascend by difficulty

CTB McGraw-Hill

Item Map

Order of difficulty (easy to hard)	Item No.	Item No.	Domain	Language Skills Assessed	Item Type	Score	What does this item measure? That is, what do you know about a student who can respond successfully to this item?	Why is this item more difficult than the preceding item?
1	543	29	Reading	determining main idea	MC	A		NA
2	563	26	Reading	making inferences, understanding literary terms	MC	C		
3	570	2	Reading	decoding phonemes to determine words that rhyme	MC	B		
4	583	23	Reading	using dictionary skills to understand words with multiple meanings	MC	B		
5	586	34	Reading	inferring author's intent, understanding figurative language	MC	B		
6	587	32	Reading	understanding vocabulary and complex answer choices	MC	D		
7	604	35	Reading	understanding literary devices, e.g. sound words and figurative language, e.g. foreshadowing	MC	B		
8	620	18	Reading	locating information presented in a chart, discerning details, understanding vocabulary with multiple meanings	MC	D		
9	631	25	Reading	understanding high-level vocabulary	MC	B		

CTB McGraw-Hill

Item Map

Order of difficulty (easy to hard)	Item No.	Item No.	Domain	Language Skills Assessed	Item Type	Score	What does this item measure? That is, what do you know about a student who can respond successfully to this item?	Why is this item more difficult than the preceding item?
1	543	29	Reading	determining main idea	MC	A		NA
2	563	26	Reading	making inferences, understanding literary terms	MC	C		
3	570	2	Reading	decoding phonemes to determine words that rhyme	MC	B		
4	583	23	Reading	using dictionary skills to understand words with multiple meanings	MC	B		
5	586	34	Reading	inferring author's intent, understanding figurative language	MC	B		
6	587	32	Reading	understanding vocabulary and complex answer choices	MC	D		
7	604	35	Reading	understanding literary devices, e.g. sound words and figurative language, e.g. foreshadowing	MC	B		
8	620	18	Reading	locating information presented in a chart, discerning details, understanding vocabulary with multiple meanings	MC	D		
9	631	25	Reading	understanding high-level vocabulary	MC	B		

CTB McGraw-Hill

Constructed-response Items

- Each constructed-response item will appear once in the OIB for each point.
 - Four times: twice for Grammatical, and twice for Illocutionary.
- Treat each score point like its own item.
- Consider the English-language skills demonstrated by students for each point.

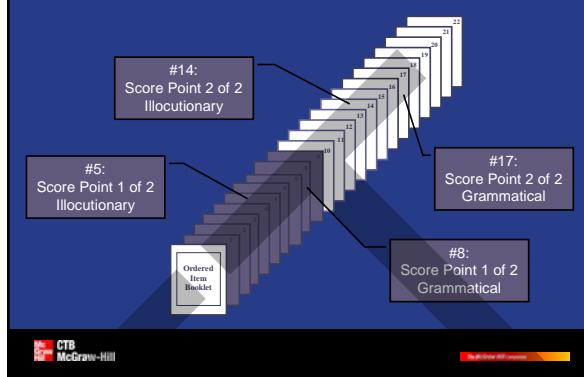
CTB McGraw-Hill

Sample Rubric

Score	Grammatical Competence	Score	Illocutionary Competence
2	Complete sentence with correct use of present progressive verbs; may use adverb(s). Errors do not interfere with meaning.	2	Clearly and completely describes the action(s) represented in the picture.
1	Complete sentence using a tense other than present progressive; no adverbs. Errors may interfere with meaning.	1	Describes action(s) represented in the picture, but the description of the action is incomplete: OR describes the picture.
0	Non-English, isolated word(s) or phrase(s), non-responsive, unintelligible, or repeats prompt.	0	Isolated word(s) or phrases unrelated to the stimulus.

CTB
McGraw-Hill

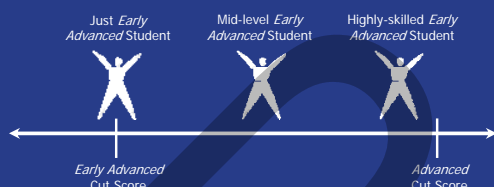
Sample OIB



CTB
McGraw-Hill

Target Student

- We want to describe the English-language skills held in *common* by *all* these students
 - These are the skills of the Just *Early Advanced* student



CTB
McGraw-Hill

Bookmark Placement

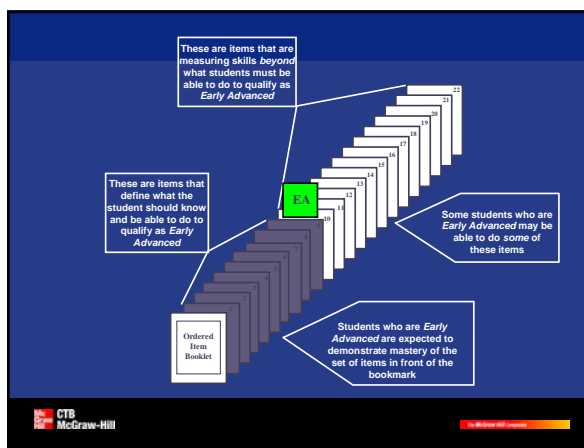
- Items preceding the Bookmark reflect content that all *Early Advanced* students should have mastery of
 - for MC items this means that *Early Advanced* students should most likely know the correct responses
 - for CR items this means that *Early Advanced* students should most likely earn the score points before the bookmark

CTB
McGraw-Hill

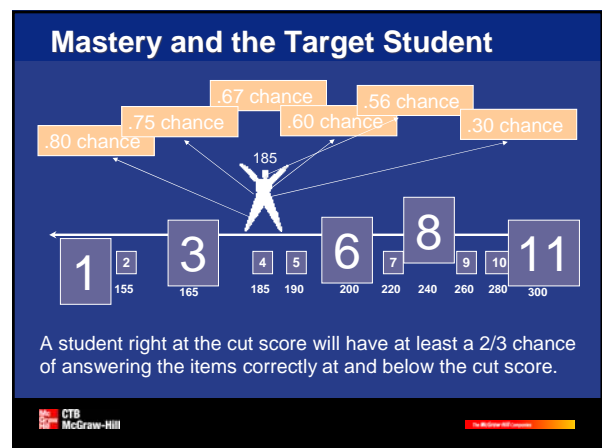
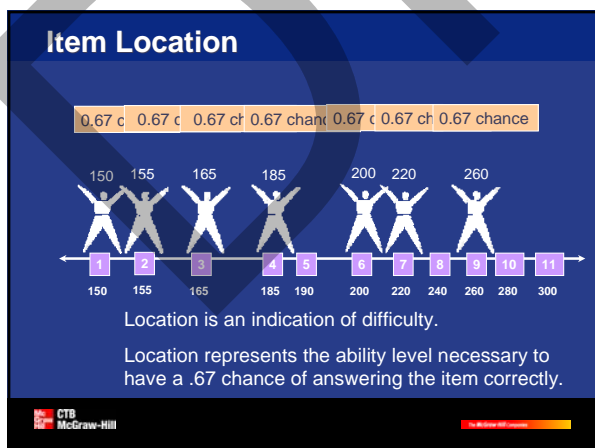
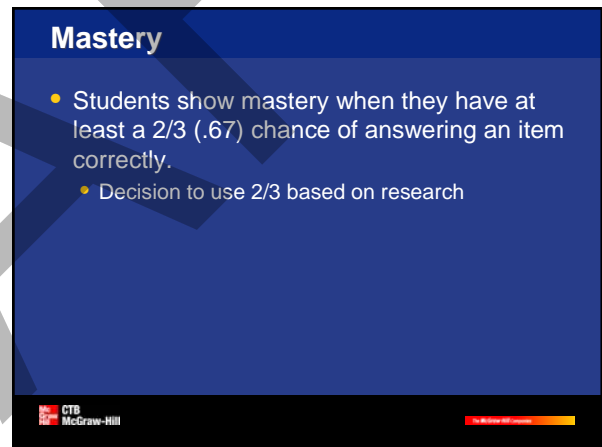
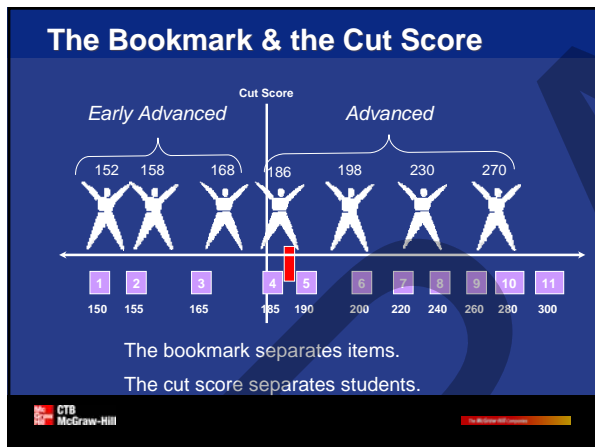
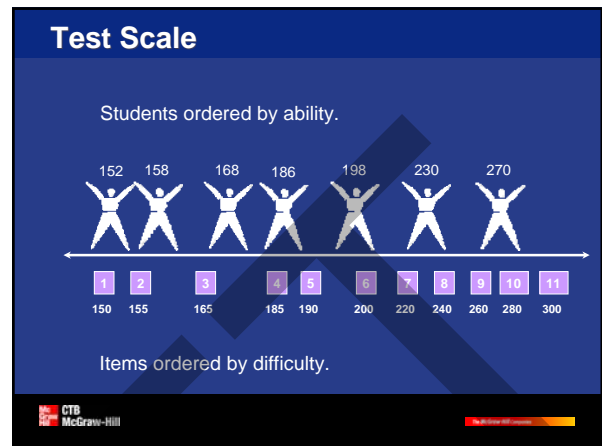
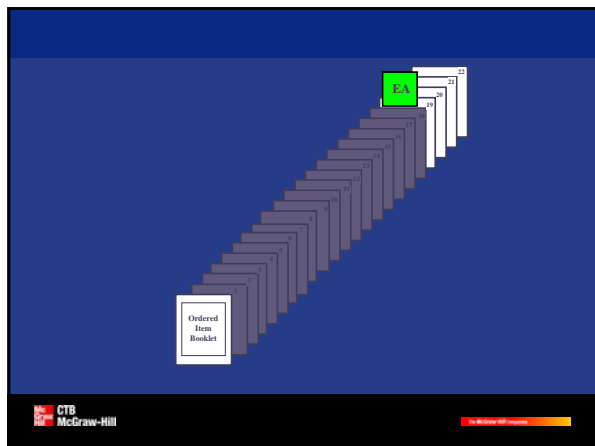
Bookmark Placement

- Place the bookmark at the first point...
- ...where you feel that a student who has mastery of the content in the items before the bookmark...
- ...has demonstrated sufficient skills...
- ...to infer that the student should be classified as *Early Advanced*.

CTB
McGraw-Hill



CTB
McGraw-Hill



Rating Form for ELPA

Print Name _____ November 2007 Oregon ELPA Standard Setting Rating Form

Grade	Round 1	Early Intermediate	Intermediate	Early Advanced	Advanced
O K					
O 1					
O 2					
O 5					
O 7					
O 11					
Content Area					
O ELPA					
Table					
O 1					

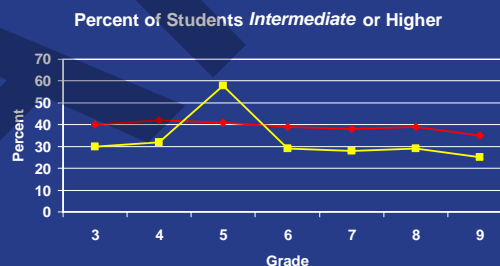
Sample Results

	Early Int. Bookmark	Intermediate Bookmark	Early Adv. Bookmark	Advanced Bookmark
Table 1	22	37	45	79
Median	22	37	45	79
Impact Data: Estimated percent of students in each performance level based on current large group median				
Beginning	Early Int.	Intermediate	Early Adv.	Advanced
0%	0%	0%	0%	0%

Articulation Discussion

- Performance standards are called *well-articulated* when the impact data associated with the cut scores form a cogent, reasonable pattern.
- After Round 3, the group will discuss the performance standards across the grades.

Articulation and Disarticulation



Agenda: This Morning

- Opening Session
- Discuss the Target Students
 - Table Activity
- Examine the test items
 - Individual Activity

Agenda: This Afternoon

- Study the Ordered Item Booklet
 - Table Activity
- Bookmark Training
- Round 1 Bookmark Placements
 - Individual Activity

Agenda: Tomorrow

- Round 2
 - Discuss Round 1 results in tables
 - Make new judgments individually
- Round 3
 - Discuss Round 2 results as a large group
 - Make new judgments individually
- Articulation Discussion
- Review performance level descriptors
- Evaluate the Standard Setting

Target Student Discussion

- The student who has *just* made it into a performance level
 - Just *Early Intermediate*, Just *Intermediate*, Just *Early Advanced*, and Just *Advanced* students
- Refer to Oregon ELP standards



Questions?

- Thank you for your participation!



Bookmark Review

Oregon

Standard Setting Training ELPA

CTB/McGraw-Hill | QUALITY ASSESSMENT SINCE 1926

Presentation Purpose

- Review the process for placing the bookmark
- Slides and content are intended to be repetitive to help you internalize the process

CTB/McGraw-Hill

Performance Levels

- Specify the knowledge, skills and abilities a student needs to know in order to be classified as *Early Intermediate*, *Intermediate*, *Early Advanced*, and *Advanced* in relation to the proficiency standards.

CTB/McGraw-Hill

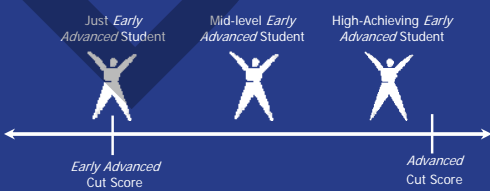
Purpose of the Standard Setting

- You will recommend four cut scores on the test scale.
- One cut score for
 - *Early Intermediate*
 - *Intermediate*
 - *Early Advanced*
 - *Advanced*
- Decisions will be based on Oregon ELP standards.

CTB/McGraw-Hill

Target Student

- We want to describe the skills held in *common* by *all* these students
 - These are the skills of the Just *Early Advanced* student

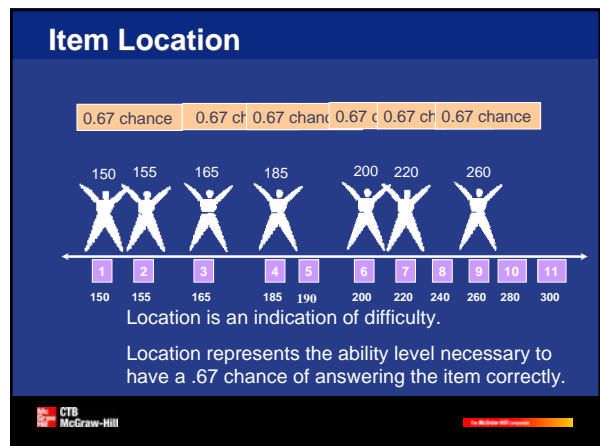
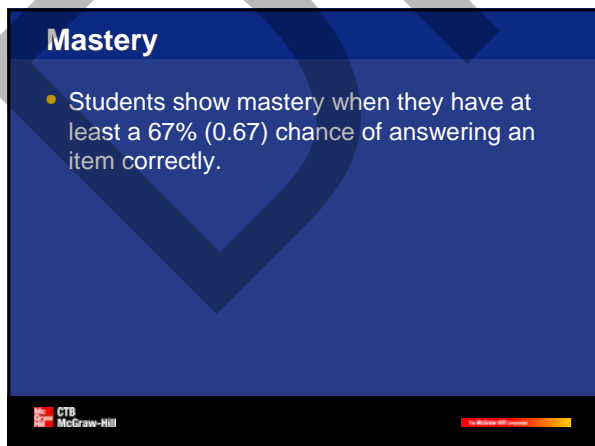
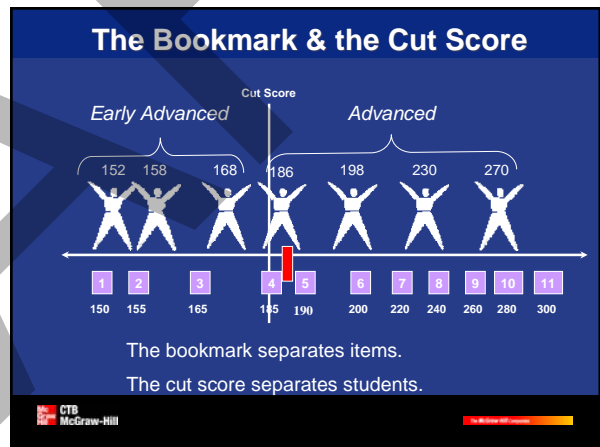
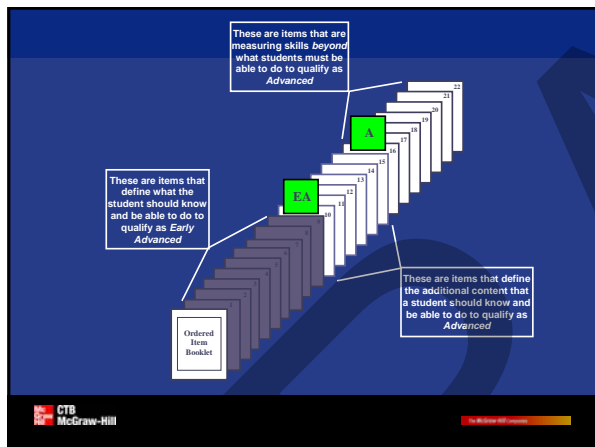
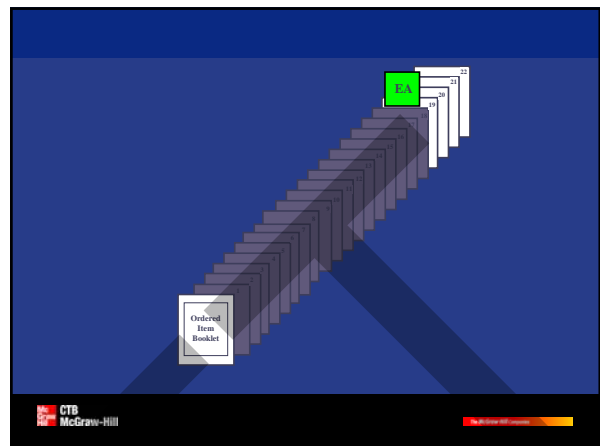
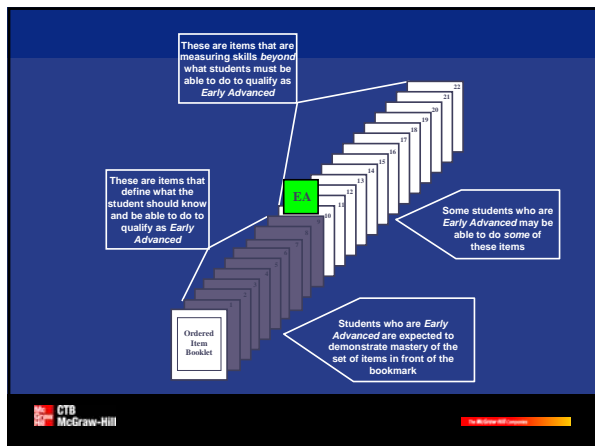


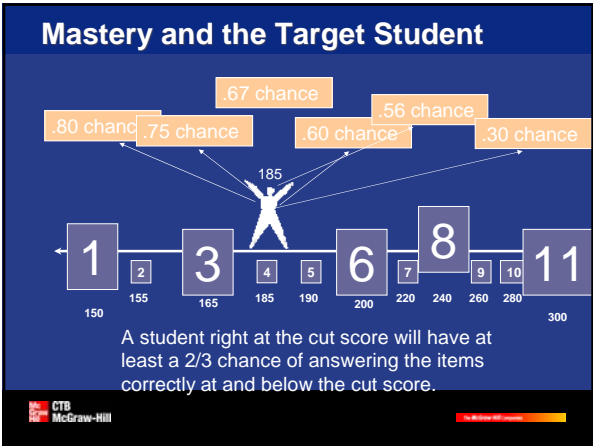
CTB/McGraw-Hill

Bookmark Placement cont...

- Place the bookmark at the first point...
- ...where you feel that a student who has mastery of the content in the items before the bookmark...
- ...has demonstrated sufficient skills...
- ...to infer that the student should be classified as *Early Advanced*.

CTB/McGraw-Hill





Rating Form

November 2007 Oregon ELPA Standard Setting Rating Form

Print Name _____

Grade
☐ K
☐ 1
☐ 2
☐ 5
☐ 7
☐ 11

Round
 1

Content Area
☐ ELPA

Table
☐ 1

	Early Intermediate	Intermediate	Early Advanced	Advanced
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				

CTB McGraw-Hill

SECTION E

Detailed Bookmark Judgments

DRAFT

Oregon ELPA November 2007 Grade K ELPA MC Extra
Round 1 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	5	10	17	25
1	2	4	10	16	33
1	3	8	16	22	33

Overall	Median	5	10	17	33
	Minimum	4	10	16	25
	Maximum	8	16	22	33
	SD	2.08	3.46	3.21	4.62

Oregon ELPA November 2007 Grade K ELPA MC Extra
Round 1 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	481	486	493	496
1	2	479	486	492	499
1	3	485	492	496	499

Overall	Median	481	486	493	499
	Minimum	479	486	492	496
	Maximum	485	492	496	499
	SD	3.06	3.46	2.08	1.73

Oregon ELPA November 2007 Grade K ELPA MC Extra
Round 1 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	5	10	17	33
Median	Overall	5	10	17	33
Minimum	1	4	10	16	25
Minimum	Overall	4	10	16	25
Maximum	1	8	16	22	33
Maximum	Overall	8	16	22	33
SD	1	2.08	3.46	3.21	4.62
SD	Overall	2.08	3.46	3.21	4.62

Overall	Median	5	10	17	33
	Minimum	4	10	16	25
	Maximum	8	16	22	33
	SD	2.08	3.46	3.21	4.62

Oregon ELPA November 2007 Grade K ELPA MC Extra
Round 1 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	481	486	493	499
Median	Overall	481	486	493	499
Minimum	1	479	486	492	496
Minimum	Overall	479	486	492	496
Maximum	1	485	492	496	499
Maximum	Overall	485	492	496	499
SD	1	3.06	3.46	2.08	1.73
SD	Overall	3.06	3.46	2.08	1.73

Overall	Median	481	486	493	499
	Minimum	479	486	492	496
	Maximum	485	492	496	499
	SD	3.06	3.46	2.08	1.73

Oregon ELPA November 2007 Grade K ELPA MC Extra Round 1 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	5	10	17	33
Overall	5	10	17	33

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	12.3	19.6	33.3	19.4	15.4

Oregon ELPA November 2007 Grade K ELPA MC Extra
Round 2 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	4	9	21	31
1	2	4	10	20	33
1	3	7	16	25	37

Overall	Median	4	10	21	33
	Minimum	4	9	20	31
	Maximum	7	16	25	37
	SD	1.73	3.79	2.65	3.06

Oregon ELPA November 2007 Grade K ELPA MC Extra
Round 2 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	479	485	495	498
1	2	479	486	495	499
1	3	484	492	496	500

Overall	Median	479	486	495	499
	Minimum	479	485	495	498
	Maximum	484	492	496	500
	SD	2.89	3.79	0.58	1.00

Oregon ELPA November 2007 Grade K ELPA MC Extra
Round 2 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	4	10	21	33
Median	Overall	4	10	21	33
Minimum	1	4	9	20	31
Minimum	Overall	4	9	20	31
Maximum	1	7	16	25	37
Maximum	Overall	7	16	25	37
SD	1	1.73	3.79	2.65	3.06
SD	Overall	1.73	3.79	2.65	3.06

Overall	Median	4	10	21	33
	Minimum	4	9	20	31
	Maximum	7	16	25	37
	SD	1.73	3.79	2.65	3.06

Oregon ELPA November 2007 Grade K ELPA MC Extra
Round 2 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	479	486	495	499
Median	Overall	479	486	495	499
Minimum	1	479	485	495	498
Minimum	Overall	479	485	495	498
Maximum	1	484	492	496	500
Maximum	Overall	484	492	496	500
SD	1	2.89	3.79	0.58	1.00
SD	Overall	2.89	3.79	0.58	1.00

Overall	Median	479	486	495	499
	Minimum	479	485	495	498
	Maximum	484	492	496	500
	SD	2.89	3.79	0.58	1.00

Oregon ELPA November 2007 Grade K ELPA MC Extra Round 2 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	4	10	21	33
Overall	4	10	21	33

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	6.2	25.7	41.5	11.2	15.4

Oregon ELPA November 2007 Grade K ELPA MC Extra
Round 3 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	4	14	21	40
1	2	4	14	21	42
1	3	4	15	23	39

Overall	Median	4	14	21	40
	Minimum	4	14	21	39
	Maximum	4	15	23	42
	SD	0.00	0.58	1.15	1.53

Oregon ELPA November 2007 Grade K ELPA MC Extra
Round 3 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	479	489	495	500
1	2	479	489	495	501
1	3	479	491	496	500

Overall	Median	479	489	495	500
	Minimum	479	489	495	500
	Maximum	479	491	496	501
	SD	0.00	1.15	0.58	0.58

Oregon ELPA November 2007 Grade K ELPA MC Extra
Round 3 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	4	14	21	40
Median	Overall	4	14	21	40
Minimum	1	4	14	21	39
Minimum	Overall	4	14	21	39
Maximum	1	4	15	23	42
Maximum	Overall	4	15	23	42
SD	1	0.00	0.58	1.15	1.53
SD	Overall	0.00	0.58	1.15	1.53

Overall	Median	4	14	21	40
	Minimum	4	14	21	39
	Maximum	4	15	23	42
	SD	0.00	0.58	1.15	1.53

Oregon ELPA November 2007 Grade K ELPA MC Extra
Round 3 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	479	489	495	500
Median	Overall	479	489	495	500
Minimum	1	479	489	495	500
Minimum	Overall	479	489	495	500
Maximum	1	479	491	496	501
Maximum	Overall	479	491	496	501
SD	1	0.00	1.15	0.58	0.58
SD	Overall	0.00	1.15	0.58	0.58

Overall	Median	479	489	495	500
	Minimum	479	489	495	500
	Maximum	479	491	496	501
	SD	0.00	1.15	0.58	0.58

Oregon ELPA November 2007 Grade K ELPA MC Extra Round 3 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	4	14	21	40
Overall	4	14	21	40

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	6.2	39.9	27.3	12.9	13.7

Oregon ELPA November 2007 Grade 1 ELPA MC Extra
Round 1 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	17	32	39	58
1	2	17	33	45	53
1	3	16	31	39	64

Overall	Median	17	32	39	58
	Minimum	16	31	39	53
	Maximum	17	33	45	64
	SD	0.58	1.00	3.46	5.51

Oregon ELPA November 2007 Grade 1 ELPA MC Extra
Round 1 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	493	498	500	505
1	2	493	499	501	503
1	3	492	498	500	506

Overall	Median	493	498	500	505
	Minimum	492	498	500	503
	Maximum	493	499	501	506
	SD	0.58	0.58	0.58	1.53

Oregon ELPA November 2007 Grade 1 ELPA MC Extra
Round 1 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	17	32	39	58
Median	Overall	17	32	39	58
Minimum	1	16	31	39	53
Minimum	Overall	16	31	39	53
Maximum	1	17	33	45	64
Maximum	Overall	17	33	45	64
SD	1	0.58	1.00	3.46	5.51
SD	Overall	0.58	1.00	3.46	5.51

Overall	Median	17	32	39	58
	Minimum	16	31	39	53
	Maximum	17	33	45	64
	SD	0.58	1.00	3.46	5.51

Oregon ELPA November 2007 Grade 1 ELPA MC Extra
Round 1 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	493	498	500	505
Median	Overall	493	498	500	505
Minimum	1	492	498	500	503
Minimum	Overall	492	498	500	503
Maximum	1	493	499	501	506
Maximum	Overall	493	499	501	506
SD	1	0.58	0.58	0.58	1.53
SD	Overall	0.58	0.58	0.58	1.53

Overall	Median	493	498	500	505
	Minimum	492	498	500	503
	Maximum	493	499	501	506
	SD	0.58	0.58	0.58	1.53

Oregon ELPA November 2007 Grade 1 ELPA MC Extra Round 1 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	17	32	39	58
Overall	17	32	39	58

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	16.3	11.5	5.7	16.0	50.4

Oregon ELPA November 2007 Grade 1 ELPA MC Extra
Round 2 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	14	32	49	79
1	2	14	33	50	76
1	3	14	36	50	76

Overall	Median	14	33	50	76
	Minimum	14	32	49	76
	Maximum	14	36	50	79
	SD	0.00	2.08	0.58	1.73

Oregon ELPA November 2007 Grade 1 ELPA MC Extra
Round 2 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	489	498	503	518
1	2	489	499	503	515
1	3	489	499	503	515

Overall	Median	489	499	503	515
	Minimum	489	498	503	515
	Maximum	489	499	503	518
	SD	0.00	0.58	0.00	1.73

Oregon ELPA November 2007 Grade 1 ELPA MC Extra
Round 2 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	14	33	50	76
Median	Overall	14	33	50	76
Minimum	1	14	32	49	76
Minimum	Overall	14	32	49	76
Maximum	1	14	36	50	79
Maximum	Overall	14	36	50	79
SD	1	0.00	2.08	0.58	1.73
SD	Overall	0.00	2.08	0.58	1.73

Overall	Median	14	33	50	76
	Minimum	14	32	49	76
	Maximum	14	36	50	79
	SD	0.00	2.08	0.58	1.73

Oregon ELPA November 2007 Grade 1 ELPA MC Extra
Round 2 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	489	499	503	515
Median	Overall	489	499	503	515
Minimum	1	489	498	503	515
Minimum	Overall	489	498	503	515
Maximum	1	489	499	503	518
Maximum	Overall	489	499	503	518
SD	1	0.00	0.58	0.00	1.73
SD	Overall	0.00	0.58	0.00	1.73

Overall	Median	489	499	503	515
	Minimum	489	498	503	515
	Maximum	489	499	503	518
	SD	0.00	0.58	0.00	1.73

Oregon ELPA November 2007 Grade 1 ELPA MC Extra Round 2 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	14	33	50	76
Overall	14	33	50	76

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	8.8	22.3	11.5	37.0	20.3

Oregon ELPA November 2007 Grade 1 ELPA MC Extra
Round 3 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	16	32	64	81
1	2	14	33	64	80
1	3	16	38	62	81

Overall	Median	16	33	64	81
	Minimum	14	32	62	80
	Maximum	16	38	64	81
	SD	1.15	3.21	1.15	0.58

Oregon ELPA November 2007 Grade 1 ELPA MC Extra
Round 3 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	492	498	506	523
1	2	489	499	506	519
1	3	492	500	506	523

Overall	Median	492	499	506	523
	Minimum	489	498	506	519
	Maximum	492	500	506	523
	SD	1.73	1.00	0.00	2.31

Oregon ELPA November 2007 Grade 1 ELPA MC Extra
Round 3 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	16	33	64	81
Median	Overall	16	33	64	81
Minimum	1	14	32	62	80
Minimum	Overall	14	32	62	80
Maximum	1	16	38	64	81
Maximum	Overall	16	38	64	81
SD	1	1.15	3.21	1.15	0.58
SD	Overall	1.15	3.21	1.15	0.58

Overall	Median	16	33	64	81
	Minimum	14	32	62	80
	Maximum	16	38	64	81
	SD	1.15	3.21	1.15	0.58

Oregon ELPA November 2007 Grade 1 ELPA MC Extra
Round 3 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	492	499	506	523
Median	Overall	492	499	506	523
Minimum	1	489	498	506	519
Minimum	Overall	489	498	506	519
Maximum	1	492	500	506	523
Maximum	Overall	492	500	506	523
SD	1	1.73	1.00	0.00	2.31
SD	Overall	1.73	1.00	0.00	2.31

Overall	Median	492	499	506	523
	Minimum	489	498	506	519
	Maximum	492	500	506	523
	SD	1.73	1.00	0.00	2.31

Oregon ELPA November 2007 Grade 1 ELPA MC Extra Round 3 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	16	33	64	81
Overall	16	33	64	81

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	14.0	17.1	21.3	41.5	6.0

Oregon ELPA November 2007 Grade 2 ELPA
Round 1 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	13	54	69	99
1	2	12	27	88	93
1	3	16	49	70	99

Overall	Median	13	49	70	99
	Minimum	12	27	69	93
	Maximum	16	54	88	99
	SD	2.08	14.36	10.69	3.46

Oregon ELPA November 2007 Grade 2 ELPA
Round 1 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	494	510	514	543
1	2	493	499	523	530
1	3	495	508	515	543

Overall	Median	494	508	515	543
	Minimum	493	499	514	530
	Maximum	495	510	523	543
	SD	1.00	5.86	4.93	7.51

Oregon ELPA November 2007 Grade 2 ELPA
Round 1 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	13	49	70	99
Median	Overall	13	49	70	99
Minimum	1	12	27	69	93
Minimum	Overall	12	27	69	93
Maximum	1	16	54	88	99
Maximum	Overall	16	54	88	99
SD	1	2.08	14.36	10.69	3.46
SD	Overall	2.08	14.36	10.69	3.46

Overall	Median	13	49	70	99
	Minimum	12	27	69	93
	Maximum	16	54	88	99
	SD	2.08	14.36	10.69	3.46

Oregon ELPA November 2007 Grade 2 ELPA
Round 1 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	494	508	515	543
Median	Overall	494	508	515	543
Minimum	1	493	499	514	530
Minimum	Overall	493	499	514	530
Maximum	1	495	510	523	543
Maximum	Overall	495	510	523	543
SD	1	1.00	5.86	4.93	7.51
SD	Overall	1.00	5.86	4.93	7.51

Overall	Median	494	508	515	543
	Minimum	493	499	514	530
	Maximum	495	510	523	543
	SD	1.00	5.86	4.93	7.51

Oregon ELPA November 2007 Grade 2 ELPA Round 1 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	13	49	70	99
Overall	13	49	70	99

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	11.6	41.0	22.4	25.0	0.0

Oregon ELPA November 2007 Grade 2 ELPA
Round 2 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	16	49	69	96
1	2	12	27	88	98
1	3	16	49	69	97

Overall	Median	16	49	69	97
	Minimum	12	27	69	96
	Maximum	16	49	88	98
	SD	2.31	12.70	10.97	1.00

Oregon ELPA November 2007 Grade 2 ELPA
Round 2 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	495	508	514	534
1	2	493	499	523	540
1	3	495	508	514	535

Overall	Median	495	508	514	535
	Minimum	493	499	514	534
	Maximum	495	508	523	540
	SD	1.15	5.20	5.20	3.21

Oregon ELPA November 2007 Grade 2 ELPA
Round 2 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	16	49	69	97
Median	Overall	16	49	69	97
Minimum	1	12	27	69	96
Minimum	Overall	12	27	69	96
Maximum	1	16	49	88	98
Maximum	Overall	16	49	88	98
SD	1	2.31	12.70	10.97	1.00
SD	Overall	2.31	12.70	10.97	1.00

Overall	Median	16	49	69	97
	Minimum	12	27	69	96
	Maximum	16	49	88	98
	SD	2.31	12.70	10.97	1.00

Oregon ELPA November 2007 Grade 2 ELPA
Round 2 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	495	508	514	535
Median	Overall	495	508	514	535
Minimum	1	493	499	514	534
Minimum	Overall	493	499	514	534
Maximum	1	495	508	523	540
Maximum	Overall	495	508	523	540
SD	1	1.15	5.20	5.20	3.21
SD	Overall	1.15	5.20	5.20	3.21

Overall	Median	495	508	514	535
	Minimum	493	499	514	534
	Maximum	495	508	523	540
	SD	1.15	5.20	5.20	3.21

Oregon ELPA November 2007 Grade 2 ELPA Round 2 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	16	49	69	97
Overall	16	49	69	97

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	13.9	38.7	19.9	27.1	0.4

Oregon ELPA November 2007 Grade 2 ELPA
Round 3 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	16	49	69	89
1	2	12	58	88	90
1	3	16	49	69	89

Overall	Median	16	49	69	89
	Minimum	12	49	69	89
	Maximum	16	58	88	90
	SD	2.31	5.20	10.97	0.58

Oregon ELPA November 2007 Grade 2 ELPA
Round 3 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	495	508	514	523
1	2	493	511	523	527
1	3	495	508	514	523

Overall	Median	495	508	514	523
	Minimum	493	508	514	523
	Maximum	495	511	523	527
	SD	1.15	1.73	5.20	2.31

Oregon ELPA November 2007 Grade 2 ELPA
Round 3 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	16	49	69	89
Median	Overall	16	49	69	89
Minimum	1	12	49	69	89
Minimum	Overall	12	49	69	89
Maximum	1	16	58	88	90
Maximum	Overall	16	58	88	90
SD	1	2.31	5.20	10.97	0.58
SD	Overall	2.31	5.20	10.97	0.58

Overall	Median	16	49	69	89
	Minimum	12	49	69	89
	Maximum	16	58	88	90
	SD	2.31	5.20	10.97	0.58

Oregon ELPA November 2007 Grade 2 ELPA
Round 3 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	495	508	514	523
Median	Overall	495	508	514	523
Minimum	1	493	508	514	523
Minimum	Overall	493	508	514	523
Maximum	1	495	511	523	527
Maximum	Overall	495	511	523	527
SD	1	1.15	1.73	5.20	2.31
SD	Overall	1.15	1.73	5.20	2.31

Overall	Median	495	508	514	523
	Minimum	493	508	514	523
	Maximum	495	511	523	527
	SD	1.15	1.73	5.20	2.31

Oregon ELPA November 2007 Grade 2 ELPA Round 3 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	16	49	69	89
Overall	16	49	69	89

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	13.9	38.7	19.9	20.9	6.6

Oregon ELPA November 2007 Grade 5 ELPA
Round 1 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	5	20	47	79
1	2	8	15	54	74

Overall	Median	6.5	17.5	50.5	76.5
	Minimum	5	15	47	74
	Maximum	8	20	54	79
	SD	2.12	3.54	4.95	3.54

Oregon ELPA November 2007 Grade 5 ELPA
Round 1 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	495	501	512	527
1	2	497	499	514	520

Overall	Median	495	500	513	523
	Minimum	495	499	512	520
	Maximum	497	501	514	527
	SD	1.41	1.41	1.41	4.95

Oregon ELPA November 2007 Grade 5 ELPA
Round 1 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	6.5	17.5	50.5	76.5
Median	Overall	6.5	17.5	50.5	76.5
Minimum	1	5	15	47	74
Minimum	Overall	5	15	47	74
Maximum	1	8	20	54	79
Maximum	Overall	8	20	54	79
SD	1	2.12	3.54	4.95	3.54
SD	Overall	2.12	3.54	4.95	3.54

Overall	Median	6.5	17.5	50.5	76.5
	Minimum	5	15	47	74
	Maximum	8	20	54	79
	SD	2.12	3.54	4.95	3.54

Oregon ELPA November 2007 Grade 5 ELPA
Round 1 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	495	500	513	523
Median	Overall	495	500	513	523
Minimum	1	495	499	512	520
Minimum	Overall	495	499	512	520
Maximum	1	497	501	514	527
Maximum	Overall	497	501	514	527
SD	1	1.41	1.41	1.41	4.95
SD	Overall	1.41	1.41	1.41	4.95

Overall	Median	495	500	513	523
	Minimum	495	499	512	520
	Maximum	497	501	514	527
	SD	1.41	1.41	1.41	4.95

Oregon ELPA November 2007 Grade 5 ELPA Round 1 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	6.5	17.5	50.5	76.5
Overall	6.5	17.5	50.5	76.5

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	5.1	4.8	28.3	45.6	16.2

Oregon ELPA November 2007 Grade 5 ELPA
Round 2 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	7	31	57	77
1	2	5	33	56	77

Overall	Median	6	32	56.5	77
	Minimum	5	31	56	77
	Maximum	7	33	57	77
	SD	1.41	1.41	0.71	0.00

Oregon ELPA November 2007 Grade 5 ELPA
Round 2 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	496	507	515	523
1	2	495	508	515	523

Overall	Median	495	508	515	523
	Minimum	495	507	515	523
	Maximum	496	508	515	523
	SD	0.71	0.71	0.00	0.00

Oregon ELPA November 2007 Grade 5 ELPA
Round 2 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	6	32	56.5	77
Median	Overall	6	32	56.5	77
Minimum	1	5	31	56	77
Minimum	Overall	5	31	56	77
Maximum	1	7	33	57	77
Maximum	Overall	7	33	57	77
SD	1	1.41	1.41	0.71	0.00
SD	Overall	1.41	1.41	0.71	0.00

Overall	Median	6	32	56.5	77
	Minimum	5	31	56	77
	Maximum	7	33	57	77
	SD	1.41	1.41	0.71	0.00

Oregon ELPA November 2007 Grade 5 ELPA
Round 2 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	495	508	515	523
Median	Overall	495	508	515	523
Minimum	1	495	507	515	523
Minimum	Overall	495	507	515	523
Maximum	1	496	508	515	523
Maximum	Overall	496	508	515	523
SD	1	0.71	0.71	0.00	0.00
SD	Overall	0.71	0.71	0.00	0.00

Overall	Median	495	508	515	523
	Minimum	495	507	515	523
	Maximum	496	508	515	523
	SD	0.71	0.71	0.00	0.00

Oregon ELPA November 2007 Grade 5 ELPA Round 2 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	6	32	56.5	77
Overall	6	32	56.5	77

Impact Data

	Beginning	Early Intermediate	Intermediate	Early Advanced	Advanced
Overall	5.1	18.0	23.7	37.0	16.2

Oregon ELPA November 2007 Grade 5 ELPA
Round 3 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	9	33	59	77
1	2	9	33	59	77

Overall	Median	9	33	59	77
	Minimum	9	33	59	77
	Maximum	9	33	59	77
	SD	0.00	0.00	0.00	0.00

Oregon ELPA November 2007 Grade 5 ELPA
Round 3 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	497	508	516	523
1	2	497	508	516	523

Overall	Median	497	508	516	523
	Minimum	497	508	516	523
	Maximum	497	508	516	523
	SD	0.00	0.00	0.00	0.00

Oregon ELPA November 2007 Grade 5 ELPA
Round 3 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	9	33	59	77
Median	Overall	9	33	59	77
Minimum	1	9	33	59	77
Minimum	Overall	9	33	59	77
Maximum	1	9	33	59	77
Maximum	Overall	9	33	59	77
SD	1	0.00	0.00	0.00	0.00
SD	Overall	0.00	0.00	0.00	0.00

Overall	Median	9	33	59	77
	Minimum	9	33	59	77
	Maximum	9	33	59	77
	SD	0.00	0.00	0.00	0.00

Oregon ELPA November 2007 Grade 5 ELPA
Round 3 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	497	508	516	523
Median	Overall	497	508	516	523
Minimum	1	497	508	516	523
Minimum	Overall	497	508	516	523
Maximum	1	497	508	516	523
Maximum	Overall	497	508	516	523
SD	1	0.00	0.00	0.00	0.00
SD	Overall	0.00	0.00	0.00	0.00

Overall	Median	497	508	516	523
	Minimum	497	508	516	523
	Maximum	497	508	516	523
	SD	0.00	0.00	0.00	0.00

Oregon ELPA November 2007 Grade 5 ELPA Round 3 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	9	33	59	77
Overall	9	33	59	77

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	6.9	16.3	28.5	32.2	16.1

Oregon ELPA November 2007 Grade 7 ELPA
Round 1 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	6	15	53	73
1	2	11	21	47	75
1	3	12	24	57	80

Overall	Median	11	21	53	75
	Minimum	6	15	47	73
	Maximum	12	24	57	80
	SD	3.21	4.58	5.03	3.61

Oregon ELPA November 2007 Grade 7 ELPA
Round 1 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	490	498	511	519
1	2	496	499	509	523
1	3	497	501	513	527

Overall	Median	496	499	511	523
	Minimum	490	498	509	519
	Maximum	497	501	513	527
	SD	3.79	1.53	2.00	4.00

Oregon ELPA November 2007 Grade 7 ELPA
Round 1 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	11	21	53	75
Median	Overall	11	21	53	75
Minimum	1	6	15	47	73
Minimum	Overall	6	15	47	73
Maximum	1	12	24	57	80
Maximum	Overall	12	24	57	80
SD	1	3.21	4.58	5.03	3.61
SD	Overall	3.21	4.58	5.03	3.61

Overall	Median	11	21	53	75
	Minimum	6	15	47	73
	Maximum	12	24	57	80
	SD	3.21	4.58	5.03	3.61

Oregon ELPA November 2007 Grade 7 ELPA
Round 1 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	496	499	511	523
Median	Overall	496	499	511	523
Minimum	1	490	498	509	519
Minimum	Overall	490	498	509	519
Maximum	1	497	501	513	527
Maximum	Overall	497	501	513	527
SD	1	3.79	1.53	2.00	4.00
SD	Overall	3.79	1.53	2.00	4.00

Overall	Median	496	499	511	523
	Minimum	490	498	509	519
	Maximum	497	501	513	527
	SD	3.79	1.53	2.00	4.00

Oregon ELPA November 2007 Grade 7 ELPA Round 1 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	11	21	53	75
Overall	11	21	53	75

Impact Data

	Beginning	Early Intermediate	Intermediate	Early Advanced	Advanced
Overall	6.6	2.4	21.0	52.2	17.8

Oregon ELPA November 2007 Grade 7 ELPA
Round 2 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	12	36	66	79
1	2	12	34	61	77
1	3	12	34	61	79

Overall	Median	12	34	61	79
	Minimum	12	34	61	77
	Maximum	12	36	66	79
	SD	0.00	1.15	2.89	1.15

Oregon ELPA November 2007 Grade 7 ELPA
Round 2 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	497	506	518	526
1	2	497	505	516	524
1	3	497	505	516	526

Overall	Median	497	505	516	526
	Minimum	497	505	516	524
	Maximum	497	506	518	526
	SD	0.00	0.58	1.15	1.15

Oregon ELPA November 2007 Grade 7 ELPA
Round 2 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	12	34	61	79
Median	Overall	12	34	61	79
Minimum	1	12	34	61	77
Minimum	Overall	12	34	61	77
Maximum	1	12	36	66	79
Maximum	Overall	12	36	66	79
SD	1	0.00	1.15	2.89	1.15
SD	Overall	0.00	1.15	2.89	1.15

Overall	Median	12	34	61	79
	Minimum	12	34	61	77
	Maximum	12	36	66	79
	SD	0.00	1.15	2.89	1.15

Oregon ELPA November 2007 Grade 7 ELPA
Round 2 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	497	505	516	526
Median	Overall	497	505	516	526
Minimum	1	497	505	516	524
Minimum	Overall	497	505	516	524
Maximum	1	497	506	518	526
Maximum	Overall	497	506	518	526
SD	1	0.00	0.58	1.15	1.15
SD	Overall	0.00	0.58	1.15	1.15

Overall	Median	497	505	516	526
	Minimum	497	505	516	524
	Maximum	497	506	518	526
	SD	0.00	0.58	1.15	1.15

Oregon ELPA November 2007 Grade 7 ELPA Round 2 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	12	34	61	79
Overall	12	34	61	79

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	7.5	8.2	34.7	41.0	8.6

Oregon ELPA November 2007 Grade 7 ELPA
Round 3 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	12	40	67	76
1	2	12	39	67	77
1	3	12	39	70	77

Overall	Median	12	39	67	77
	Minimum	12	39	67	76
	Maximum	12	40	70	77
	SD	0.00	0.58	1.73	0.58

Oregon ELPA November 2007 Grade 7 ELPA
Round 3 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	497	507	518	523
1	2	497	507	518	524
1	3	497	507	519	524

Overall	Median	497	507	518	524
	Minimum	497	507	518	523
	Maximum	497	507	519	524
	SD	0.00	0.00	0.58	0.58

Oregon ELPA November 2007 Grade 7 ELPA
Round 3 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	12	39	67	77
Median	Overall	12	39	67	77
Minimum	1	12	39	67	76
Minimum	Overall	12	39	67	76
Maximum	1	12	40	70	77
Maximum	Overall	12	40	70	77
SD	1	0.00	0.58	1.73	0.58
SD	Overall	0.00	0.58	1.73	0.58

Overall	Median	12	39	67	77
	Minimum	12	39	67	76
	Maximum	12	40	70	77
	SD	0.00	0.58	1.73	0.58

Oregon ELPA November 2007 Grade 7 ELPA
Round 3 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	497	507	518	524
Median	Overall	497	507	518	524
Minimum	1	497	507	518	523
Minimum	Overall	497	507	518	523
Maximum	1	497	507	519	524
Maximum	Overall	497	507	519	524
SD	1	0.00	0.00	0.58	0.58
SD	Overall	0.00	0.00	0.58	0.58

Overall	Median	497	507	518	524
	Minimum	497	507	518	523
	Maximum	497	507	519	524
	SD	0.00	0.00	0.58	0.58

Oregon ELPA November 2007 Grade 7 ELPA Round 3 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	12	39	67	77
Overall	12	39	67	77

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	7.5	11.7	40.3	27.1	13.4

Oregon ELPA November 2007 Grade 11 ELPA
Round 1 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	8	21	42	74
1	2	2	8	37	74
1	3	2	21	54	71

Overall	Median	2	21	42	74
	Minimum	2	8	37	71
	Maximum	8	21	54	74
	SD	3.46	7.51	8.74	1.73

Oregon ELPA November 2007 Grade 11 ELPA
Round 1 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	492	501	508	522
1	2	476	492	506	522
1	3	476	501	512	520

Overall	Median	476	501	508	522
	Minimum	476	492	506	520
	Maximum	492	501	512	522
	SD	9.24	5.20	3.06	1.15

Oregon ELPA November 2007 Grade 11 ELPA
Round 1 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	2	21	42	74
Median	Overall	2	21	42	74
Minimum	1	2	8	37	71
Minimum	Overall	2	8	37	71
Maximum	1	8	21	54	74
Maximum	Overall	8	21	54	74
SD	1	3.46	7.51	8.74	1.73
SD	Overall	3.46	7.51	8.74	1.73

Overall	Median	2	21	42	74
	Minimum	2	8	37	71
	Maximum	8	21	54	74
	SD	3.46	7.51	8.74	1.73

Oregon ELPA November 2007 Grade 11 ELPA
Round 1 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	476	501	508	522
Median	Overall	476	501	508	522
Minimum	1	476	492	506	520
Minimum	Overall	476	492	506	520
Maximum	1	492	501	512	522
Maximum	Overall	492	501	512	522
SD	1	9.24	5.20	3.06	1.15
SD	Overall	9.24	5.20	3.06	1.15

Overall	Median	476	501	508	522
	Minimum	476	492	506	520
	Maximum	492	501	512	522
	SD	9.24	5.20	3.06	1.15

Oregon ELPA November 2007 Grade 11 ELPA Round 1 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	2	21	42	74
Overall	2	21	42	74

Impact Data

	Beginnin g	Early Intermedi ate	Intermedi ate	Early Advance d	Advance d
Overall	0.0	16.4	10.2	47.0	26.4

Oregon ELPA November 2007 Grade 11 ELPA
Round 2 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	12	24	56	77
1	2	12	21	56	80
1	3	9	21	53	75

Overall	Median	12	21	56	77
	Minimum	9	21	53	75
	Maximum	12	24	56	80
	SD	1.73	1.73	1.73	2.52

Oregon ELPA November 2007 Grade 11 ELPA
Round 2 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	497	502	513	524
1	2	497	501	513	528
1	3	494	501	512	523

Overall	Median	497	501	513	524
	Minimum	494	501	512	523
	Maximum	497	502	513	528
	SD	1.73	0.58	0.58	2.65

Oregon ELPA November 2007 Grade 11 ELPA
Round 2 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	12	21	56	77
Median	Overall	12	21	56	77
Minimum	1	9	21	53	75
Minimum	Overall	9	21	53	75
Maximum	1	12	24	56	80
Maximum	Overall	12	24	56	80
SD	1	1.73	1.73	1.73	2.52
SD	Overall	1.73	1.73	1.73	2.52

Overall	Median	12	21	56	77
	Minimum	9	21	53	75
	Maximum	12	24	56	80
	SD	1.73	1.73	1.73	2.52

Oregon ELPA November 2007 Grade 11 ELPA
Round 2 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	497	501	513	524
Median	Overall	497	501	513	524
Minimum	1	494	501	512	523
Minimum	Overall	494	501	512	523
Maximum	1	497	502	513	528
Maximum	Overall	497	502	513	528
SD	1	1.73	0.58	0.58	2.65
SD	Overall	1.73	0.58	0.58	2.65

Overall	Median	497	501	513	524
	Minimum	494	501	512	523
	Maximum	497	502	513	528
	SD	1.73	0.58	0.58	2.65

Oregon ELPA November 2007 Grade 11 ELPA Round 2 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	12	21	56	77
Overall	12	21	56	77

Impact Data

	Beginning	Early Intermediate	Intermediate	Early Advanced	Advanced
Overall	11.7	4.7	23.7	42.3	17.6

Oregon ELPA November 2007 Grade 11 ELPA
Round 3 Bookmark Placements

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	9	24	58	80
1	2	9	10	58	80
1	3	9	21	58	80

Overall	Median	9	21	58	80
	Minimum	9	10	58	80
	Maximum	9	24	58	80
	SD	0.00	7.37	0.00	0.00

Oregon ELPA November 2007 Grade 11 ELPA
Round 3 Cut Scores

Table	Participant	Early Intermediate	Intermediate	Early Advanced	Advanced
1	1	494	502	515	528
1	2	494	495	515	528
1	3	494	501	515	528

Overall	Median	494	501	515	528
	Minimum	494	495	515	528
	Maximum	494	502	515	528
	SD	0.00	3.79	0.00	0.00

Oregon ELPA November 2007 Grade 11 ELPA
Round 3 Summary of Bookmark Placements

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	9	21	58	80
Median	Overall	9	21	58	80
Minimum	1	9	10	58	80
Minimum	Overall	9	10	58	80
Maximum	1	9	24	58	80
Maximum	Overall	9	24	58	80
SD	1	0.00	7.37	0.00	0.00
SD	Overall	0.00	7.37	0.00	0.00

Overall	Median	9	21	58	80
	Minimum	9	10	58	80
	Maximum	9	24	58	80
	SD	0.00	7.37	0.00	0.00

Oregon ELPA November 2007 Grade 11 ELPA
Round 3 Summary of Cut Scores

Statistic	Table	Early Intermediate	Intermediate	Early Advanced	Advanced
Median	1	494	501	515	528
Median	Overall	494	501	515	528
Minimum	1	494	495	515	528
Minimum	Overall	494	495	515	528
Maximum	1	494	502	515	528
Maximum	Overall	494	502	515	528
SD	1	0.00	3.79	0.00	0.00
SD	Overall	0.00	3.79	0.00	0.00

Overall	Median	494	501	515	528
	Minimum	494	495	515	528
	Maximum	494	502	515	528
	SD	0.00	3.79	0.00	0.00

Oregon ELPA November 2007 Grade 11 ELPA Round 3 Median Bookmark Summary

Table	Early Intermediate	Intermediate	Early Advanced	Advanced
1	9	21	58	80
Overall	9	21	58	80

Impact Data

	Beginning	Early Intermediate	Intermediate	Early Advanced	Advanced
Overall	8.8	7.6	30.0	46.2	7.4

SECTION F

Proficiency Level Descriptions

DRAFT



Oregon English Language Proficiency Assessment (ELPA)

Proficiency Level Descriptors (PLDs)

Prepared by CTB/McGraw-Hill
For the Oregon Department of Education
November 19, 2007

ELPA PROFICIENCY LEVEL DESCRIPTORS

Kindergarten	
Pre-production Level	<p>Students at the Pre-production level may demonstrate minimal comprehension of high-frequency words or phrases. On the ELPA, they</p> <ul style="list-style-type: none"> comprehend picture-referenced and highly contextualized words or very simple phrases. repeat, but with comprehension of only isolated words or high-frequency phrases. may use gestures to communicate meaning.
Beginning Level	<p>Students at the Beginning level are able to read and listen to basic information with very limited comprehension. They are able to speak and write using simple language with very limited accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> may be able decode and identify letter-sound correspondence in simple words. comprehend and respond to basic information in highly context-embedded, school-based social situations. orally respond to prompts with a very limited range of simple language and very limited accuracy.
Early Intermediate Level	<p>Students at the Early Intermediate level are able to read and listen to simple or highly contextualized grade-level information with limited comprehension. They are able to speak and write using simple language with limited accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> decode and identify letter-sound correspondence in simple words. comprehend simple information across a limited variety of social situations and subject areas in school-based situations. respond appropriately to simple prompts and orally express ideas with frequent grammatical and syntactical errors. comprehend details in reduced-complexity listening passages with a limited degree of comprehension. use a limited range of vocabulary and grammatical forms (e.g., simple verb forms, short phrases, and simple sentences).
Intermediate Level	<p>Students at the Intermediate are able to read and listen to some grade-level information with comprehension. They are able to speak using some complex language and write simple academic language with some accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> read a few sight words. use and understand vocabulary across a range of school-based situations and some academic subject areas. respond appropriately to prompts and orally express ideas with some grammatical and syntactical errors. comprehend main ideas and concrete details from short listening passages on a variety of topics. use sentences containing simple academic language (simple and compound sentences, basic verb tenses, prepositions) with increasing accuracy.

ELPA PROFICIENCY LEVEL DESCRIPTORS

Kindergarten	
Early Advanced Level	<p>Students at the Early Advanced level are able to read and listen to most grade-level information with comprehension. They are able to speak with increasingly complex language, and write academic language with some accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> • decode and identify sound-letter correspondence in a wider range of words. • use and understand most concepts and vocabulary from a variety of content areas. • comprehend details and main ideas in short listening passages. • comprehend and use simple language with accuracy, and some complex language (compound sentences, adverbials, a range of verb tenses) with occasional errors that do not interfere with academic performance. • read and complete short sentences with appropriate vocabulary.
Advanced (Proficient) Level	<p>Students at the Advanced (Proficient) level are able to consistently read and listen to an extensive range of complex grade-level information with comprehension. They are able to speak and write using an extensive range of complex language with a level of accuracy and fluency approximating native English speakers. On the ELPA, they</p> <ul style="list-style-type: none"> • use and understand concepts and vocabulary from a variety of content areas. • identify main ideas and details in listening passages containing complex language and academic vocabulary. • comprehend and use complex grammatical structures (e.g., relative clauses, tense/aspect structures) with only minor errors. • read sentences to answer questions about reading and listening passages.

ELPA PROFICIENCY LEVEL DESCRIPTORS

Grade 1	
Pre-production Level	<p>Students at the Pre-production level may demonstrate minimal comprehension of high-frequency words or phrases. On the ELPA, they</p> <ul style="list-style-type: none"> comprehend picture-referenced and highly contextualized words or very simple phrases. repeat, but with comprehension of only isolated words or high-frequency phrases. may use gestures to communicate meaning.
Beginning Level	<p>Students at the Beginning level are able to read and listen to basic information with very limited comprehension. They are able to speak and write using simple language with very limited accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> may be able to decode and identify letter-sound correspondence in simple words. comprehend and respond to basic information in highly context-embedded, school-based social situations. orally respond to prompts with a very limited range of simple language and very limited accuracy.
Early Intermediate Level	<p>Students at the Early Intermediate level are able to read and listen to simple or highly contextualized grade-level information with limited comprehension. They are able to speak and write using simple language with limited accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> read words and short simple sentences. comprehend simple information across a limited variety of social situations and subject areas in school-based situations. respond appropriately to simple prompts and orally express ideas with frequent grammatical and syntactical errors. comprehend details in reduced complexity listening passages with a limited degree of comprehension. use a limited range of vocabulary and grammatical forms (e.g., simple verb forms, short phrases, and simple sentences).
Intermediate Level	<p>Students at the Intermediate level are able to read and listen to some grade-level information with comprehension. They are able to speak using some complex language and write simple academic language with some accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> decode and identify sound-letter correspondence in words. use and understand vocabulary across a range of school-based situations and some academic subject areas. respond appropriately to prompts and orally express ideas with some grammatical and syntactical errors. comprehend main ideas and concrete details from short listening passages on a variety of topics. speak in sentences containing simple academic language (simple and compound sentences, basic verb tenses, prepositions) with increasing accuracy.

ELPA PROFICIENCY LEVEL DESCRIPTORS

Grade 1	
Early Advanced Level	<p>Students at the Early Advanced level are able to read and listen to most grade-level information with comprehension. They are able to speak with increasingly complex language, and write academic language with some accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> • use and understand most concepts and vocabulary from a variety of content areas. • comprehend details and main ideas in short listening passages. • comprehend and use simple language with accuracy, and some complex language. (compound sentences, adverbials, a range of verb tenses) with occasional errors that do not interfere with academic performance. • read and complete short sentences with appropriate vocabulary.
Advanced (Proficient) Level	<p>Students at the Advanced (Proficient) level are able to consistently read and listen to an extensive range of complex and abstract grade-level information with comprehension. They are able to speak and write using an extensive range of complex language with a level of accuracy and fluency approximating native English speakers. On the ELPA, they</p> <ul style="list-style-type: none"> • use and understand concepts and vocabulary from a variety of content areas. • identify main ideas and details in listening passages containing complex language and academic vocabulary. • read short sentences. • comprehend and use complex grammatical structures (e.g., relative clauses, tense/aspect structures) with only minor errors.

ELPA PROFICIENCY LEVEL DESCRIPTORS

Grades 2–3	
Pre-production Level	<p>Students at the Pre-production level may demonstrate minimal comprehension of high-frequency words or phrases. On the ELPA, they</p> <ul style="list-style-type: none"> comprehend picture-referenced and highly contextualized words or very simple phrases. repeat, but with comprehension of only isolated words or high-frequency phrases. may use gestures to communicate meaning.
Beginning Level	<p>Students at the Beginning level are able to read and listen to basic information with very limited comprehension. They are able to speak and write using simple language with very limited accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> decode and identify letter-sound correspondence in simple words and phrases. comprehend and respond to basic information in highly context-embedded, school-based social situations. orally respond to prompts with a very limited range of simple language and very limited accuracy. write using memorized vocabulary and simple phrases with very limited accuracy.
Early Intermediate Level	<p>Students at the Early Intermediate level are able to read and listen to simple or highly contextualized grade-level information with limited comprehension. They are able to speak and write using simple language with limited accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> use context and known vocabulary to decode text. comprehend simple information across a limited variety of social situations and subject areas in school-based situations. respond appropriately to simple prompts and orally express ideas with frequent grammatical and syntactical errors. comprehend details in reduced complexity reading and listening passages. write and speak using a limited range of vocabulary and grammatical forms (e.g., simple verb forms, short phrases, and simple sentences).
Intermediate Level	<p>Students at the Intermediate level are able to read and listen to some grade-level information with comprehension. They are able to speak using some complex language and write simple academic language with some accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> use and understand vocabulary across a range of school-based situations and some academic subject areas. respond appropriately to prompts and orally express ideas with some grammatical and syntactical errors. comprehend main ideas and concrete details from reading and listening passages on a variety of topics. write and speak using a limited number of sentences containing simple academic language (simple and compound sentences, basic verb tenses, prepositions) with increasing accuracy.

ELPA PROFICIENCY LEVEL DESCRIPTORS

Grades 2–3	
Early Advanced Level	<p>Students at the Early Advanced level are able to read and listen to most grade-level information with comprehension. They are able to speak with increasingly complex language, and write academic language with some accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> • use and understand most concepts and vocabulary from a variety of content areas. • comprehend academic reading and listening passages by demonstrating some of the following skills: (1) identify details and main ideas; (2) identify sequence of events and processes; (3) interpret meaning. • comprehend and use simple language with accuracy, and some complex language (compound sentences, adverbials, a range of verb tenses) with occasional errors that do not interfere with academic performance. • organize written and spoken information into clear sentences with mostly appropriate transitions and some supporting details.
Advanced (Proficient) Level	<p>Students at the Advanced (Proficient) level are able to consistently read and listen to an extensive range of complex and abstract grade-level information with comprehension. They are able to speak and write using an extensive range of complex language with a level of accuracy and fluency approximating native English speakers. On the ELPA, they</p> <ul style="list-style-type: none"> • use and understand concepts and vocabulary from an extensive variety of content areas. • comprehend reading and listening passages containing complex language and academic vocabulary by demonstrating the following skills: (1) determine the meaning of vocabulary from complex context; (2) identify details and main ideas; (3) make inferences; (4) distinguish between fact and opinion; (5) evaluate the purpose of text; (6) summarize and paraphrase information from listening and reading passages. • comprehend and use complex grammatical structures (e.g., relative clauses, tense/aspect structures) with only minor errors. • organize written and spoken information into coherent discourse with appropriate transitions and supporting details.

ELPA PROFICIENCY LEVEL DESCRIPTORS

Grades 4–5	
Pre-production Level	<p>Students at the Pre-production level may demonstrate minimal comprehension of high-frequency words or phrases. On the ELPA, they</p> <ul style="list-style-type: none"> comprehend picture-referenced and highly contextualized words or very simple phrases. repeat, but with comprehension of only isolated words or high-frequency phrases. may use gestures to communicate meaning.
Beginning Level	<p>Students at the Beginning level are able to read and listen to basic information with very limited comprehension. They are able to speak and write using simple language with very limited accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> decode and accurately identify letter-sound correspondence in simple words and phrases. comprehend and respond to basic information in highly context-embedded, school-based social situations. orally respond to prompts with a very limited range of simple language and very limited accuracy. write using memorized vocabulary and simple phrases with very limited accuracy.
Early Intermediate Level	<p>Students at the Early Intermediate level are able to read and listen to simple or highly contextualized grade-level information with limited comprehension. They are able to speak and write using simple language with limited accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> use context and known vocabulary to decode text. comprehend simple information across a limited variety of social situations and subject areas in school-based situations. respond appropriately to simple prompts and orally express ideas with frequent grammatical and syntactical errors. comprehend details in reduced complexity reading and listening passages with a limited degree of comprehension. write and speak using a limited range of vocabulary and grammatical forms (e.g., simple verb forms, short phrases, and simple sentences).
Intermediate Level	<p>Students at the Intermediate level are able to read and listen to some grade-level information with comprehension. They are able to speak using some complex language and write simple academic language with some accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> use and understand vocabulary across a range of school-based situations and some academic subject areas. respond appropriately to prompts and orally express ideas with some grammatical and syntactical errors. comprehend main ideas and concrete details from reading and listening passages on a variety of topics. write and speak using a limited number of sentences containing simple academic language (simple and compound sentences, basic verb tenses, prepositions) with increasing accuracy.

ELPA PROFICIENCY LEVEL DESCRIPTORS

Grades 4–5	
Early Advanced Level	<p>Students at the Early Advanced are able to read and listen to most grade-level information with comprehension. They are able to speak with increasingly complex language, and write academic language with some accuracy and fluency. On the ELPA, they</p> <p>use and understand most concepts and vocabulary from a variety of content areas.</p> <p>comprehend academic reading and listening passages by demonstrating some of the following skills: (1) identify details and main ideas; (2) identify sequence of events and processes; (3) interpret meaning.</p> <p>comprehend and use simple language with accuracy, and some complex language (compound sentences, adverbials, a range of verb tenses) with occasional errors that do not interfere with academic performance.</p> <p>organize written and spoken information into clear sentences with mostly appropriate transitions and some supporting details.</p>
Advanced (Proficient) Level	<p>Students at the Advanced (Proficient) level are able to consistently read and listen to an extensive range of complex and abstract grade-level information with comprehension. They are able to speak and write using an extensive range of complex language with a level of accuracy and fluency approximating native English speakers. On the ELPA, they</p> <p>use and understand concepts and vocabulary from an extensive variety of content areas.</p> <p>comprehend reading and listening passages containing complex language and academic vocabulary by demonstrating the following skills: (1) determine the meaning of vocabulary from complex context; (2) identify details and main ideas; (3) make inferences; (4) distinguish between fact and opinion; (5) evaluate the purpose of text; (6) summarize and paraphrase information from listening and reading passages.</p> <p>comprehend and use complex grammatical structures (e.g., relative clauses, tense/aspect structures) with only minor errors.</p> <p>organize written and spoken information into coherent discourse with appropriate transitions and supporting details.</p>

ELPA PROFICIENCY LEVEL DESCRIPTORS

Grades 6–8	
Pre-production Level	<p>Students at the Pre-production level may demonstrate minimal comprehension of high-frequency words or phrases. On the ELPA, they</p> <ul style="list-style-type: none"> comprehend picture-referenced and highly contextualized words or very simple phrases. repeat, but with comprehension of only isolated words or high-frequency phrases. may use gestures to communicate meaning.
Beginning Level	<p>Students at the Beginning level are able to read and listen to basic information with very limited comprehension. They are able to speak and write using simple language with very limited accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> decode and accurately identify letter-sound correspondence in simple words and phrases. comprehend and respond to basic information in highly context-embedded, school-based social situations. orally respond to prompts with a very limited range of simple language and very limited accuracy. write using memorized vocabulary and simple phrases with very limited accuracy.
Early Intermediate Level	<p>Students at the Early Intermediate level are able to read and listen to simple or highly contextualized grade-level information with limited comprehension. They are able to speak and write using simple language with limited accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> use context and known vocabulary to decode text. comprehend simple information across a limited variety of social situations and subject areas in school-based situations. respond appropriately to simple prompts and orally express ideas with frequent grammatical and syntactical errors. comprehend details in reduced complexity reading and listening passages with a limited degree of comprehension. write and speak using a limited range of vocabulary and grammatical forms (e.g., simple verb forms, short phrases, and simple sentences).
Intermediate Level	<p>Students at the Intermediate level require (considerable) ongoing instructional support in all content areas. They are able to read and listen to some grade-level information with comprehension. They are able to speak using some complex language and write simple academic language with some accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> use and understand vocabulary across a range of school-based situations and some academic subject areas. respond appropriately to prompts and orally express ideas with some grammatical and syntactical errors. comprehend main ideas and concrete details from reading and listening passages on a variety of topics. write and speak using a limited number of sentences containing simple academic language (simple and compound sentences, basic verb tenses, prepositions) with increasing accuracy.

ELPA PROFICIENCY LEVEL DESCRIPTORS

Grades 6–8	
Early Advanced Level	<p>Students at the Early Advanced level are able to read and listen to most grade-level information with comprehension. They are able to speak with increasingly complex language, and write academic language with some accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> • use and understand most concepts and vocabulary from a variety of content areas. • comprehend academic reading and listening passages by demonstrating some of the following skills: (1) identify details and main ideas; (2) identify sequence of events and processes; (3) interpret meaning. • comprehend and use simple language with accuracy, and some complex language (compound sentences, adverbials, a range of verb tenses) with occasional errors that do not interfere with academic performance. • organize written and spoken information into clear sentences with mostly appropriate transitions and some supporting details.
Advanced (Proficient) Level	<p>Students at the Advanced (Proficient) level are able to consistently read and listen to an extensive range of complex and abstract grade-level information with comprehension. They are able to speak and write using an extensive range of complex language with a level of accuracy and fluency approximating native English speakers. On the ELPA, they</p> <ul style="list-style-type: none"> • use and understand concepts and vocabulary from an extensive variety of content areas. • comprehend reading and listening passages containing complex language and academic vocabulary by demonstrating the following skills: (1) determine the meaning of vocabulary from complex context; (2) identify details and main ideas; (3) make inferences; (4) distinguish between fact and opinion; (5) evaluate the purpose of text; (6) summarize and paraphrase information from listening and reading passages. • comprehend and use complex grammatical structures (e.g., relative clauses, tense/aspect structures) with only minor errors. • organize written and spoken information into coherent discourse with appropriate transitions and supporting details.

ELPA PROFICIENCY LEVEL DESCRIPTORS

Grades 9–12	
Pre-production Level	<p>Students at the Pre-production level may demonstrate minimal comprehension of high-frequency words or phrases. On the ELPA, they</p> <ul style="list-style-type: none"> comprehend picture-referenced and highly contextualized words or very simple phrases. repeat, but with comprehension of only isolated words or high-frequency phrases. may use gestures to communicate meaning.
Beginning Level	<p>Students at the Beginning level are able to read and listen to basic information with very limited comprehension. They are able to speak and write using simple language with very limited accuracy and fluency. On the ELPA, they</p> <p>decode and accurately identify letter-sound correspondence in simple words and phrases.</p> <p>comprehend and respond to basic information in highly context-embedded, school-based social situations.</p> <p>orally respond to prompts with a very limited range of simple language and very limited accuracy.</p> <p>write using memorized vocabulary and simple phrases with very limited accuracy.</p>
Early Intermediate Level	<p>Students at the Early Intermediate level are able to read and listen to simple or highly contextualized grade-level information with limited comprehension. They are able to speak and write using simple language with limited accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> use context and known vocabulary to decode text. comprehend simple information across a limited variety of social situations and subject areas in school-based situations. respond appropriately to simple prompts and orally express ideas with frequent grammatical and syntactical errors. comprehend details in reduced complexity reading and listening passages with a limited degree of comprehension. write and speak using a limited range of vocabulary and grammatical forms (e.g., simple verb forms, short phrases, and simple sentences).
Intermediate Level	<p>Students at the Intermediate level are able to read and listen to some grade-level information with comprehension. They are able to speak using some complex language and write simple academic language with some accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> use and understand vocabulary across a range of school-based situations and some academic subject areas. respond appropriately to prompts and orally express ideas with some grammatical and syntactical errors. comprehend main ideas and concrete details from reading and listening passages on a variety of topics. write and speak using a limited number of sentences containing simple academic language (simple and compound sentences, basic verb tenses, prepositions) with increasing accuracy.

ELPA PROFICIENCY LEVEL DESCRIPTORS

Grades 9–12	
Early Advanced Level	<p>Students at the Early Advanced level are able to read and listen to most grade-level information with comprehension. They are able to speak with increasingly complex language, and write academic language with some accuracy and fluency. On the ELPA, they</p> <ul style="list-style-type: none"> • use and understand most concepts and vocabulary from a variety of content areas. • comprehend academic reading and listening passages by demonstrating some of the following skills: (1) identify details and main ideas; (2) identify sequence of events and processes; (3) interpret meaning. • Comprehend and use simple language with accuracy, and some complex language (compound sentences, adverbials, a range of verb tenses) with occasional errors that do not interfere with academic performance. • organize written and spoken information into clear sentences with mostly appropriate transitions and some supporting details.
Advanced (Proficient) Level	<p>Students at the Advanced (Proficient) level are able to consistently read and listen to an extensive range of complex and abstract grade-level information with comprehension. They are able to speak and write using an extensive range of complex language with a level of accuracy and fluency approximating native English speakers. On the ELPA, they</p> <ul style="list-style-type: none"> • use and understand concepts and vocabulary from an extensive variety of content areas. • comprehend reading and listening passages containing complex language and academic vocabulary by demonstrating the following skills: (1) determine the meaning of vocabulary from complex context; (2) identify details and main ideas; (3) make inferences; (4) distinguish between fact and opinion; (5) evaluate the purpose of text; (6) summarize and paraphrase information from listening and reading passages. • comprehend and use complex grammatical structures (e.g., relative clauses, tense/aspect structures) with only minor errors. • organize written and spoken information into coherent discourse with appropriate transitions and supporting details.

SECTION G

Graphical Representations of Participants' Judgments

DRAFT

Oregon ELPA November 2007 Grade K Early Intermediate Cut Point

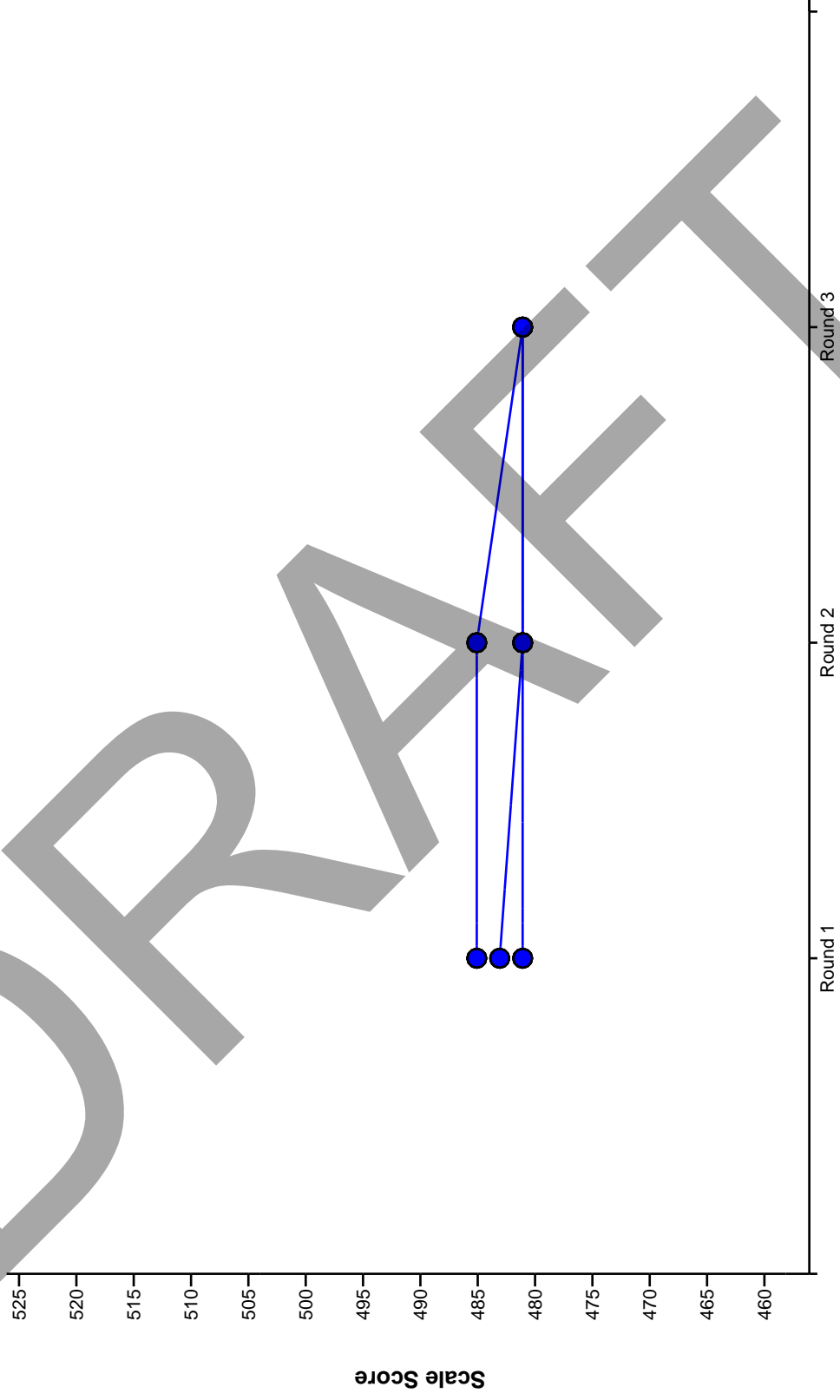


Table 1

Oregon ELPA November 2007 Grade K Intermediate Cut Point

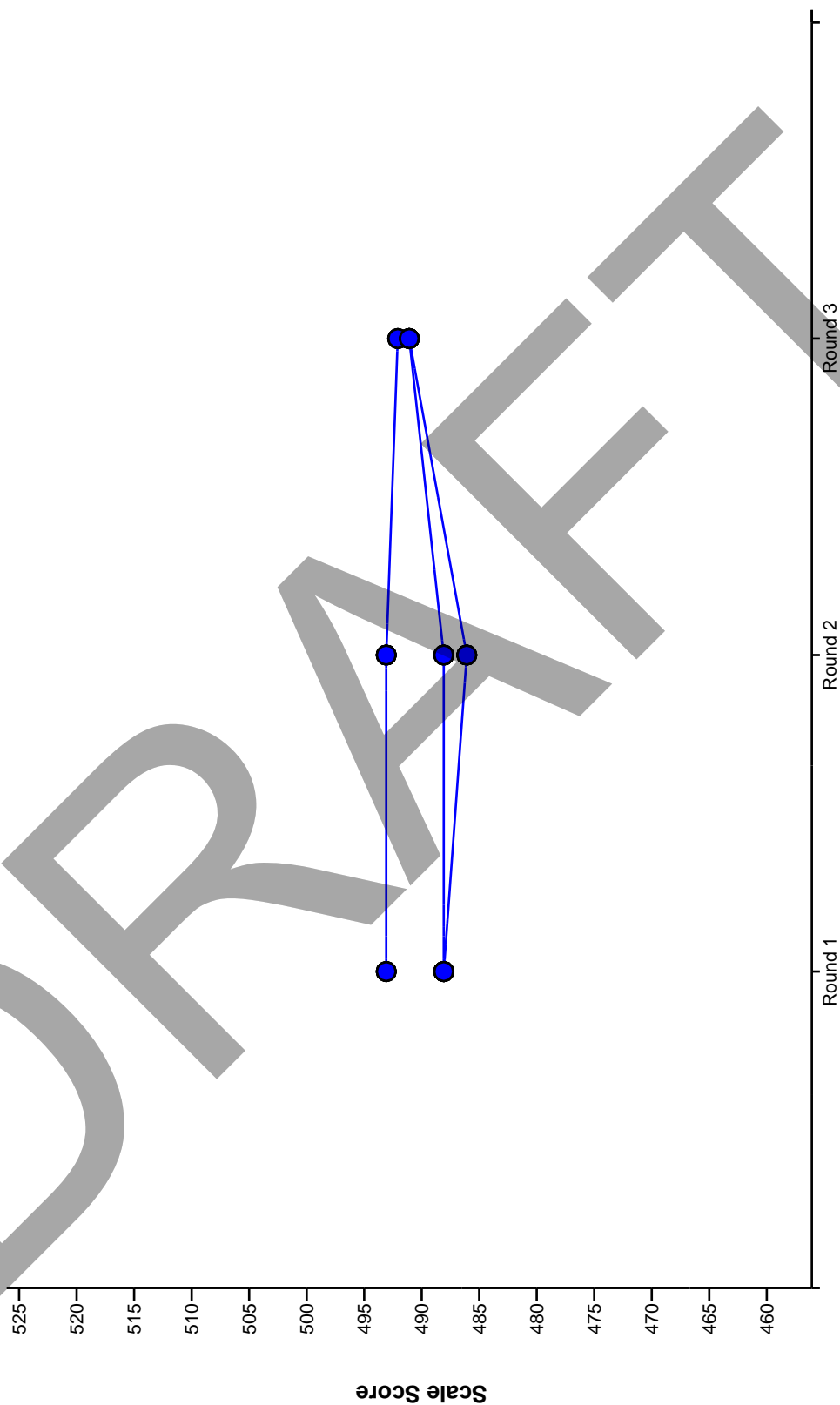


Table 1

Oregon ELPA November 2007 Grade K Early Advanced Cut Point

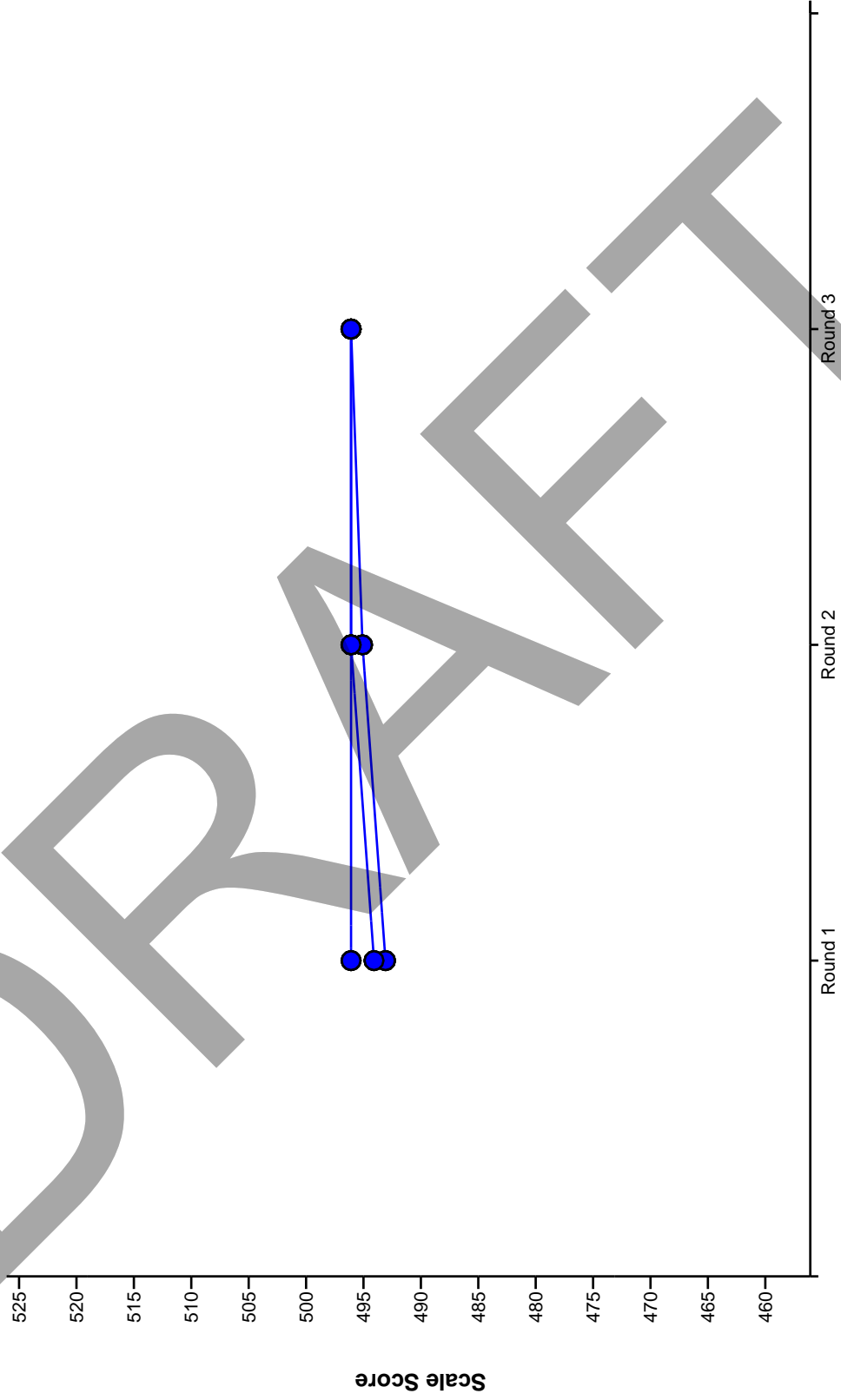


Table 1

Oregon ELPA November 2007 Grade K Advanced Cut Point

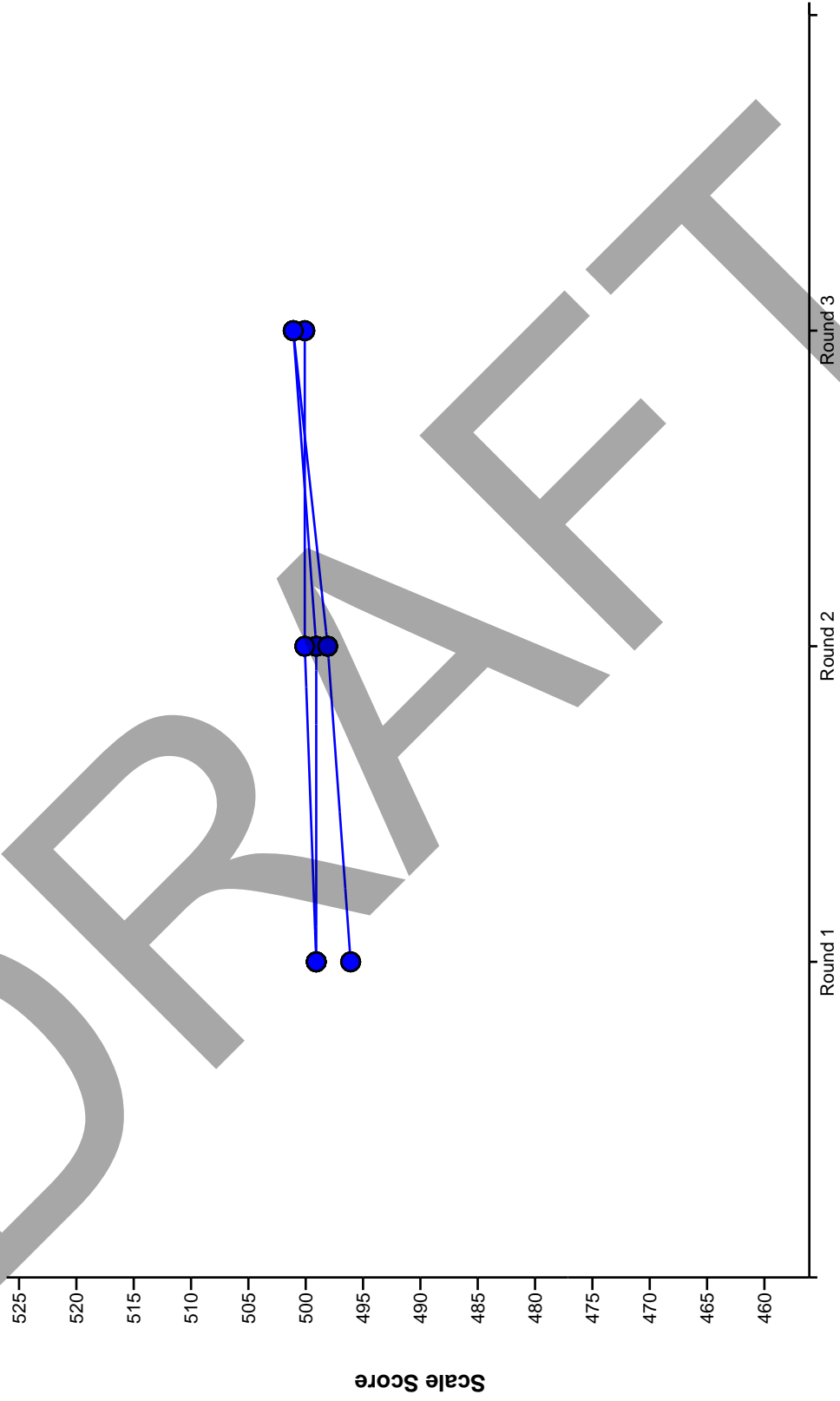


Table 1

Oregon ELPA November 2007 Grade 1 Early Intermediate Cut Point

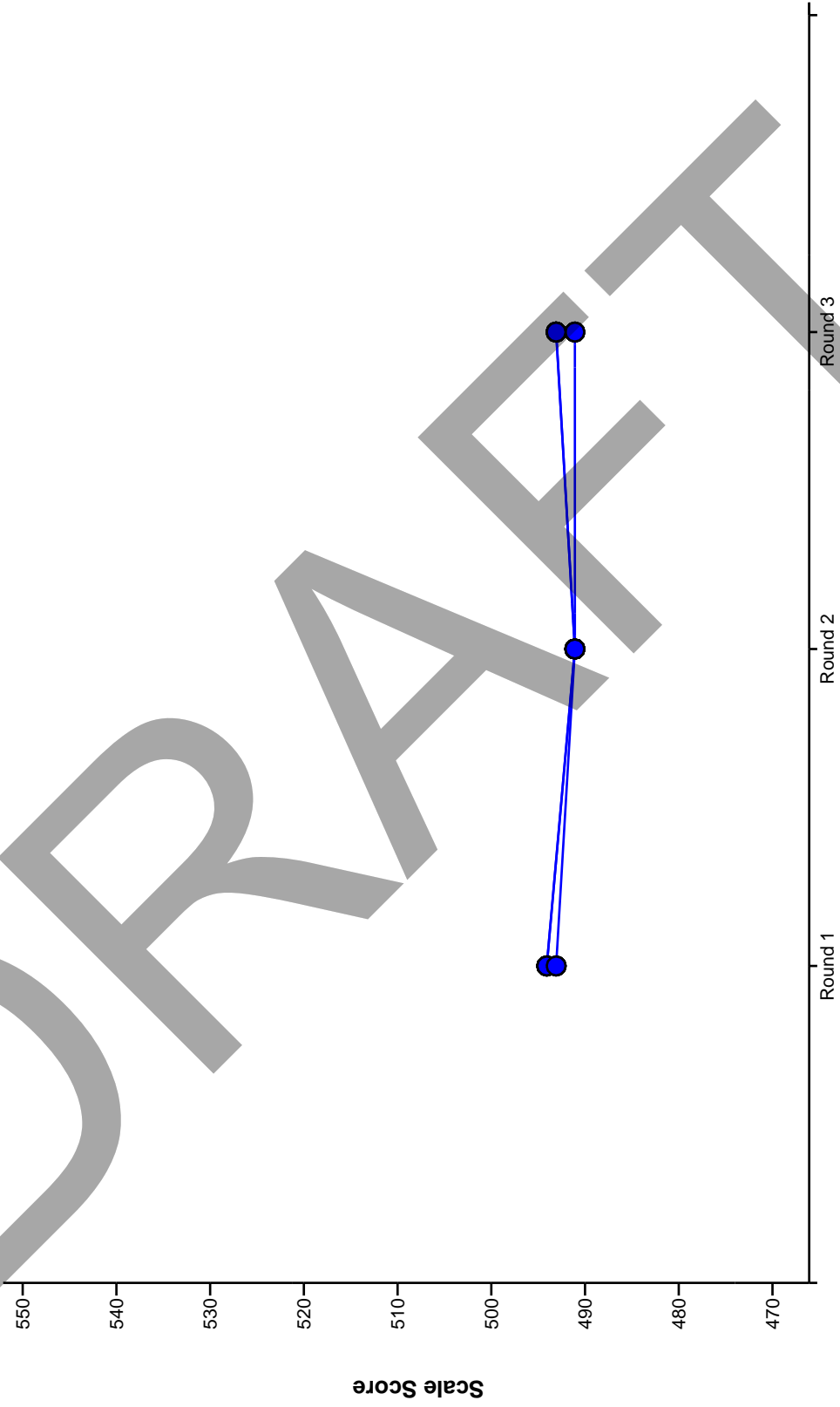


Table 1

Oregon ELPA November 2007 Grade 1 Intermediate Cut Point

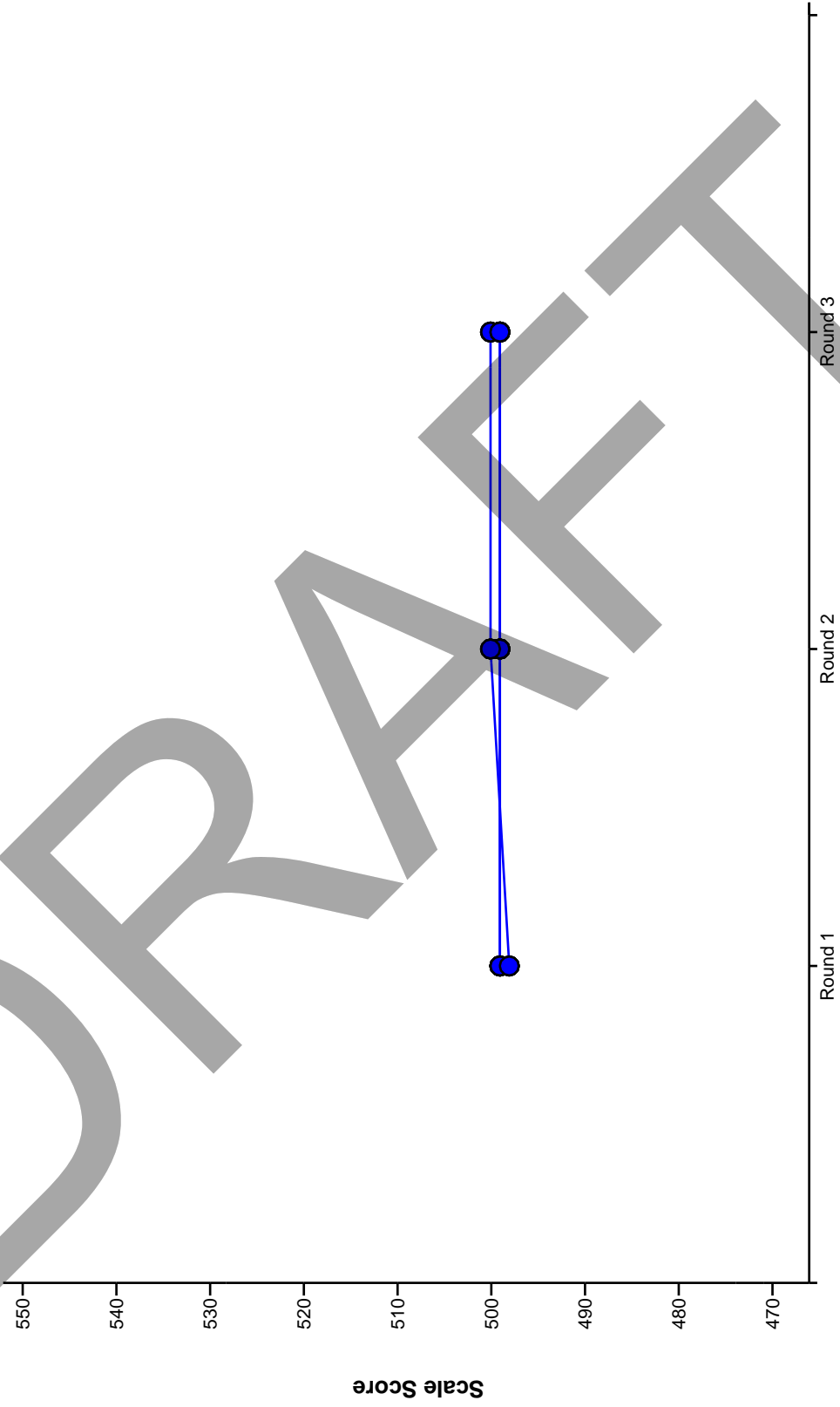


Table 1

Oregon ELPA November 2007 Grade 1 Early Advanced Cut Point

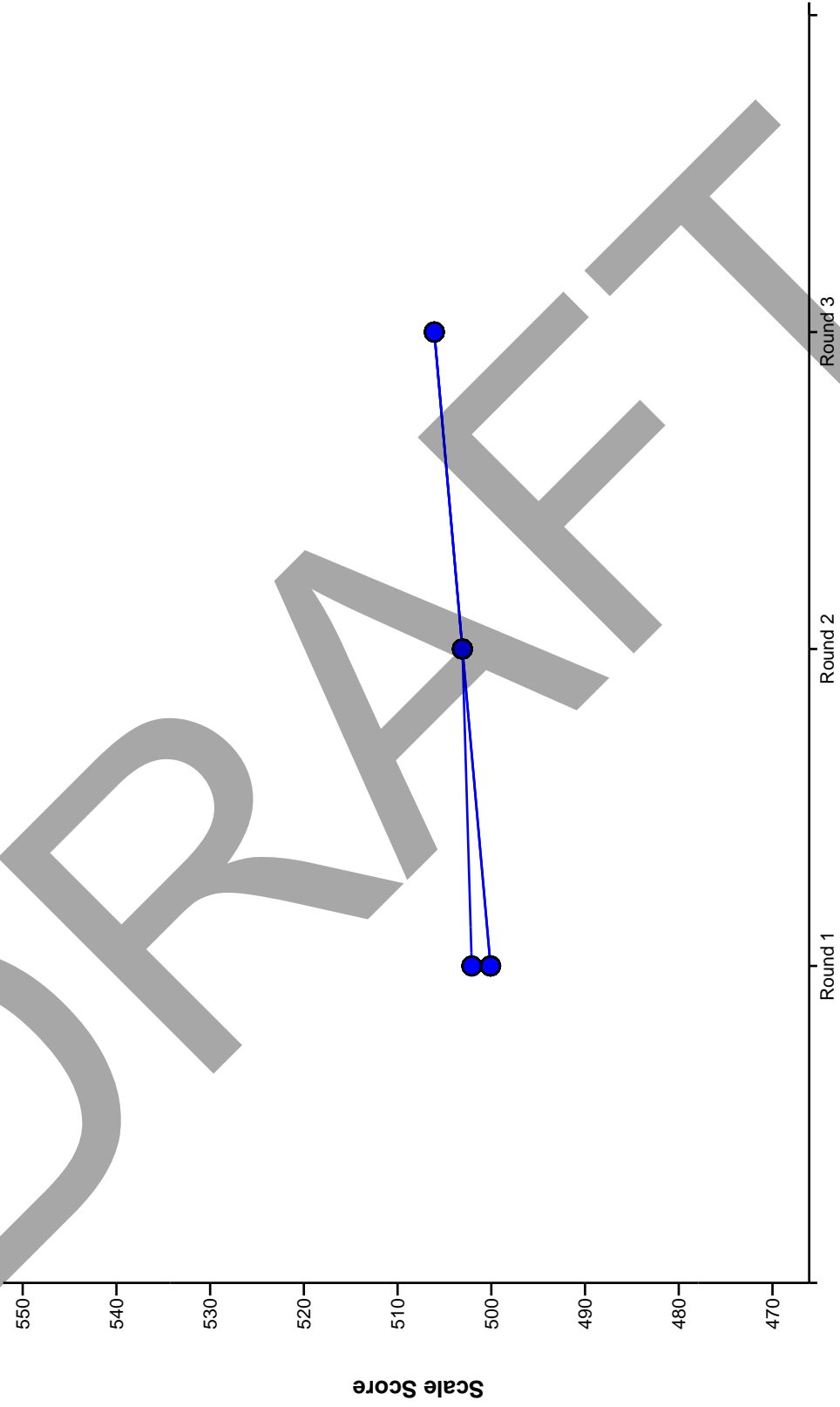


Table 1

Oregon ELPA November 2007 Grade 1 Advanced Cut Point

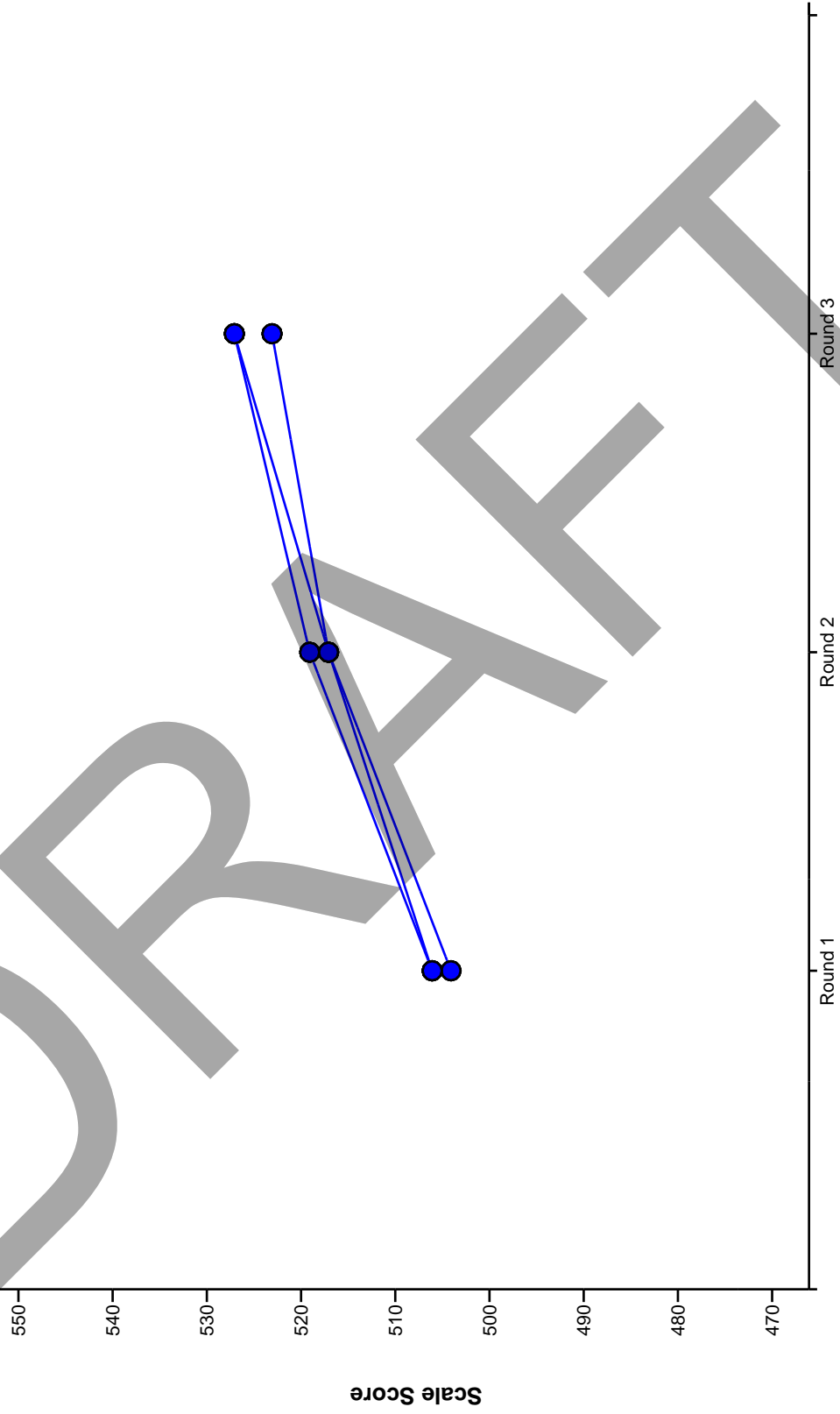


Table 1

Oregon ELPA November 2007 Grade 2 Early Intermediate Cut Point

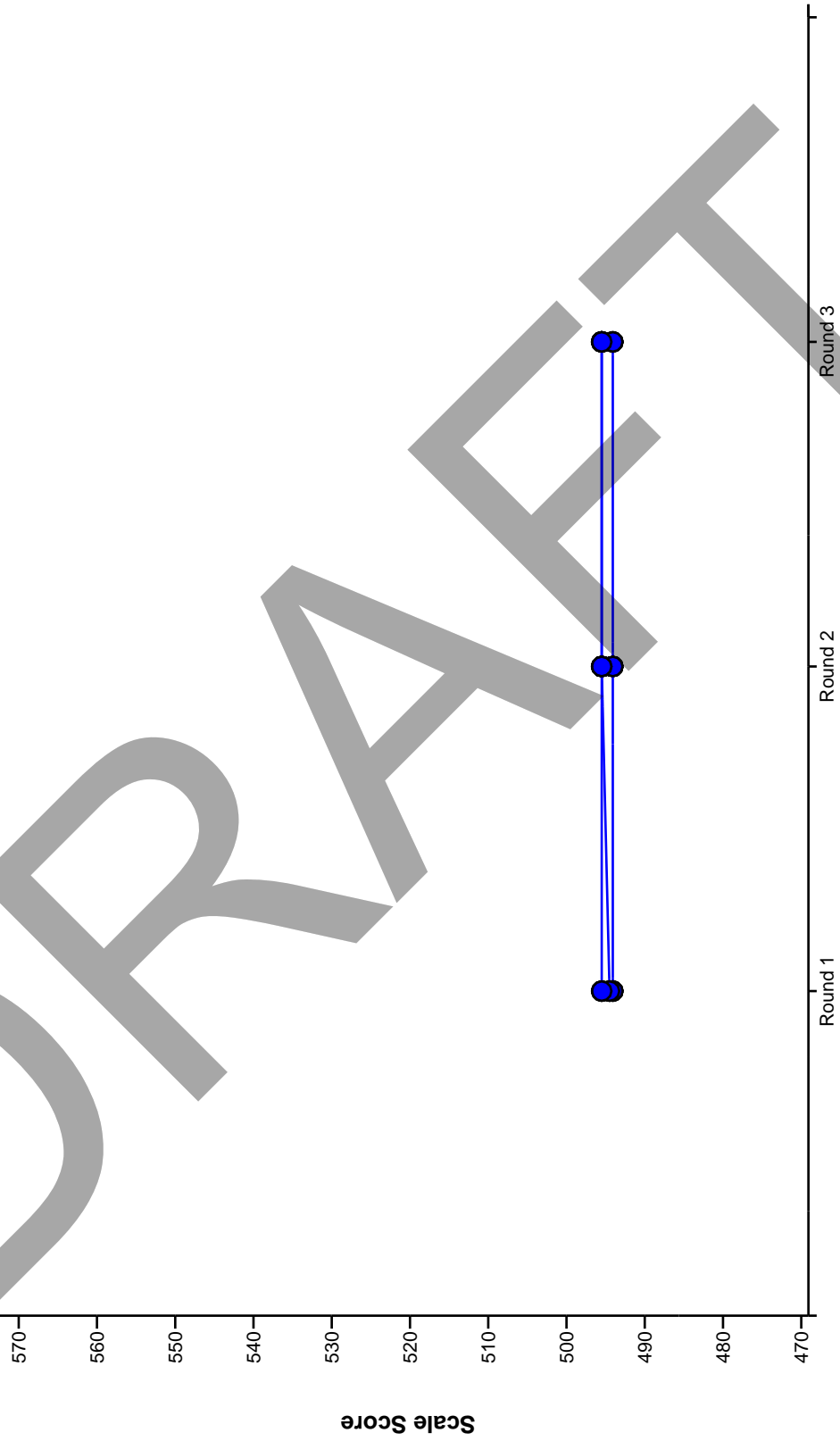


Table 1

Oregon ELPA November 2007 Grade 2 Intermediate Cut Point

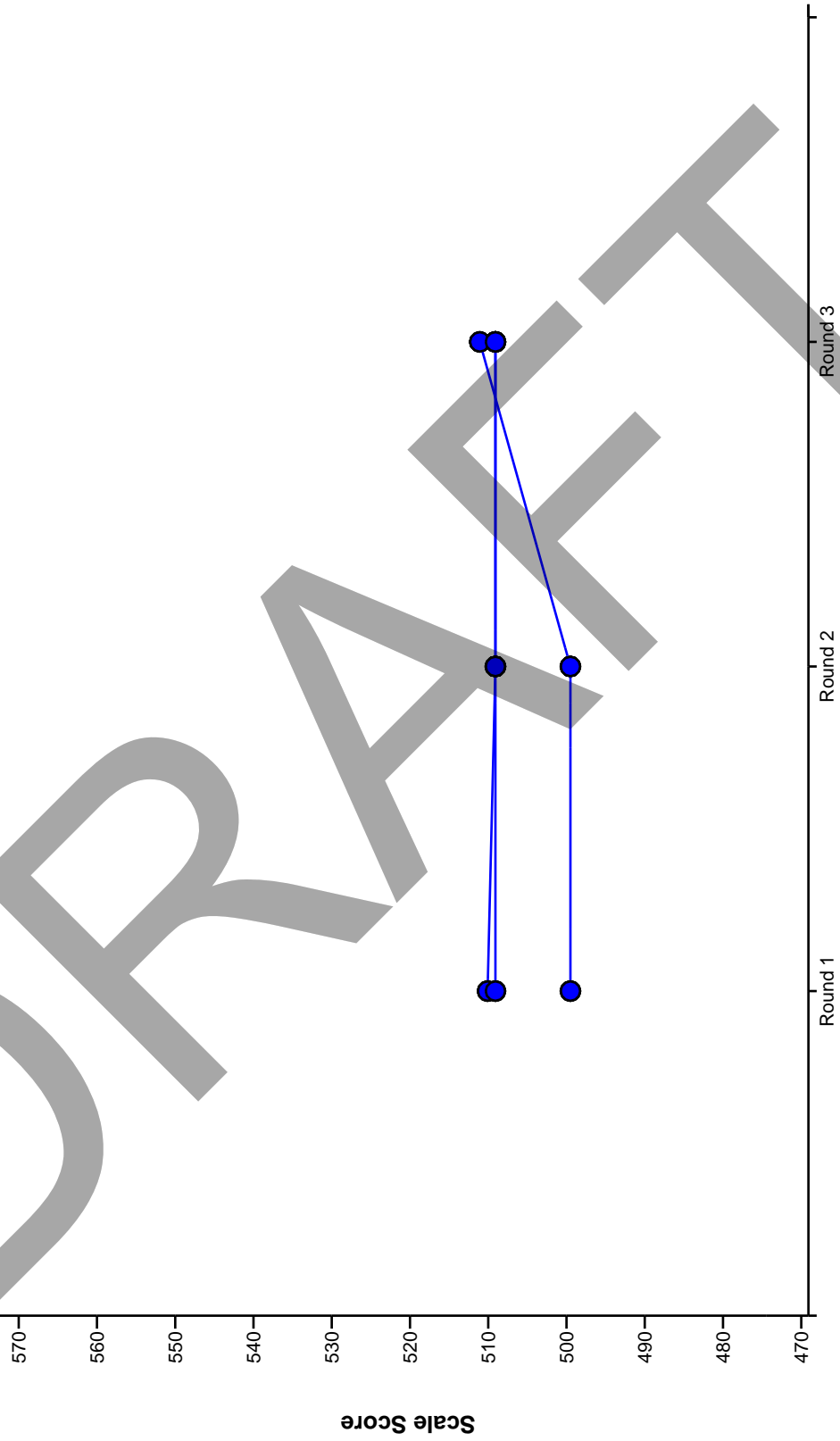


Table 1

Oregon ELPA November 2007 Grade 2 Early Advanced Cut Point

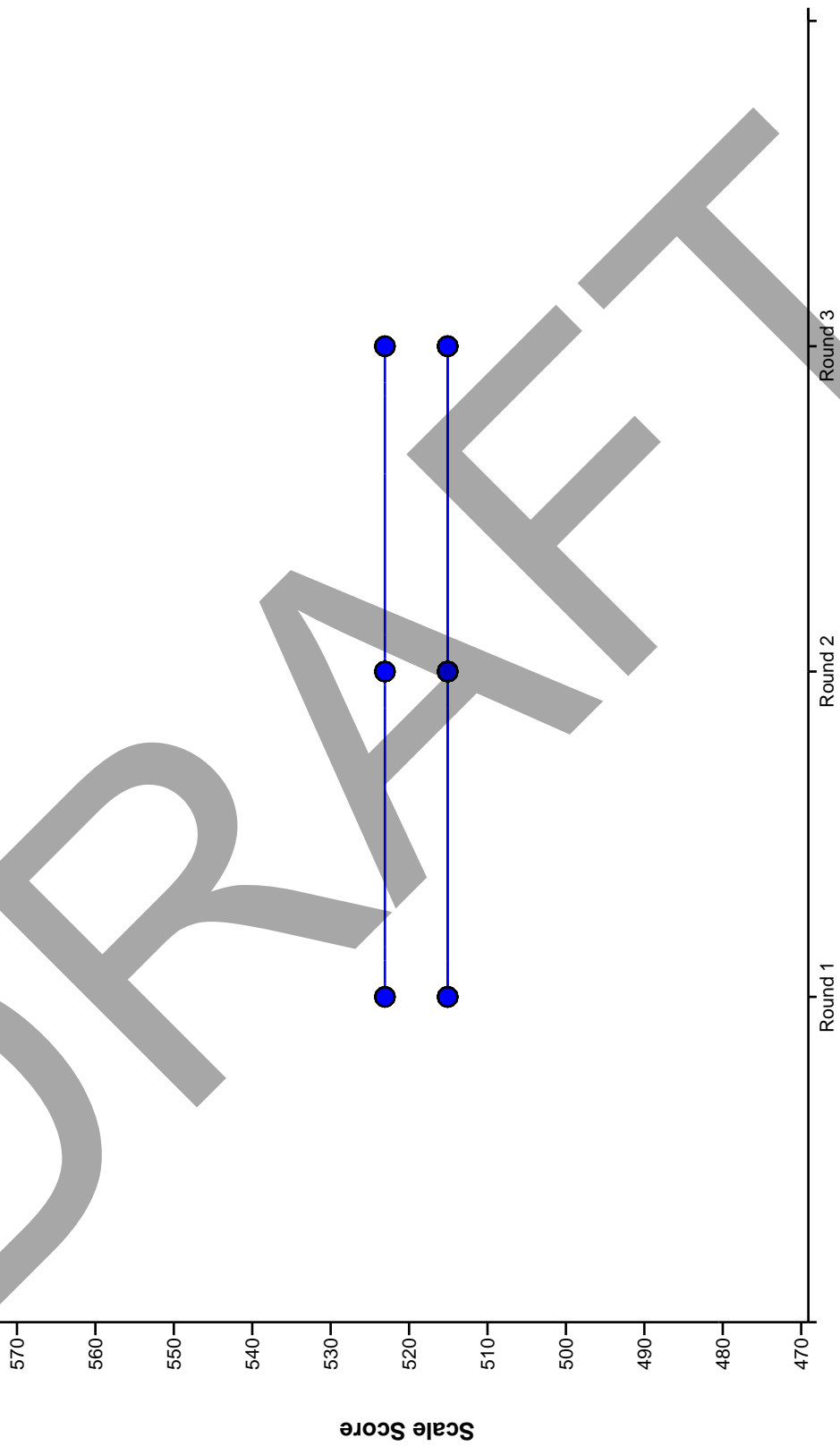


Table 1

Oregon ELPA November 2007 Grade 2 Advanced Cut Point

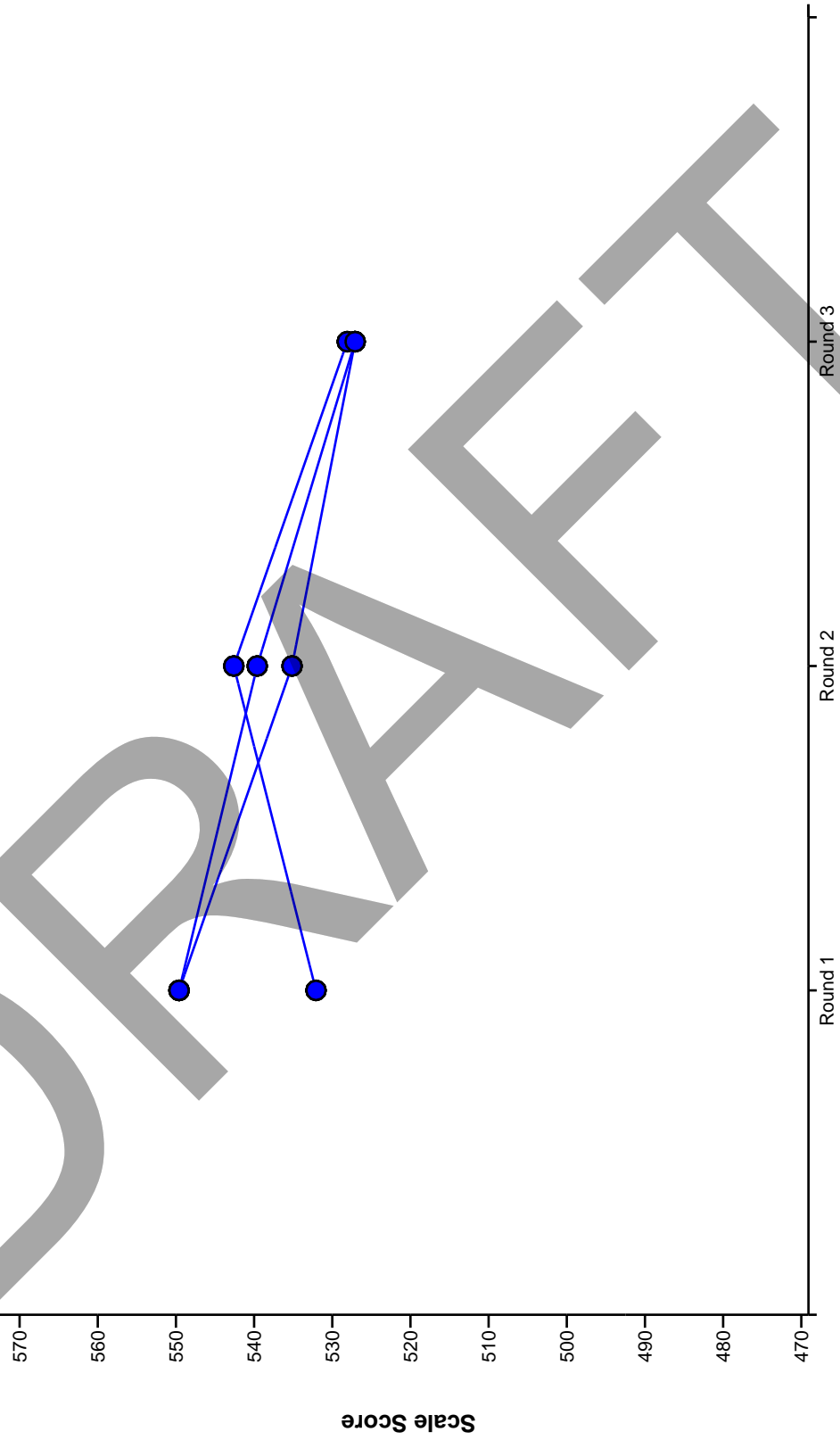


Table 1

Oregon ELPA November 2007 Grade 5 Early Intermediate Cut Point

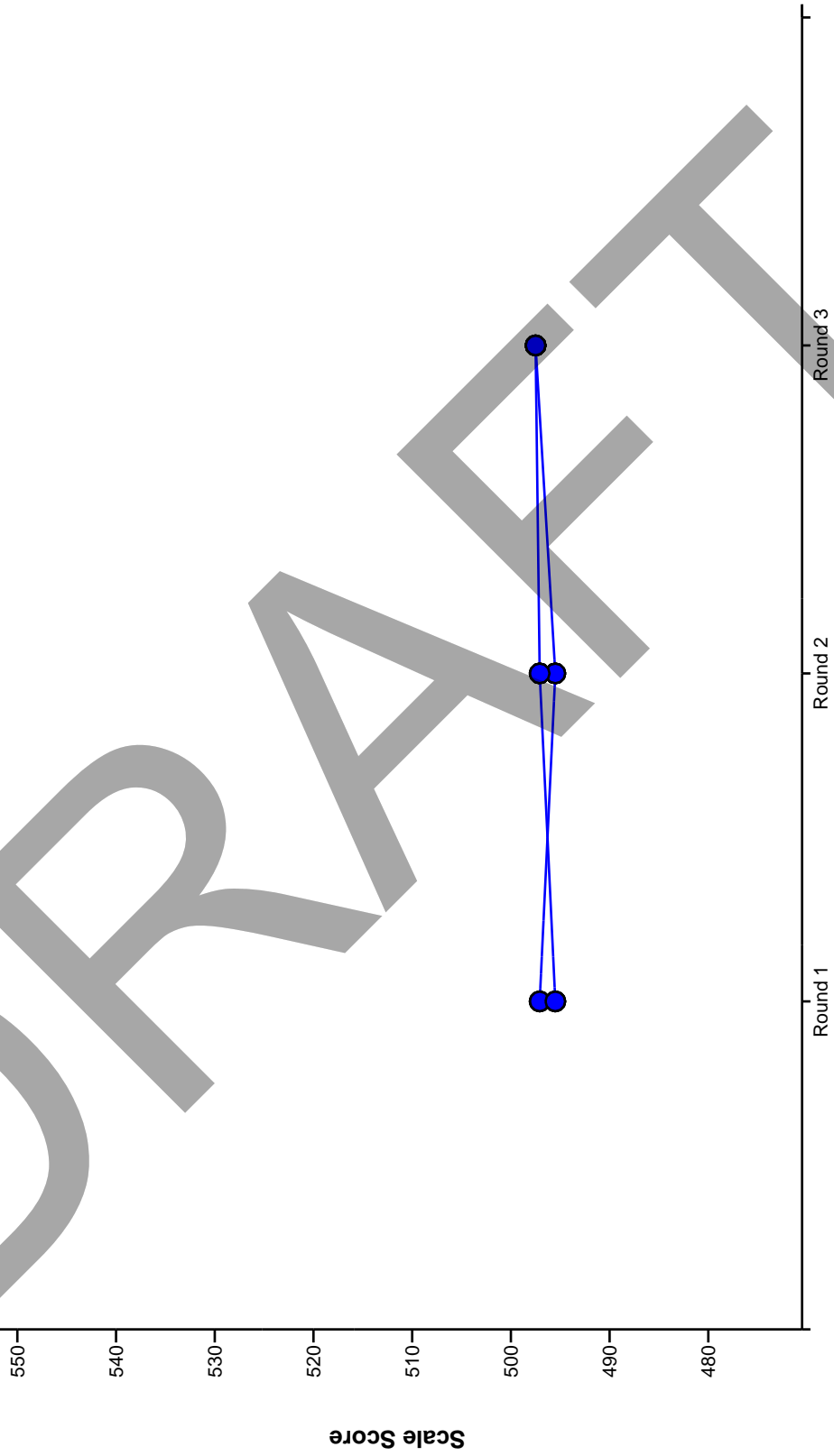


Table 1

Oregon ELPA November 2007 Grade 5 Intermediate Cut Point

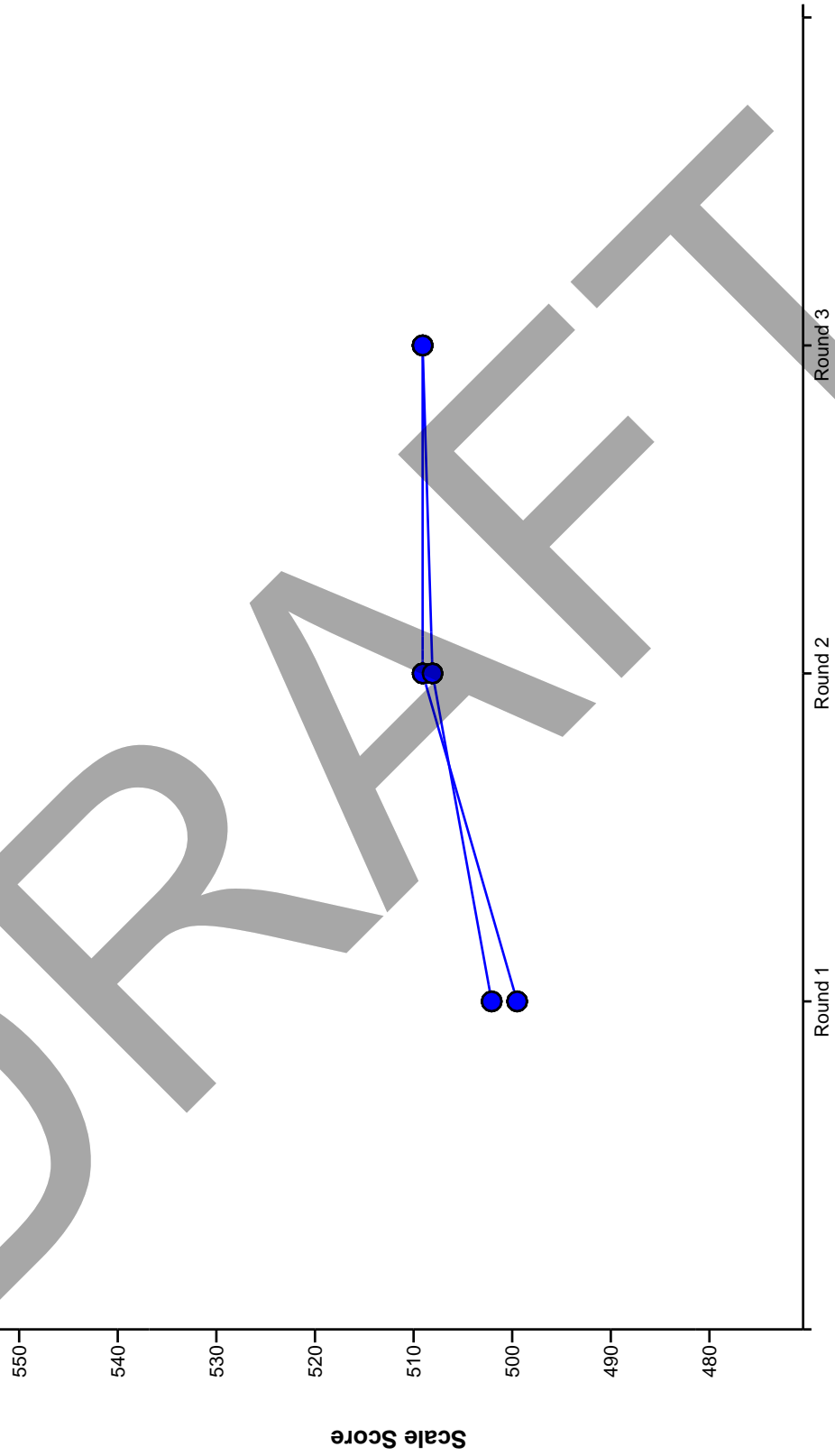


Table 1

Oregon ELPA November 2007 Grade 5 Early Advanced Cut Point

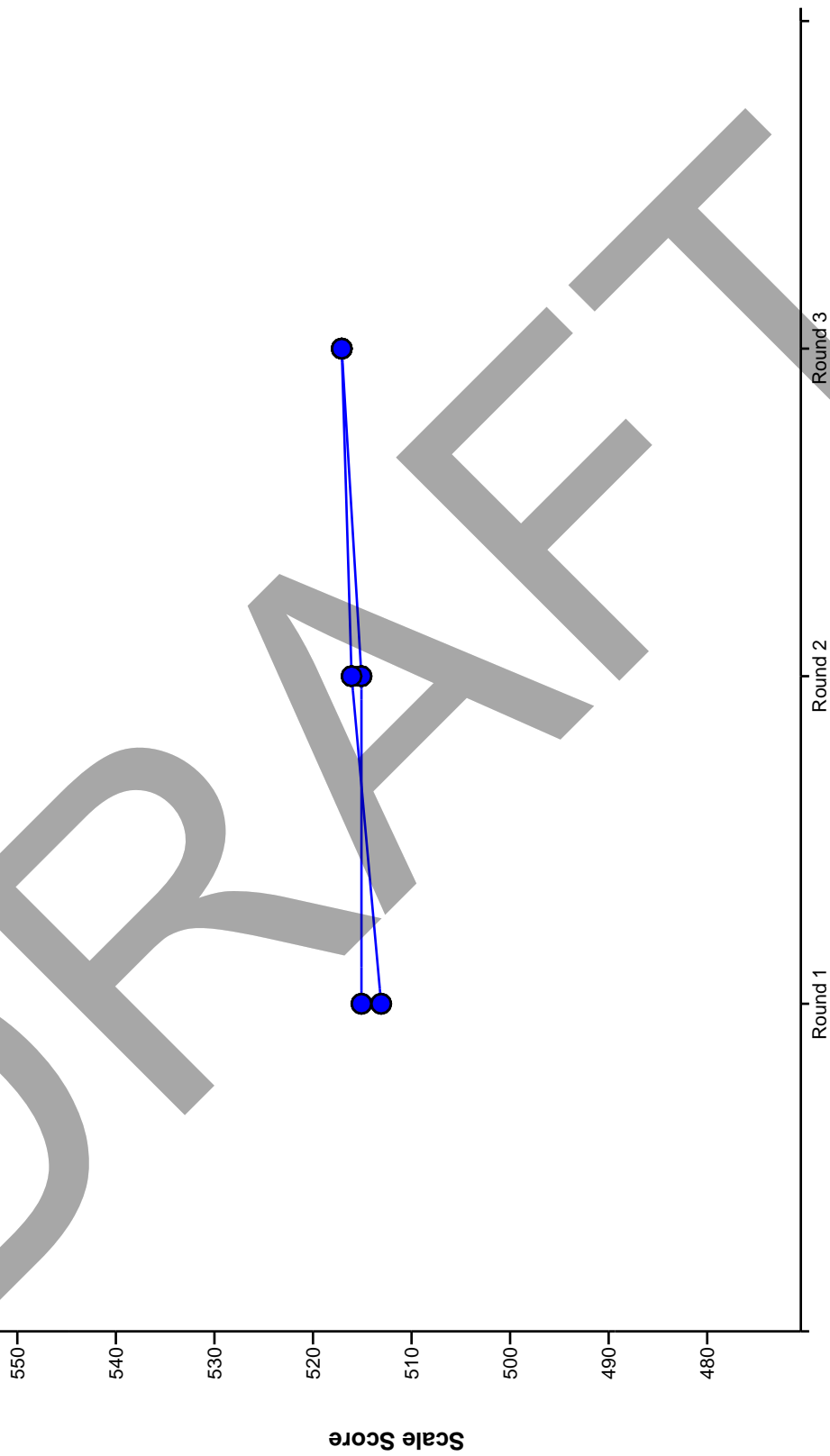
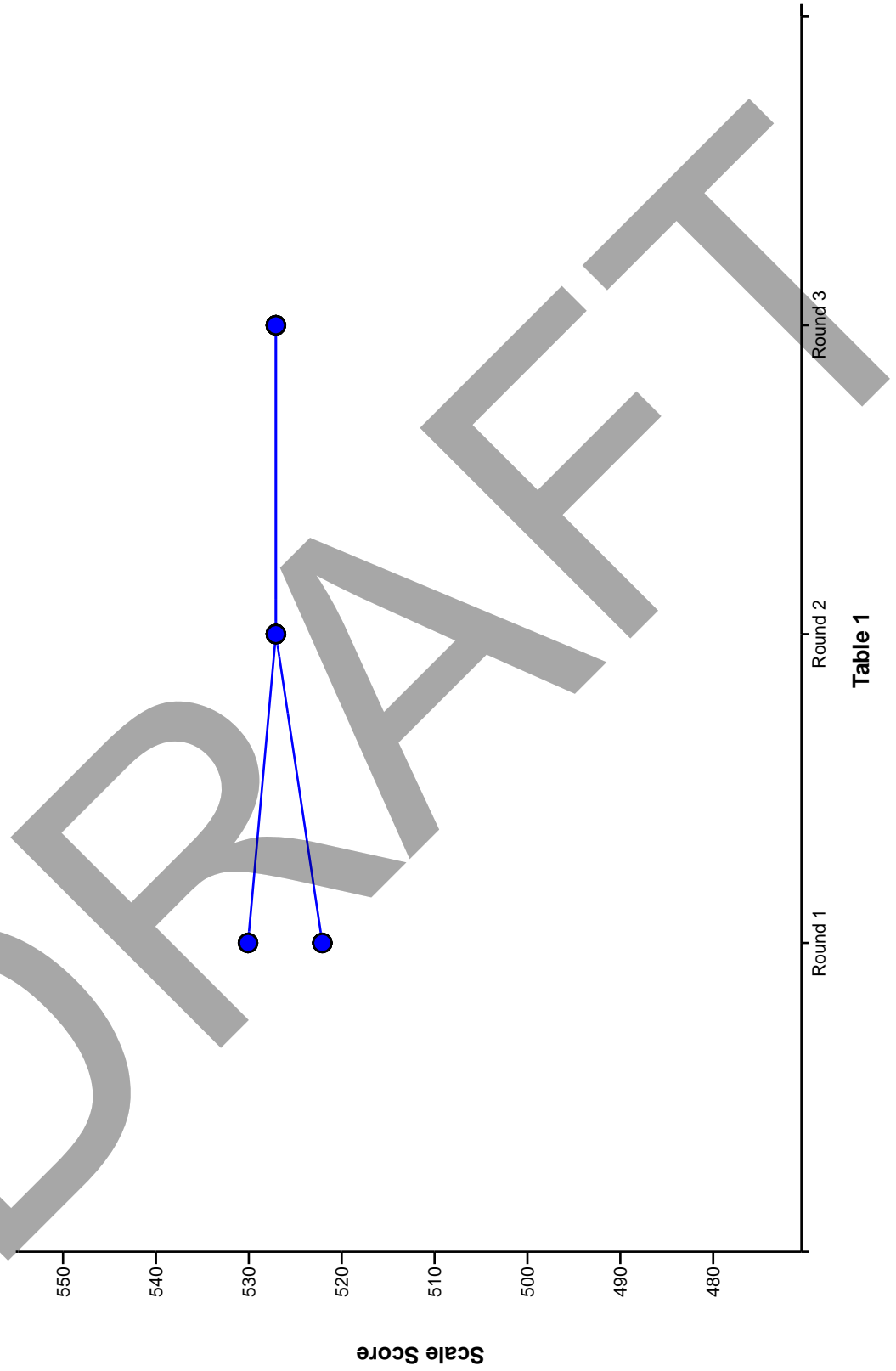


Table 1

Oregon ELPA November 2007 Grade 5 Advanced Cut Point



Oregon ELPA November 2007 Grade 7 Early Intermediate Cut Point

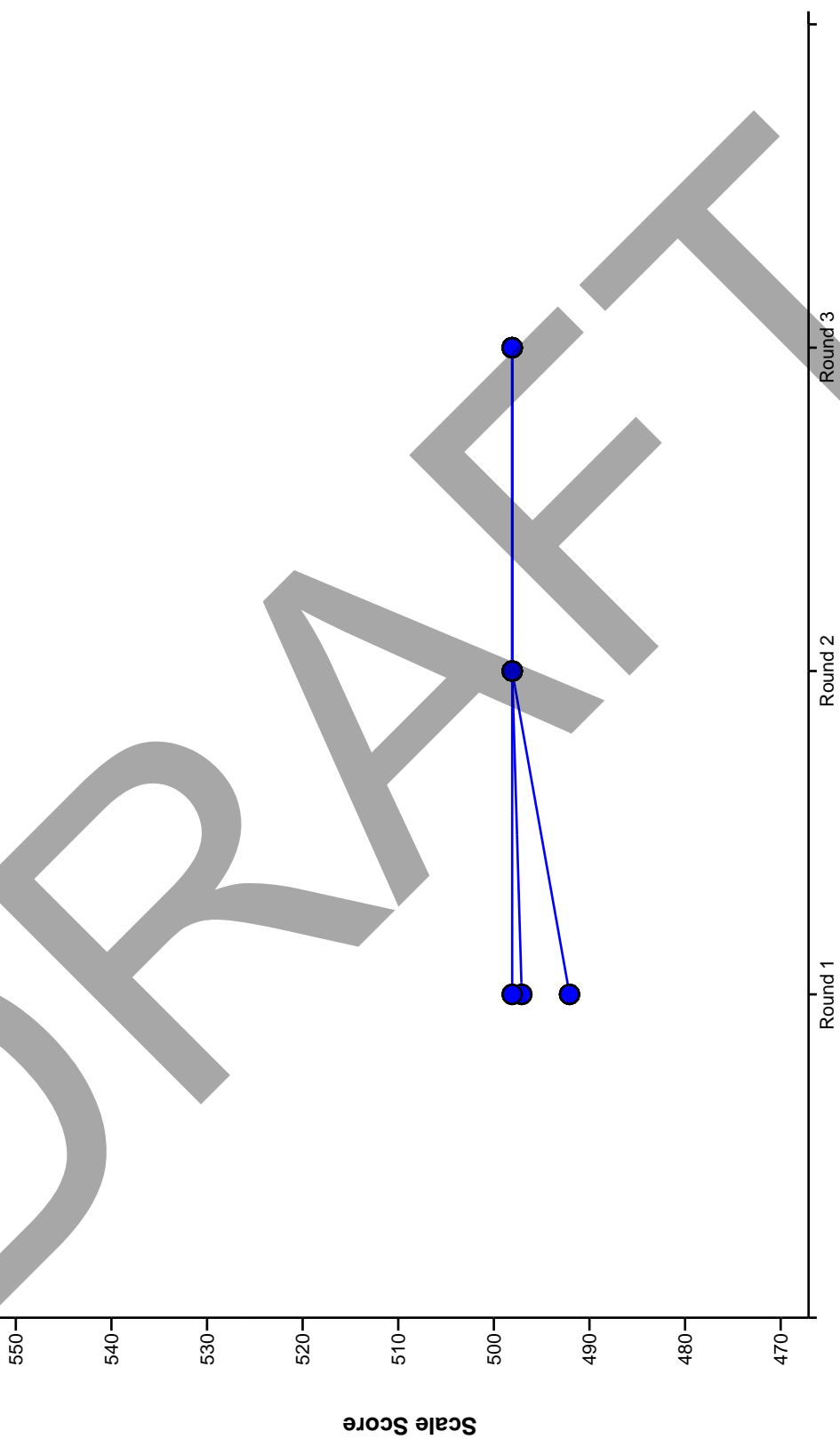


Table 1

Oregon ELPA November 2007 Grade 7 Intermediate Cut Point

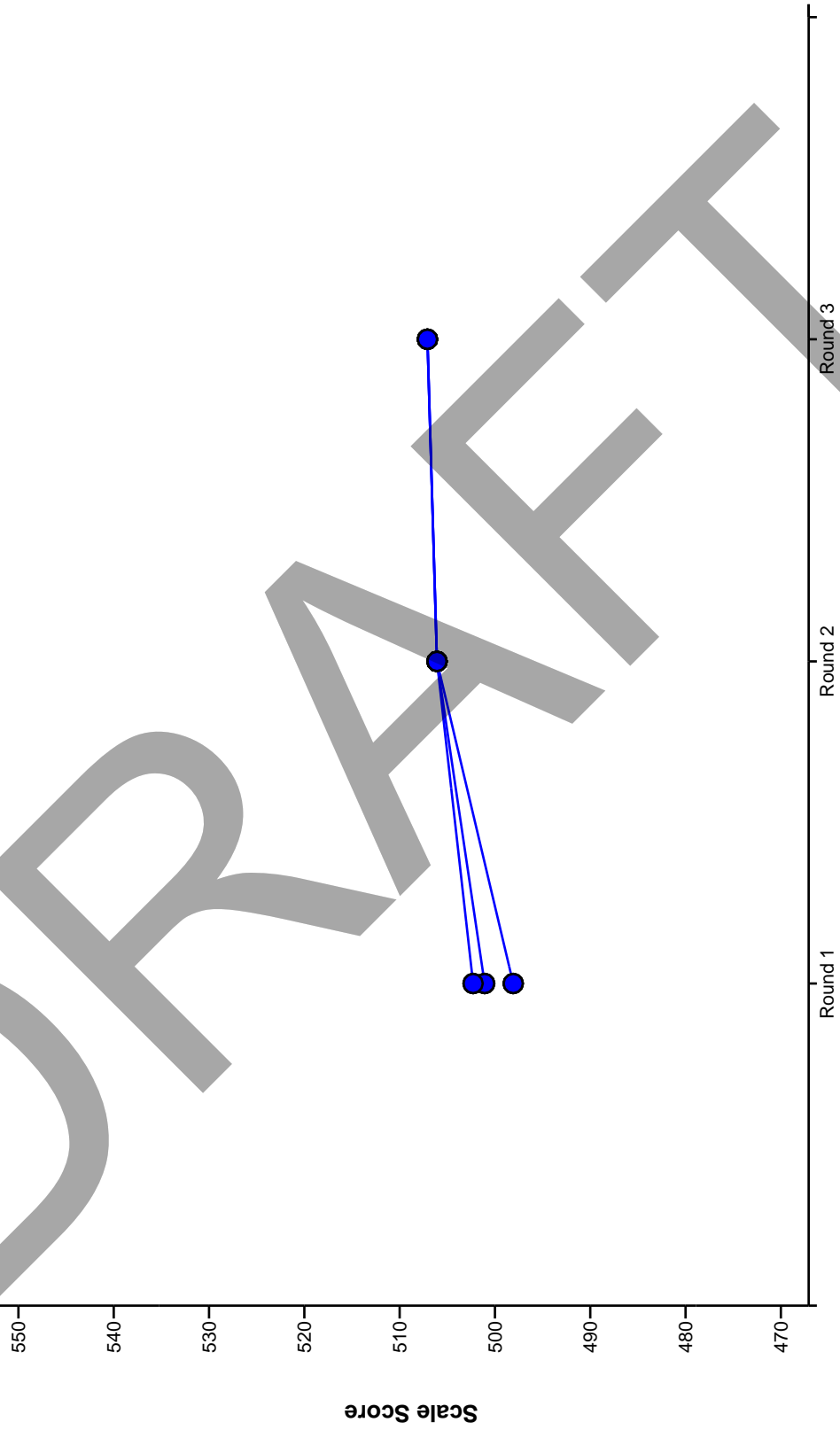


Table 1

Oregon ELPA November 2007 Grade 7 Early Advanced Cut Point

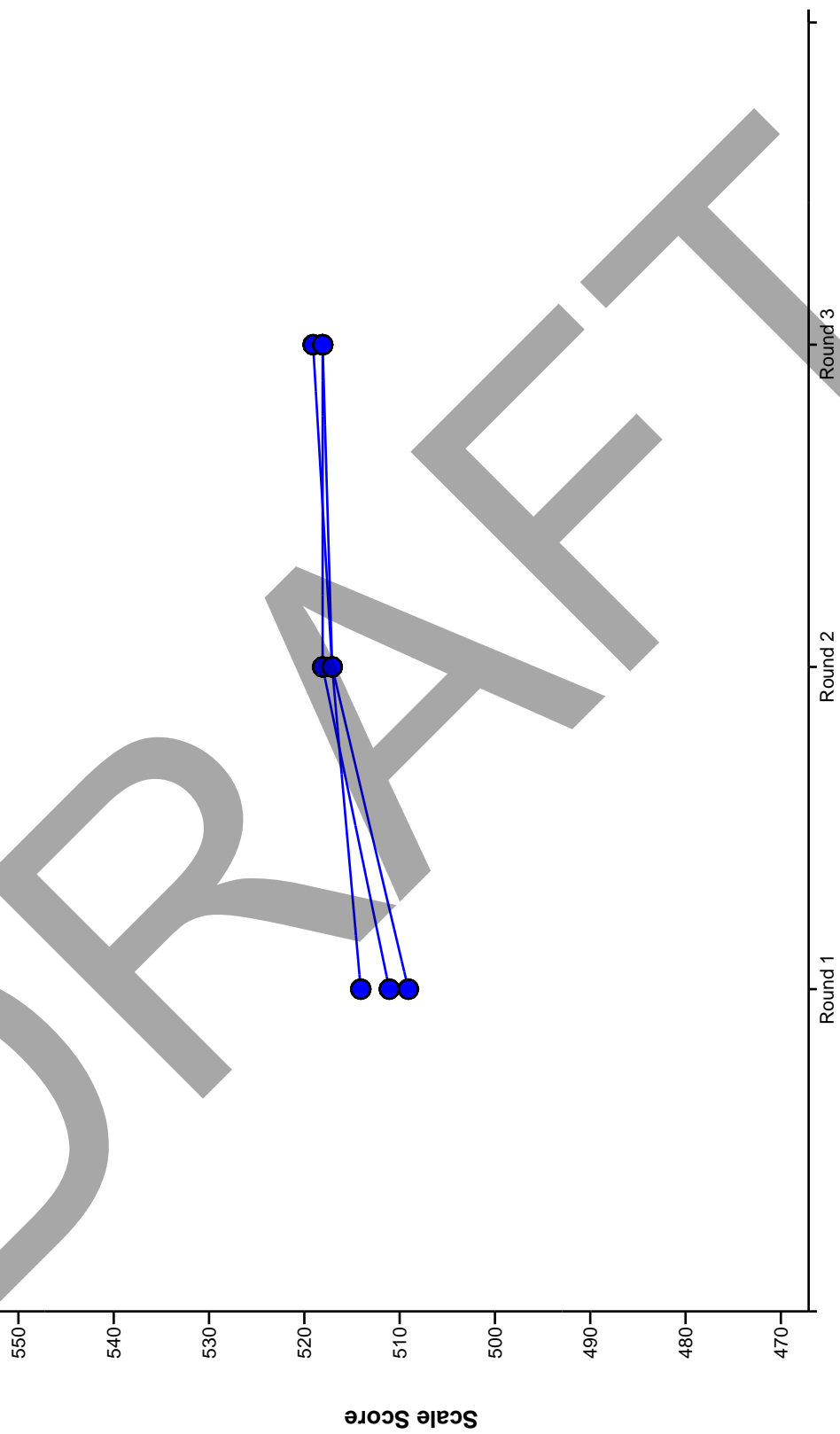


Table 1

Oregon ELPA November 2007 Grade 7 Advanced Cut Point

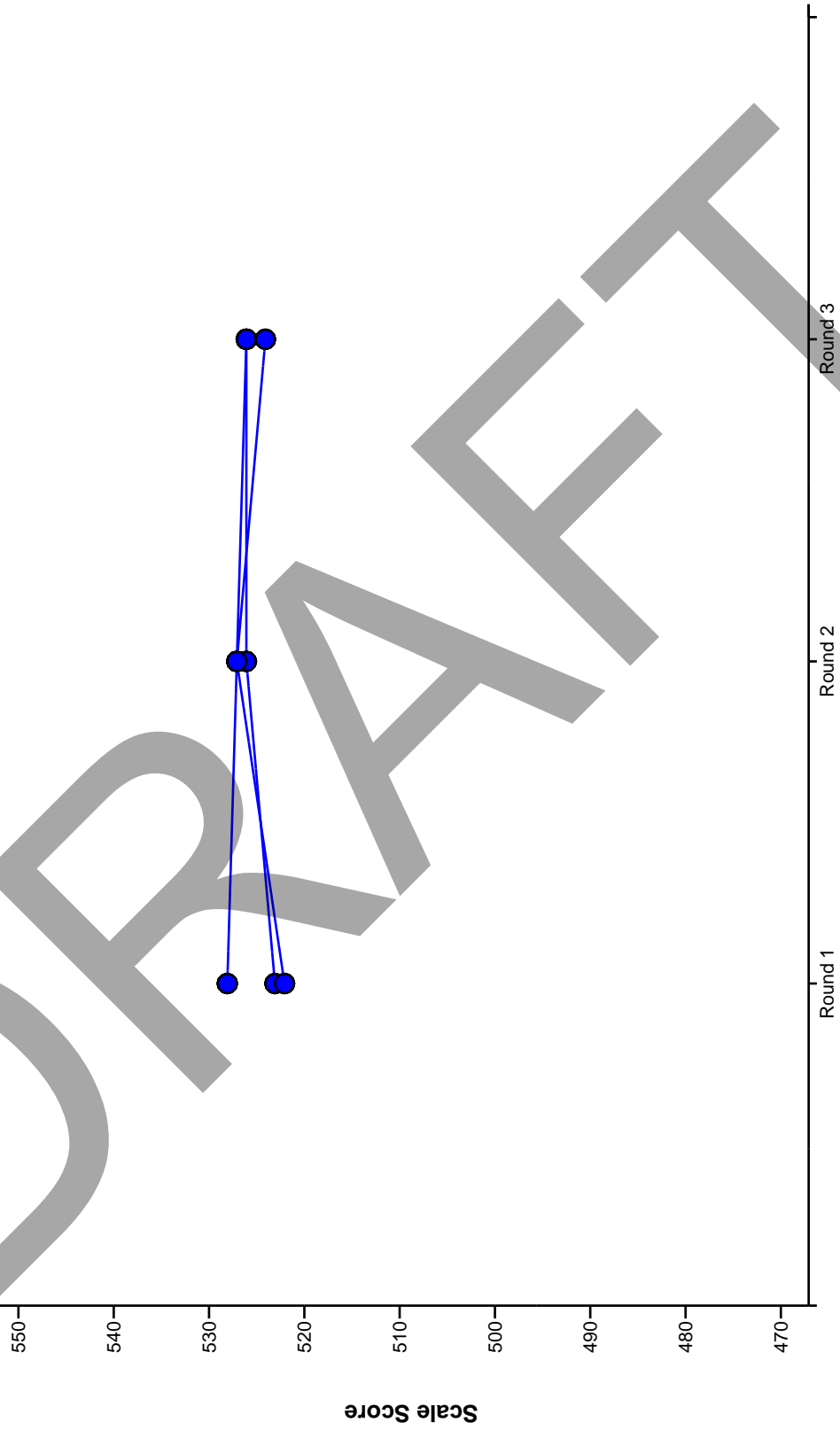
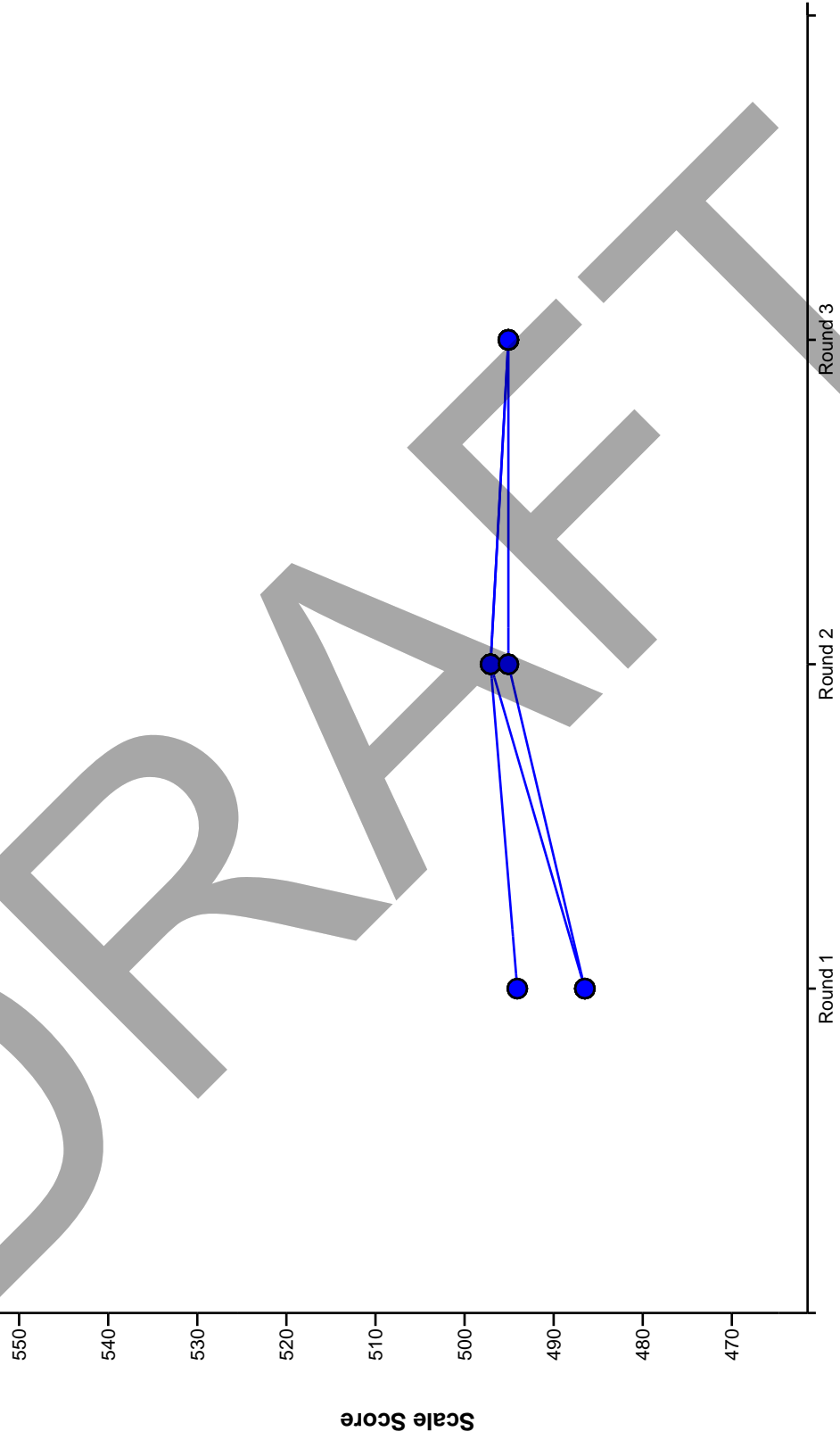


Table 1

Oregon ELPA November 2007 Grade 11 ELPA Intermediate Cut Point



Oregon ELPA November 2007 Grade 11 Intermediate Cut Point

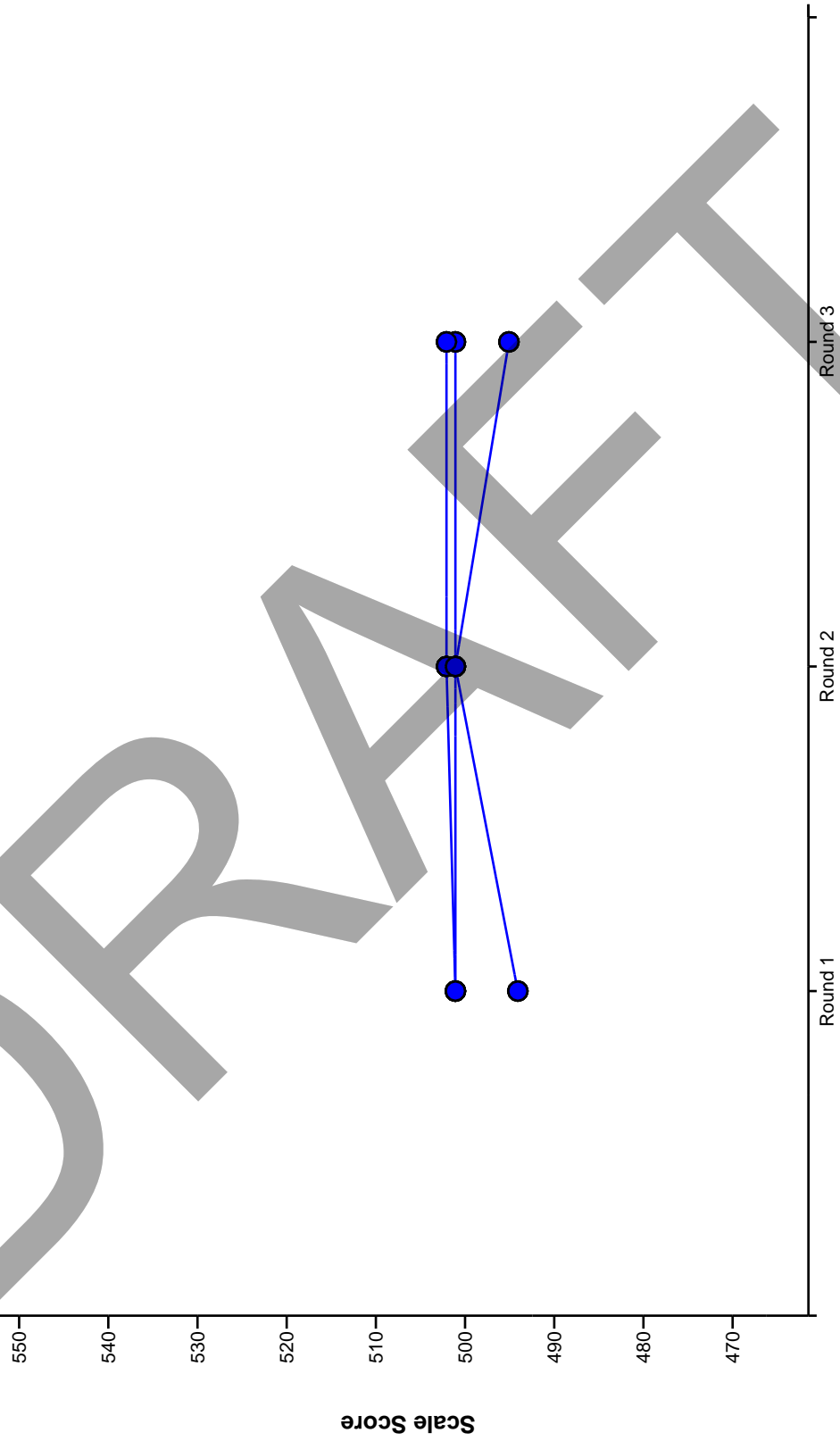


Table 1

Oregon ELPA November 2007 Grade 11 Early Advanced Cut Point

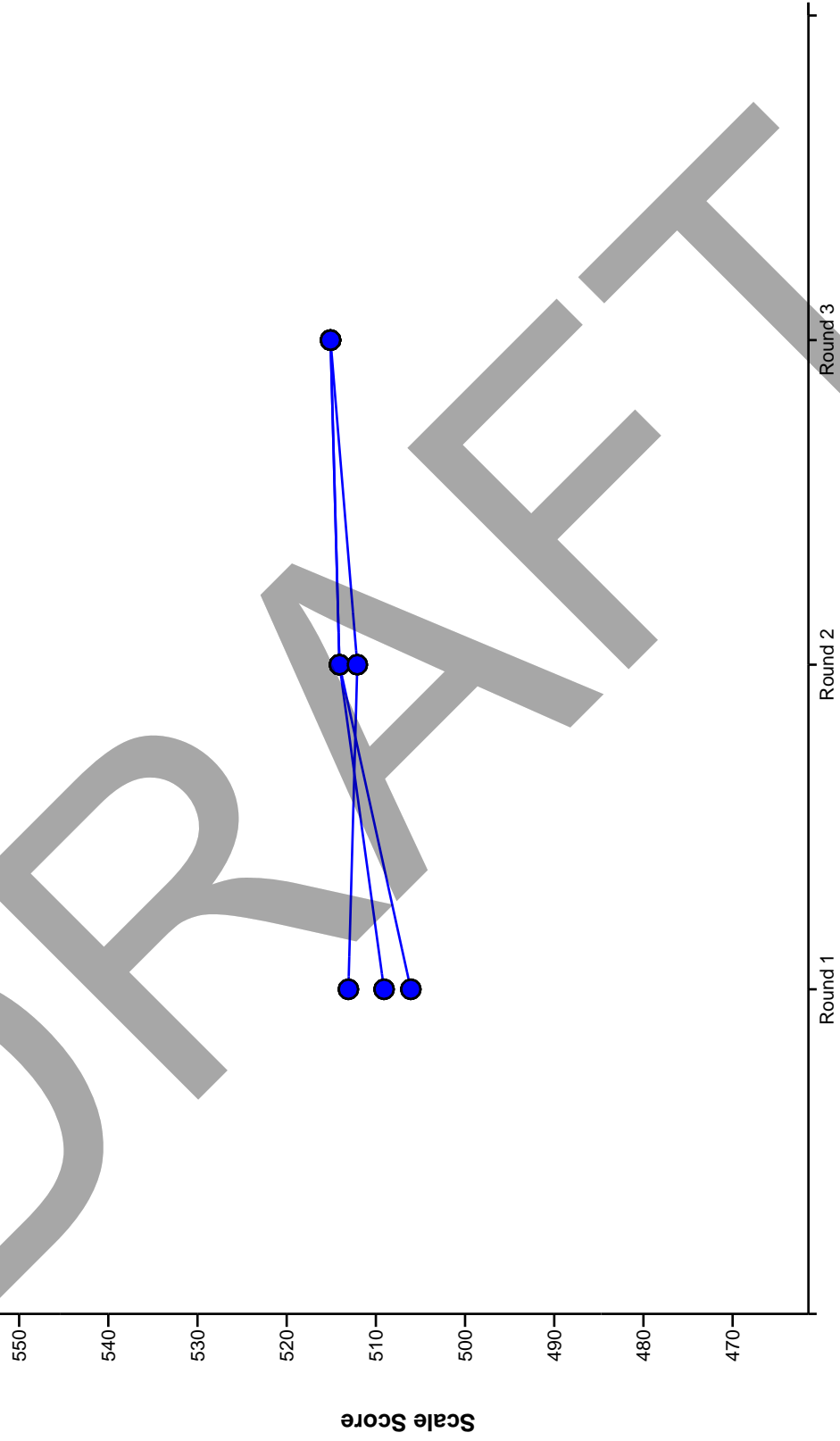
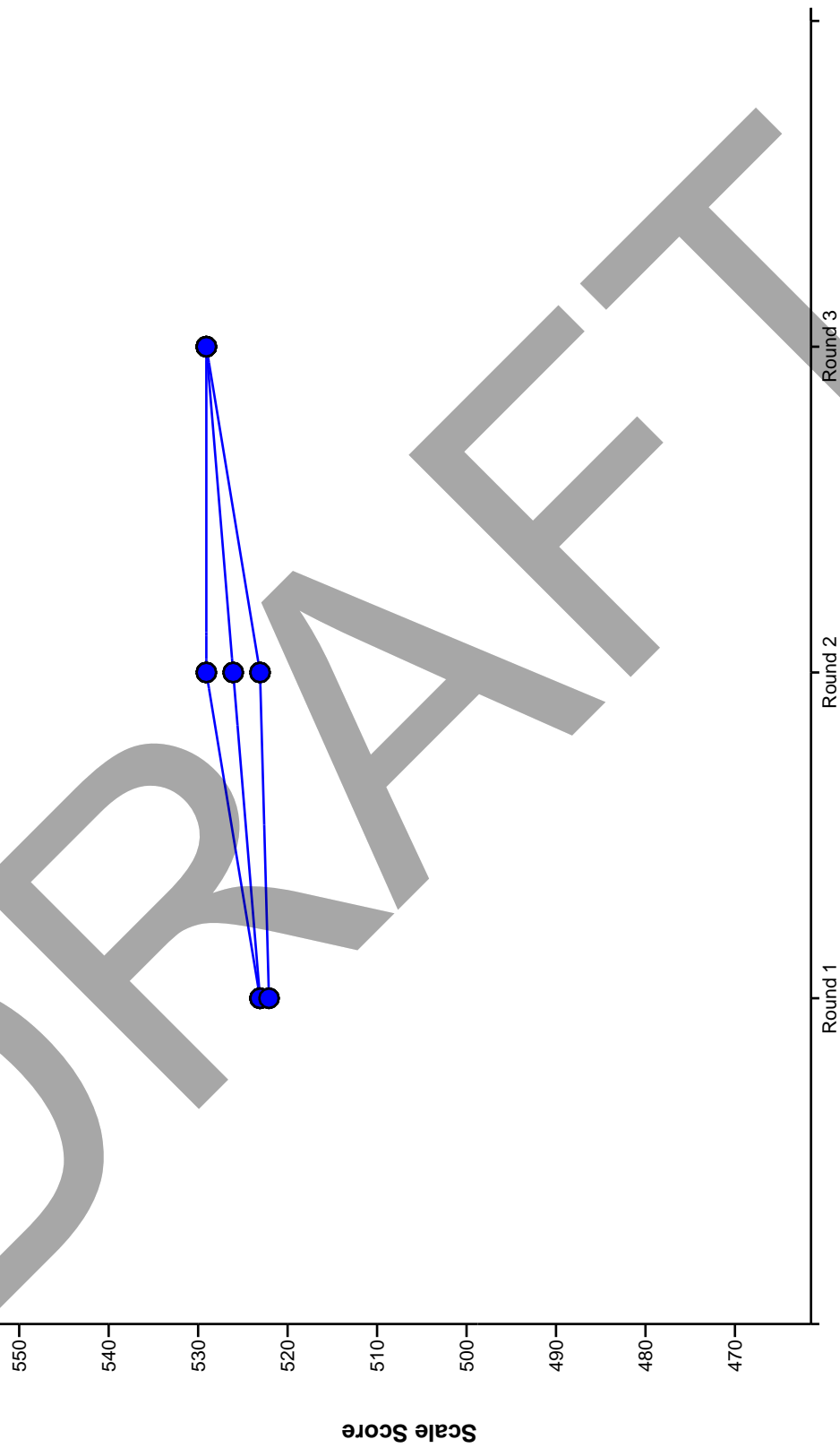


Table 1

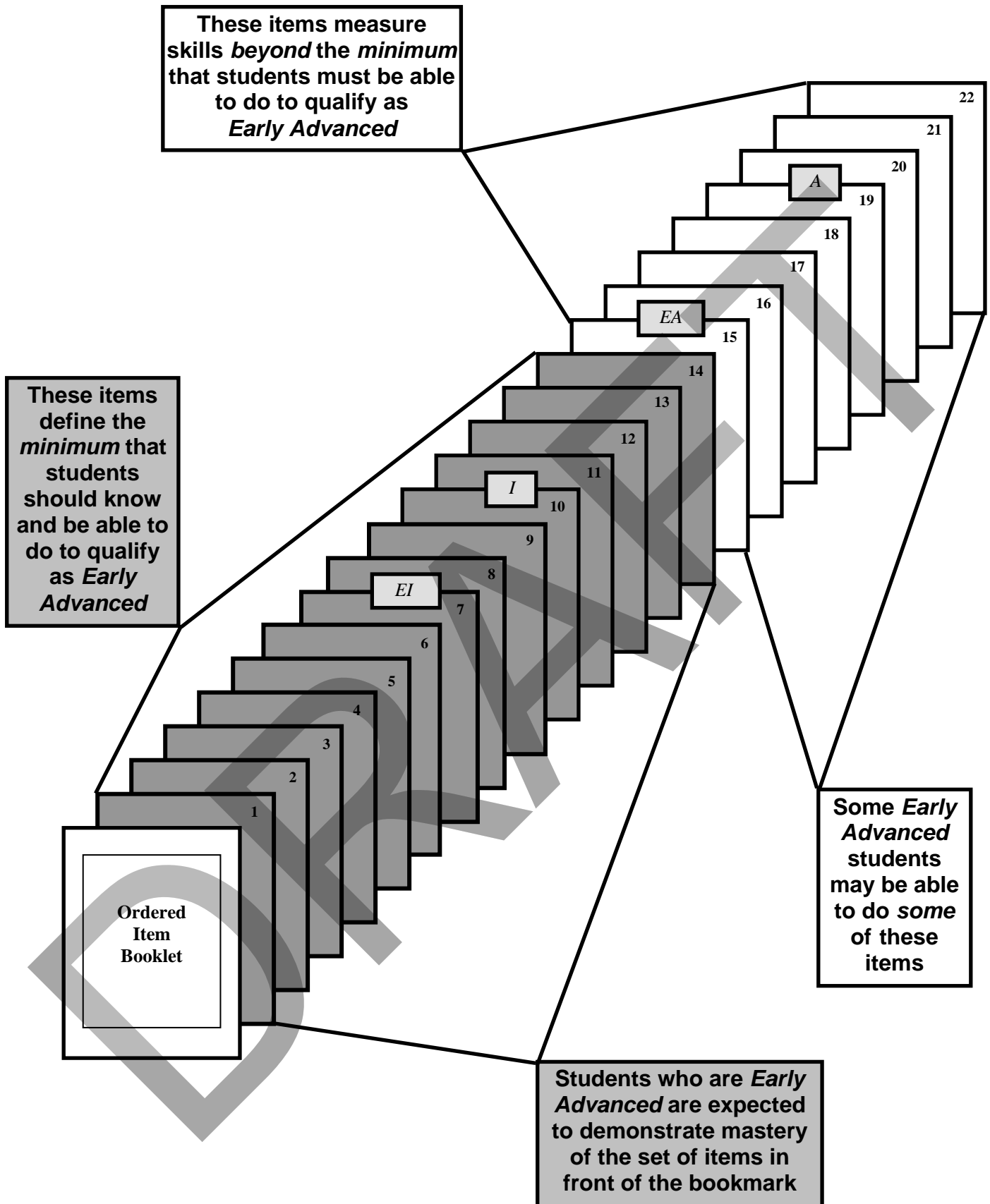
Oregon ELPA November 2007 Grade 11 Advanced Cut Point



SECTION H

Training Materials

DRAFT



Bookmark Placement

These directions are written for placing the *Early Advanced* bookmark and apply analogously to the *Early Intermediate*, *Intermediate*, and *Advanced* bookmarks.

For whom am I placing this bookmark? The Target Student

When you place your *Early Advanced* bookmark, you are separating the highest ability *Intermediate* students from the lowest ability *Early Advanced* students. In other words, you are keeping in mind the Target Student who will just make it into the *Early Advanced* level.

How do I place my bookmark? The Mechanics

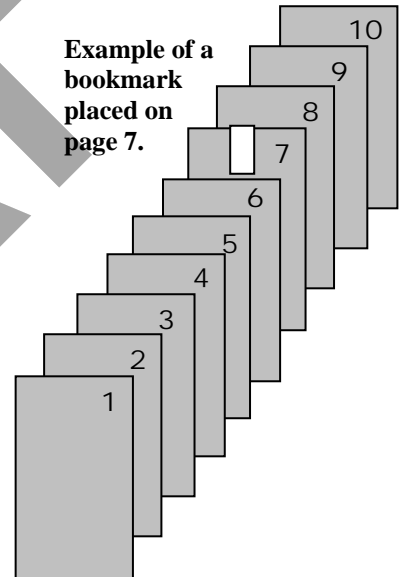
The bookmark is exactly that: a bookmark. It separates the English-language skills students are expected to master from the skills they are *not* expected to master. In the example below, a participant has placed the *Early Advanced* bookmark on page 7. With this bookmark placement, the participant says that a student must master the English-language skills represented by items 1 through 6 to be *Early Advanced*.

To place your bookmark, start at page 1 in the Ordered Item Booklet (OIB). Page through the OIB **looking at the skills covered** until you find the **first** page where you think a student has demonstrated a sufficient body of evidence to indicate that the student is *Early Advanced* relative to the standards. These are the skills you are saying an *Early Advanced* Target Student needs to master to just make it into the *Early Advanced* level.

Hold the pages that contain the English-language skills you expect the student to master in your left hand. Place your bookmark on the page **AFTER** the last item you expect the student to master. This page number is your bookmark. Write it on your Rating Form.

Hint: It may be helpful to first identify the interval of items in which you are reasonably certain the bookmark should be placed; then you can place the bookmark within that interval. If you are uncertain about where to place your bookmark, make your best decision; you will have two more rounds of voting to reconsider your bookmark.

Example of a bookmark placed on page 7.



What does my *Early Advanced* Bookmark mean? Some Answers

- You expect *Early Advanced* students to master the English-language skills contained in the items *before* your bookmark.
- *Early Advanced* students should know and be able to do the items *before* the bookmark. For multiple-choice items, *Early Advanced* students should know the correct response. For constructed-response items, *Early Advanced* students should most likely achieve the score points before the bookmark.

Is my bookmark the same as a raw score? NO

It is very important to remember that your bookmark placement is *not* equal to a raw score. In the example above, the *Early Advanced* bookmark was placed on page 7. The participant was *not* saying that a student must get six items correct to be classified as *Early Advanced*. This participant is saying that a barely *Early Advanced* student must master the English-language skills measured by the items on pages 1 through 6. The numbers in the OIB correspond to the rank order of difficulty of each item. These numbers do *not* correspond to a raw score.

Frequently Asked Questions about Bookmark Placement

These questions are written in reference to the *Early Advanced* bookmark and apply analogously to the *Early Intermediate*, *Intermediate*, and *Advanced* bookmarks.

How do I know if I placed my bookmark in the “right” place?

The “right” place is a matter of judgment, *your* judgment. You are placing your bookmark based on the content you expect students to know and be able to do.

I set my bookmark based on the content I expect students to know and be able to do, that is, the content I expect students to master. What is the definition of mastery?

We look at mastery by considering the likelihood with which students will respond correctly to the items. This question is answered in more depth in the handout “Mastery.”

If a student misses some items before the *Early Advanced* bookmark and gets some correct after the bookmark, is that student still *Early Advanced*?

A student does *not* have to get every item before the bookmark correct to be *Early Advanced*. *Early Advanced* students can miss some items *before* the bookmark and correctly respond to some items *after* the bookmark.

Does the page number on which I place my bookmark correspond to the raw score a student must get on the test?

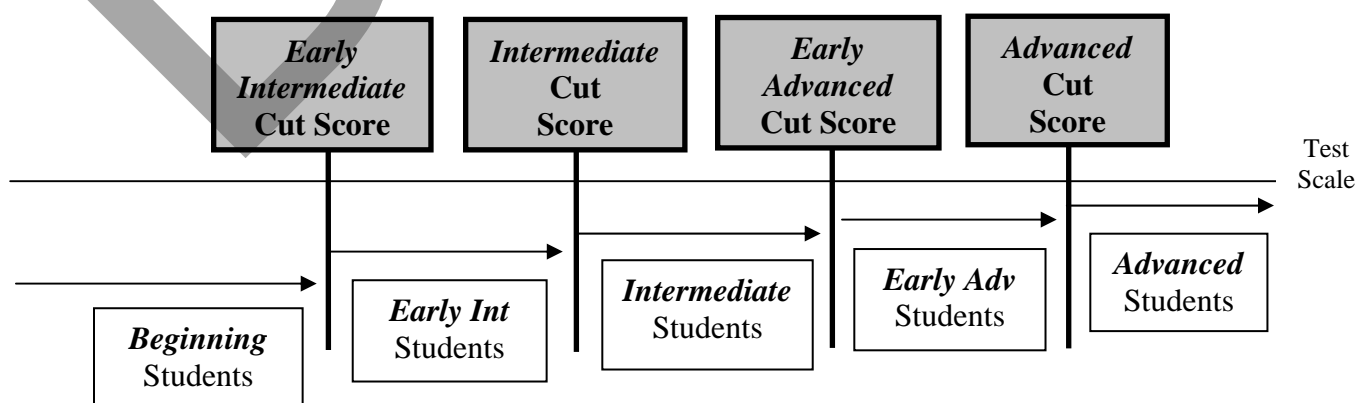
No. Remember, you are placing your bookmark based on the English-language skills you expect students to know and be able to do. You are *not* making your decision based on the number of items students must answer correctly. The bookmark is placed on a *page* in the Ordered Item Booklet. This page number corresponds to the difficulty ordering of the item, *not* to the raw score.

Should I place my bookmark in the first place in the Ordered Item Booklet where all the content standards have occurred?

Not necessarily. The test only samples the domain. In some cases, some standards will only be represented by difficult items that would be hard for most students to master.

How many bookmarks do I set?

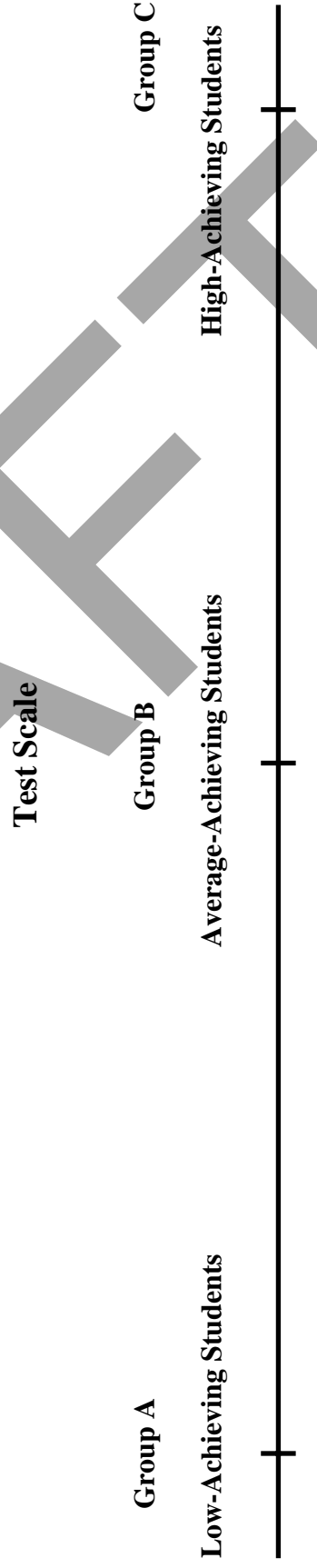
You set one less bookmark than the number of performance levels. For Oregon ELPA, you will set 4 bookmarks to separate students into 5 performance levels.



How Participants' Bookmark Judgments Relate to Expected Student Performance within Each Performance Level

You are participating in this standard setting because of your experience with students and your knowledge of the state standards, curriculum, and current instructional practices. You will be making judgments that will operationalize your expectations for the level of performance students must demonstrate in order to place in each performance level. To understand how your judgments relate to expected student performance within each performance level, consider the following examples.

Consider how students at various scale locations might perform on an imaginary assessment that consists of a total of 50 score points. In particular, we will consider the performance of groups of students who are at three specific points on the test scale. Group A consists of 100 low-achieving students, Group B consists of 100 average-achieving students, and Group C consists of 100 high-achieving students. Assume that the students have all taken the assessment and that the 100 students within each group have all obtained the exact same scale score. Note the location of the obtained scale score for each of the three groups on the test scale below.



The following three figures show how students in each of the three groups might perform on the assessment.

Figure A shows how many students in Group A responded correctly to each item in the ordered item booklet. Observe that the students in Group A performed well on the items that appear early in the ordered item booklet but performed poorly on the items that appear later in the ordered item booklet. This makes sense, because the items appear in order of difficulty, with the easiest item first and the hardest item last. For example, 99 of the 100 Group A students responded correctly to item 1, 67 of the Group A students responded correctly to item 10, but only 1 of the Group A students responded correctly to item 50.

We say that a group of like students have demonstrated mastery of the skills represented by an item if at least 2/3 of the students (about 67 out of 100) in the group can be expected to respond successfully to the item. According to Figure A, Group A students have demonstrated mastery of items 1 through 10, but have not demonstrated mastery of items 11 through 50.

Figure A. The number (or percent) of Group A students who responded correctly to each item in the ordered item booklet.

	item	item	item	item	item	item	item	item	item
	1	2	3	4	5	6	7	8	9
	10	11	12	13	14	15	16	17	18
	19	20	21	22	23	24	25	26	27
	28	29	30	31	32	33	34	35	36
	37	38	39	40	41	42	43	44	45
	46	47	48	49	50	51	52	53	54
	55	56	57	58	59	60	61	62	63
	64	65	66	67	68	69	70	71	72
	73	74	75	76	77	78	79	80	81
	82	83	84	85	86	87	88	89	90
	91	92	93	94	95	96	97	98	99
	100	100	100	100	100	100	100	100	100

[illegible][illegible]

Definition of Mastery

We say that a group of like students have demonstrated mastery of the skills represented by an item if at least 2/3 (67/100) of the students in the group can be expected to respond successfully to the item.

Figure B shows how many students in Group B responded correctly to each item in the ordered item booklet. Observe that the students in Group B performed much better than students in Group A. That makes sense because Group B students are average-achieving students while Group A students are low-achieving students.

Before you read further, use Figure B and the definition of mastery stated in the box above to determine which items Group B has mastered.

Group B students have demonstrated mastery of the skills reflected in items 1 through 30 of the ordered item booklet, but have not demonstrated mastery of the skills reflected by items 31 through 50. This is true according to the definition, because at least 67 of the 100 Group B students responded successfully to each of items 1 through 30, but fewer than 67 of them responded correctly to items 31 through 50.

Figure B. The number (or percent) of Group B students who responded correctly to each item in the ordered item booklet.

item	item	item	item	item	item	item	item	item	item
1	2	3	4	5	6	7	8	9	10
99	99	99	99	99	98	98	98	97	97
100	100	100	100	100	100	100	100	100	100

item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
96	96	95	93	89	85	84	83	83	81	79	79	78	73	72	72	71	70	69	67
100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
65	63	63	61	58	57	57	55	55	54	53	53	52	51	44	41	39	37	35	33
100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Definition of Mastery

We say that a group of like students have demonstrated mastery of the skills represented by an item if at least 2/3 (67/100) of the students in the group can be expected to respond successfully to the item.

Figure C shows how many students in Group C responded correctly to each item in the ordered item booklet. Observe that Group C performed much better than Groups A or B. That makes sense because Group C consists of high-achieving students while Groups A and B consist of low-and average-achieving students, respectively.

Before you read further, use Figure C and the definition of mastery stated in the box above to determine which items Group C has mastered. Group C students have demonstrated mastery of the skills reflected in items 1 through 45 of the ordered item booklet, but have not demonstrated mastery of the skills reflected by items 46 through 50. This is true according to the definition, because at least 67 of the 100 Group C students responded successfully to each of items 1 through 45, but fewer than 67 of them responded correctly to items 46 through 50.

Figure C. The number (or percent) of Group C students who responded correctly to each item in the ordered item booklet.

item	item	item	item	item	item	item	item	item	item
1	2	3	4	5	6	7	8	9	10
99	99	99	99	99	99	99	99	97	97
100	100	100	100	100	100	100	100	100	100

item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
97	97	95	95	94	93	92	92	91	89	89	89	88	88	88	87	87	86	85	84
100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item	item
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
83	81	81	81	80	80	79	78	77	75	74	72	70	68	67	64	58	53	49	46
100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

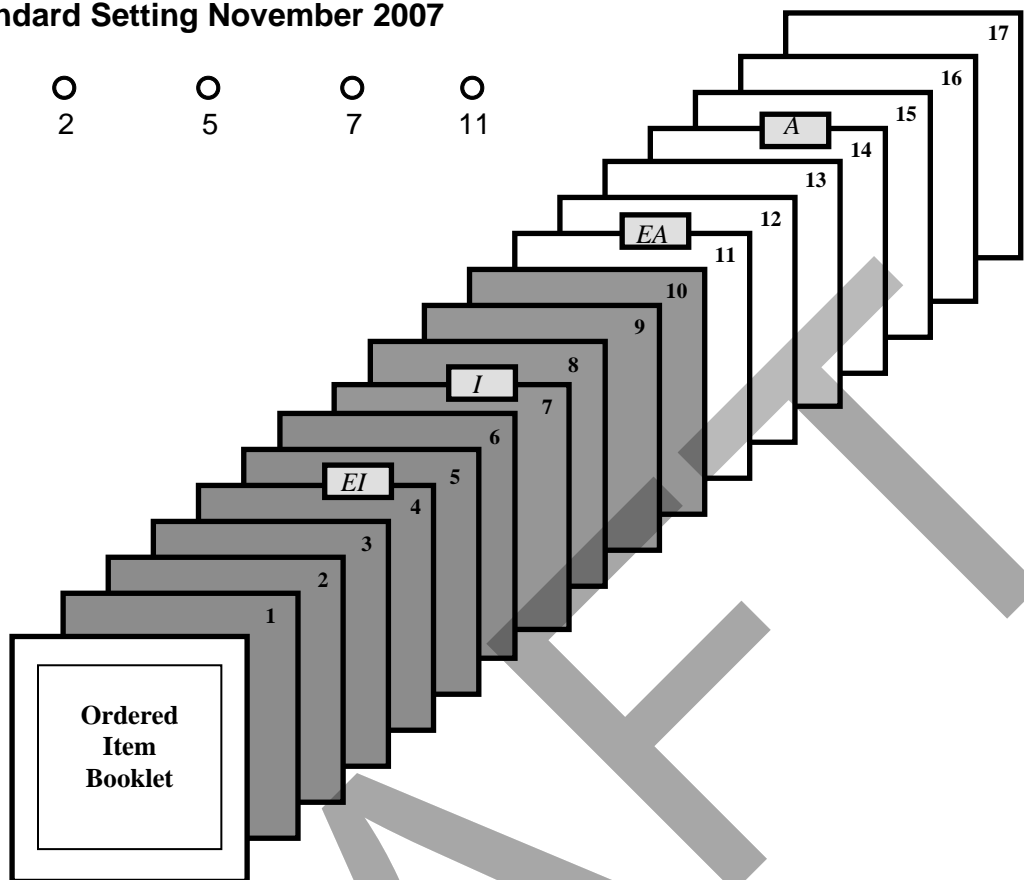
You have seen from the above examples that by using a specific definition of mastery, we can identify the skills in the ordered item booklet that students at any location of the test scale have mastered.

Also, if *you* identify a set of items in the ordered item booklet, the specific point on the test scale at which students have demonstrated mastery of the skills you have specified can be determined. This is how the various cut scores are ascertained.

As experts, you will first specify the skills in the ordered item booklet that you expect students to master in order to be classified as ***Early Advanced***. This means that you will identify the items that reflect the knowledge, skills, and abilities you expect all ***Early Advanced*** students to master. When you have made that judgment, the point on the scale at which students achieve that level of mastery can be identified.

Oregon ELPA Standard Setting November 2007

Grade: ☐ K ☐ 2 ☐ 5 ☐ 7 ☐ 11



Suppose the bookmarks were placed in this sample ordered item booklet as follows:

	<i>Early Intermediate</i> Bookmark on Page #	<i>Intermediate</i> Bookmark on Page #	<i>Early Advanced</i> Bookmark on Page #	<i>Advanced</i> Bookmark on Page #
Round 1	4	7	11	14

- Which items does a student need to master to just make it into the *Early Advanced* level?

☐ 1 to 6 ☐ 1 to 7 ☐ 1 to 10 ☐ 1 to 11
- If a student mastered only items 1 through 2, in which performance level would this student be?

☐ *Beginning* ☐ *Early Intermediate* ☐ *Intermediate* ☐ *Early Advanced* ☐ *Advanced*
- Suppose a student mastered items 1 through 13. Which performance level is this student in?

☐ *Beginning* ☐ *Early Intermediate* ☐ *Intermediate* ☐ *Early Advanced* ☐ *Advanced*
- For students who are classified as *Early Advanced*, with at least what likelihood will they be able to answer item 10?

☐ 1/3 ☐ 1/2 ☐ 2/3 ☐ 3/4
- Will the items BEFORE the *Early Advanced* bookmark be more or less difficult to answer than the items AFTER the bookmark or about the same?

☐ More difficult to answer ☐ About the same ☐ Less difficult to answer

SECTION I

Participant Evaluation

DRAFT

Evaluation - Oregon ELPA Standard Setting – November 2007

Key: SD=Strongly Disagree D=Disagree N=Neutral A=Agree SA=Strongly Agree	SD	D	N	A	SA
1. The Bookmark Procedure was well described.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. The training on bookmark placement made the task clear to me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. The training materials were helpful.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. The goals for the Bookmark Procedure were clear.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Reviewing the test items helped me place my bookmarks.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. The ordering of the items in the ordered item booklet agreed with my perception of the relative difficulty of the items.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Reviewing the Target Students helped me place my bookmarks.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. I considered the ELP standards when I placed my bookmarks.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. During Round 1, I placed my bookmarks without consulting other participants.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. I had enough time to consider my Round 1 bookmarks.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. Overall, my table's discussions were open and honest.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. Overall, I believe that my opinions were considered and valued by my group.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13. The presentation of impact data was helpful to me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14. I learned how to do the bookmark placement as I went along, so my later ones may not be comparable to my earlier ones.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15. I understood how to place my bookmarks.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
16. Overall, I am satisfied with my group's final bookmarks.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
17. I feel this procedure was fair.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
18. I am confident that the Bookmark Procedure produced valid standards.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
19. I would defend the <i>Early Intermediate</i> cut score against criticism that it is too high.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20. I would defend the <i>Early Intermediate</i> cut score against criticism that it is too low.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
21. I would defend the <i>Intermediate</i> cut score against criticism that it is too high.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
22. I would defend the <i>Intermediate</i> cut score against criticism that it is too low.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
23. I would defend the <i>Early Advanced</i> cut score against criticism that it is too high.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
24. I would defend the <i>Early Advanced</i> cut score against criticism that it is too low.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
25. I would defend the <i>Advanced</i> cut score against criticism that it is too high.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
26. I would defend the <i>Advanced</i> cut score against criticism that it is too low.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
27. Participating in the Bookmark Procedure increased my understanding of the test.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
28. This experience will help me target instruction for the students in my classroom.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
29. Overall, I valued the workshop as a professional development experience.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
30. The standard setting was well organized.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

31. What is your occupation?	<input type="radio"/> Teacher <input type="radio"/> Administrator <input type="radio"/> Instructional Assistant <input type="radio"/> Other: _____	32. How many years in your current profession?	<input type="radio"/> 1-5 <input type="radio"/> 6-10 <input type="radio"/> 11-15 <input type="radio"/> 16-20 <input type="radio"/> 21+
33. What is your education level?	<input type="radio"/> HSD or GED <input type="radio"/> Bachelor's <input type="radio"/> Master's <input type="radio"/> Doctorate	34. What is your gender?	<input type="radio"/> Male <input type="radio"/> Female
35. What is your race?	<input type="radio"/> Asian/Pacific Islander <input type="radio"/> Black/African-American <input type="radio"/> American Indian <input type="radio"/> White <input type="radio"/> Other	36. Are you of Hispanic origin?	<input type="radio"/> Yes <input type="radio"/> No
37. Have you taught Special Education?	<input type="radio"/> Yes <input type="radio"/> No	38. Have you taught ESL/ELD?	<input type="radio"/> Yes <input type="radio"/> No
39. Have you taught Vocational Education?	<input type="radio"/> Yes <input type="radio"/> No	40. Have you taught Alternative Education?	<input type="radio"/> Yes <input type="radio"/> No
41. Have you taught Adult Education?	<input type="radio"/> Yes <input type="radio"/> No	42. What is your primary role at this standard setting?	<input type="radio"/> Educator <input type="radio"/> Parent <input type="radio"/> Community Member <input type="radio"/> Business Member
43. Which grade did you work on during this standard setting?	<input type="radio"/> K-1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9 <input type="radio"/> 10 <input type="radio"/> 11		

Please add your comments on the back of this evaluation. Thank you!

Oregon ELPA Standard Setting November 2007 Evaluation Results

About these results

Each question is shown, along with its answer choices and associated response percentages. For Likert-type questions, there are five possible responses: "Strongly Disagree," "Disagree," "Neutral," "Agree," and "Strongly Agree." For each question, the number of respondents is shown in the column labeled "N."

Question 1

The Bookmark Procedure was well described.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	0.0%	42.9%	57.1%	100.0%
Kindergarten and 1	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
2	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
5	2	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%

Question 2

The training on bookmark placement made the task clear to me.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	7.1%	7.1%	35.7%	50.0%	85.7%
Kindergarten and 1	3	0.0%	33.3%	33.3%	0.0%	33.3%	33.3%
2	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
5	2	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
7	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
11	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%

Question 3

The training materials were helpful.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	7.1%	57.1%	35.7%	92.8%
Kindergarten and 1	3	0.0%	0.0%	33.3%	33.3%	33.3%	66.6%
2	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%

Question 4

The goals for the Bookmark Procedure were clear.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	21.4%	28.6%	50.0%	78.6%
Kindergarten and 1	3	0.0%	0.0%	33.3%	0.0%	66.7%	66.7%
2	3	0.0%	0.0%	33.3%	33.3%	33.3%	66.6%
5	2	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
7	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
11	3	0.0%	0.0%	33.3%	33.3%	33.3%	66.6%

Question 5

Reviewing the test items helped me place my bookmarks.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	7.1%	42.9%	50.0%	92.9%
Kindergarten and 1	3	0.0%	0.0%	33.3%	0.0%	66.7%	66.7%
2	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
5	2	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Question 6

The ordering of the items in the ordered item booklet agreed with my perception of the relative difficulty of the items.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	7.1%	21.4%	35.7%	28.6%	7.1%	35.7%
Kindergarten and 1	3	33.3%	33.3%	33.3%	0.0%	0.0%	0.0%
2	3	0.0%	0.0%	66.7%	33.3%	0.0%	33.3%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
7	3	0.0%	33.3%	66.7%	0.0%	0.0%	0.0%
11	3	0.0%	33.3%	0.0%	33.3%	33.3%	66.6%

Question 7

Reviewing the Target Student helped me place my bookmarks.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	14.3%	64.3%	21.4%	85.7%
Kindergarten and 1	3	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
2	3	0.0%	0.0%	33.3%	33.3%	33.3%	66.6%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
7	3	0.0%	0.0%	33.3%	33.3%	33.3%	66.6%
11	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%

Question 8

I considered the ELP standards when I placed my bookmarks.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
Kindergarten and 1	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
2	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
5	2	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%

Question 9

During Round 1, I placed my bookmarks without consulting other participants.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	0.0%	35.7%	64.3%	100.0%
Kindergarten and 1	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
2	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
5	2	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%
7	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
11	3	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%

Question 10

I had enough time to consider my Round 1 bookmarks.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	7.1%	21.4%	7.1%	28.6%	35.7%	64.3%
Kindergarten and 1	3	0.0%	33.3%	0.0%	33.3%	33.3%	66.6%
2	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
5	2	0.0%	50.0%	50.0%	0.0%	0.0%	0.0%
7	3	33.3%	33.3%	0.0%	33.3%	0.0%	33.3%
11	3	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%

Question 11

Overall, my table's discussions were open and honest.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	0.0%	21.4%	78.6%	100.0%
Kindergarten and 1	3	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%
2	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
5	2	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%
7	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Question 12

Overall, I believe that my opinions were considered and valued by my group.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	0.0%	35.7%	64.3%	100.0%
Kindergarten and 1	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
2	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
5	2	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Question 13

The presentation of different types of impact data was helpful to me.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
Kindergarten and 1	3	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%
2	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Question 14

I learned how to do the bookmark placement as I went along, so my later ones may not be comparable to my earlier ones.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	7.1%	28.6%	7.1%	35.7%	21.4%	57.1%
Kindergarten and 1	3	33.3%	0.0%	0.0%	33.3%	33.3%	66.6%
2	3	0.0%	33.3%	0.0%	33.3%	33.3%	66.6%
5	2	0.0%	50.0%	0.0%	50.0%	0.0%	50.0%
7	3	0.0%	33.3%	33.3%	33.3%	0.0%	33.3%
11	3	0.0%	33.3%	0.0%	33.3%	33.3%	66.6%

Question 15

I understood how to place my bookmarks.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
Kindergarten and 1	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
2	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
5	2	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
7	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Question 16

Overall, I am satisfied with my group's final bookmarks.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	7.1%	42.9%	50.0%	92.9%
Kindergarten and 1	3	0.0%	0.0%	33.3%	66.7%	0.0%	66.7%
2	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
5	2	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Question 17

I feel this procedure was fair.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	7.1%	50.0%	42.9%	92.9%
Kindergarten and 1	3	0.0%	0.0%	33.3%	66.7%	0.0%	66.7%
2	3	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Question 18

I am confident that the Bookmark Procedure produced valid standards.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	7.1%	14.3%	71.4%	7.1%	78.5%
Kindergarten and 1	3	0.0%	33.3%	33.3%	33.3%	0.0%	33.3%
2	3	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
7	3	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
11	3	0.0%	0.0%	33.3%	33.3%	33.3%	66.6%

Question 19

I would defend the *Early Intermediate* cut score against criticism that it is too high.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	7.1%	7.1%	50.0%	35.7%	85.7%
Kindergarten and 1	3	0.0%	33.3%	0.0%	33.3%	33.3%	66.6%
2	3	0.0%	0.0%	33.3%	66.7%	0.0%	66.7%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%

Question 20

I would defend the *Early Intermediate* cut score against criticism that it is too low.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	7.1%	7.1%	57.1%	28.6%	85.7%
Kindergarten and 1	3	0.0%	33.3%	0.0%	33.3%	33.3%	66.6%
2	3	0.0%	0.0%	33.3%	66.7%	0.0%	66.7%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Question 21

I would defend the *Intermediate* cut score against criticism that it is too high.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	7.1%	7.1%	64.3%	21.4%	85.7%
Kindergarten and 1	3	0.0%	33.3%	0.0%	33.3%	33.3%	66.6%
2	3	0.0%	0.0%	33.3%	66.7%	0.0%	66.7%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
7	3	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Question 22

I would defend the *Intermediate* cut score against criticism that it is too low.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	7.1%	7.1%	64.3%	21.4%	85.7%
Kindergarten and 1	3	0.0%	33.3%	0.0%	33.3%	33.3%	66.6%
2	3	0.0%	0.0%	33.3%	66.7%	0.0%	66.7%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
7	3	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Question 23

I would defend the *Early Advanced* cut score against criticism that it is too high.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	14.3%	7.1%	57.1%	21.4%	78.5%
Kindergarten and 1	3	0.0%	66.7%	0.0%	0.0%	33.3%	33.3%
2	3	0.0%	0.0%	33.3%	66.7%	0.0%	66.7%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
7	3	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Question 24

I would defend the *Early Advanced* cut score against criticism that it is too low.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	7.1%	14.3%	50.0%	28.6%	78.6%
Kindergarten and 1	3	0.0%	33.3%	33.3%	0.0%	33.3%	33.3%
2	3	0.0%	0.0%	33.3%	66.7%	0.0%	66.7%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
7	3	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%
11	3	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%

Question 25

I would defend the *Advanced* cut score against criticism that it is too high.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	7.1%	7.1%	57.1%	28.6%	85.7%
Kindergarten and 1	3	0.0%	33.3%	0.0%	66.7%	0.0%	66.7%
2	3	0.0%	0.0%	33.3%	66.7%	0.0%	66.7%
5	2	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Question 26

I would defend the *Advanced* cut score against criticism that it is too low.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	13	0.0%	15.4%	7.7%	38.5%	38.5%	77.0%
Kindergarten and 1	3	0.0%	66.7%	0.0%	33.3%	0.0%	33.3%
2	3	0.0%	0.0%	33.3%	66.7%	0.0%	66.7%
5	2	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
7	2	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
11	3	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%

Question 27

Participating in the Bookmark Procedure increased my understanding of the test.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	7.1%	35.7%	57.1%	92.8%
Kindergarten and 1	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
2	3	0.0%	0.0%	33.3%	0.0%	66.7%	66.7%
5	2	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
7	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
11	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%

Question 28

This experience will help me target instruction for the students in my classroom.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	21.4%	50.0%	28.6%	78.6%
Kindergarten and 1	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
2	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
5	2	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%
7	3	0.0%	0.0%	33.3%	66.7%	0.0%	66.7%
11	3	0.0%	0.0%	66.7%	0.0%	33.3%	33.3%

Question 29

Overall, I valued the conference as a professional development experience.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	7.1%	50.0%	42.9%	92.9%
Kindergarten and 1	3	0.0%	0.0%	33.3%	33.3%	33.3%	66.6%
2	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
5	2	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
11	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%

Question 30

The standard setting was well organized.

Grade	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree + Strongly Agree
Overall	14	0.0%	0.0%	14.3%	50.0%	35.7%	85.7%
Kindergarten and 1	3	0.0%	0.0%	33.3%	33.3%	33.3%	66.6%
2	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%
5	2	0.0%	0.0%	50.0%	50.0%	0.0%	50.0%
7	3	0.0%	0.0%	0.0%	33.3%	66.7%	100.0%
11	3	0.0%	0.0%	0.0%	66.7%	33.3%	100.0%

Question 31

What is your occupation?

Grade	N	Teacher	Administrator	Instructional Assistant	Other
Overall	14	50.0%	7.1%	0.0%	42.9%
Kindergarten and 1	3	66.7%	0.0%	0.0%	33.3%
2	3	33.3%	33.3%	0.0%	33.3%
5	2	50.0%	0.0%	0.0%	50.0%
7	3	100.0%	0.0%	0.0%	0.0%
11	3	0.0%	0.0%	0.0%	100.0%

Question 32

How many years in your current profession?

Grade	N	1-5	6-10	11-15	16-20	21+
Overall	13	15.4%	15.4%	30.8%	7.7%	30.8%
Kindergarten and 1	3	0.0%	0.0%	66.7%	0.0%	33.3%
2	2	50.0%	50.0%	0.0%	0.0%	0.0%
5	2	0.0%	0.0%	100.0%	0.0%	0.0%
7	3	0.0%	33.3%	0.0%	0.0%	66.7%
11	3	33.3%	0.0%	0.0%	33.3%	33.3%

Question 33

What is your education level?

Grade	N	HSD or GED	Bachelor's	Master's	Doctorate
Overall	14	0.0%	7.1%	85.7%	7.1%
Kindergarten and 1	3	0.0%	0.0%	100.0%	0.0%
2	3	0.0%	0.0%	100.0%	0.0%
5	2	0.0%	50.0%	50.0%	0.0%
7	3	0.0%	0.0%	100.0%	0.0%
11	3	0.0%	0.0%	66.7%	33.3%

Question 34

What is your gender?

Grade	N	Male	Female
Overall	14	14.3%	85.7%
Kindergarten and 1	3	0.0%	100.0%
2	3	0.0%	100.0%
5	2	0.0%	100.0%
7	3	0.0%	100.0%
11	3	66.7%	33.3%

Question 35

What is your race?

Grade	N	Asian/ Pacific Islander	Black/ African- American	American Indian	White	Other
Overall	14	0.0%	0.0%	0.0%	78.6%	21.4%
Kindergarten and 1	3	0.0%	0.0%	0.0%	100.0%	0.0%
2	3	0.0%	0.0%	0.0%	33.3%	66.7%
5	2	0.0%	0.0%	0.0%	100.0%	0.0%
7	3	0.0%	0.0%	0.0%	66.7%	33.3%
11	3	0.0%	0.0%	0.0%	100.0%	0.0%

Question 36

Are you of Hispanic origin?

Grade	N	Yes	No
Overall	14	21.4%	78.6%
Kindergarten and 1	3	0.0%	100.0%
2	3	66.7%	33.3%
5	2	0.0%	100.0%
7	3	33.3%	66.7%
11	3	0.0%	100.0%

Question 37

Have you taught Special Education?

Grade	N	Yes	No
Overall	14	28.6%	71.4%
Kindergarten and 1	3	66.7%	33.3%
2	3	0.0%	100.0%
5	2	0.0%	100.0%
7	3	66.7%	33.3%
11	3	0.0%	100.0%

Question 38

Have you taught ESL/ELD?

Grade	N	Yes	No
Overall	14	100.0%	0.0%
Kindergarten and 1	3	100.0%	0.0%
2	3	100.0%	0.0%
5	2	100.0%	0.0%
7	3	100.0%	0.0%
11	3	100.0%	0.0%

Question 39

Have you taught Vocational Education?

Grade	N	Yes	No
Overall	14	21.4%	78.6%
Kindergarten and 1	3	0.0%	100.0%
2	3	33.3%	66.7%
5	2	0.0%	100.0%
7	3	66.7%	33.3%
11	3	0.0%	100.0%

Question 40

Have you taught Alternative Education?

Grade	N	Yes	No
Overall	14	28.6%	71.4%
Kindergarten and 1	3	33.3%	66.7%
2	3	0.0%	100.0%
5	2	0.0%	100.0%
7	3	66.7%	33.3%
11	3	33.3%	66.7%

Question 41

Have you taught Adult Education?

Grade	N	Yes	No
Overall	14	64.3%	35.7%
Kindergarten and 1	3	66.7%	33.3%
2	3	33.3%	66.7%
5	2	50.0%	50.0%
7	3	66.7%	33.3%
11	3	100.0%	0.0%

Question 42

What is your primary role at this standard setting?

Grade	N	Educator	Parent	Community Member	Business Member
Overall	14	100.0%	0.0%	0.0%	0.0%
Kindergarten and 1	3	100.0%	0.0%	0.0%	0.0%
2	3	100.0%	0.0%	0.0%	0.0%
5	2	100.0%	0.0%	0.0%	0.0%
7	3	100.0%	0.0%	0.0%	0.0%
11	3	100.0%	0.0%	0.0%	0.0%

Question 43

Which grade did you work on during this standard setting?

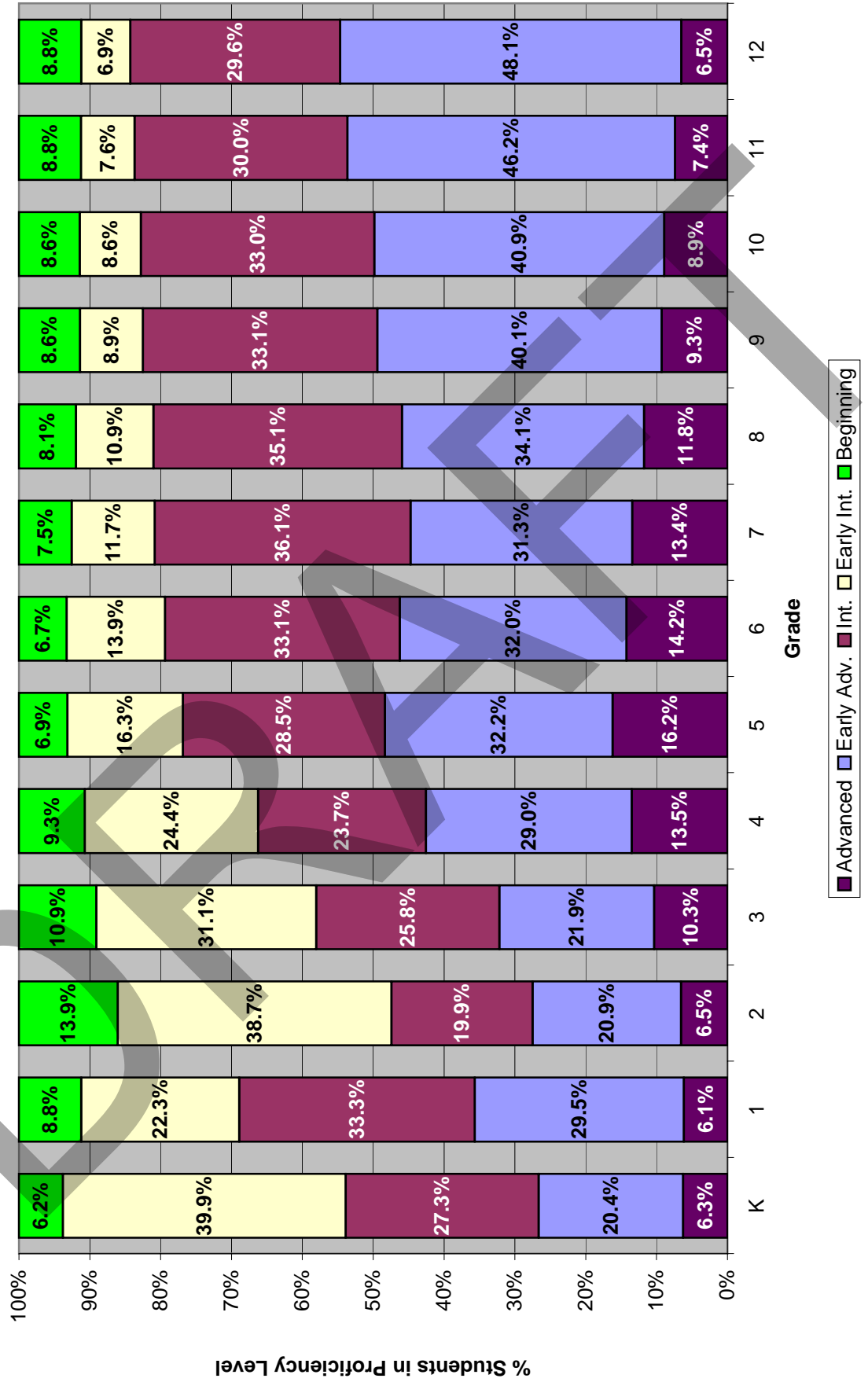
Grade	N	Overall
Kindergarten and 1	14	21.4%
2	3	21.4%
5	3	14.3%
7	2	21.4%
11	3	21.4%

SECTION J

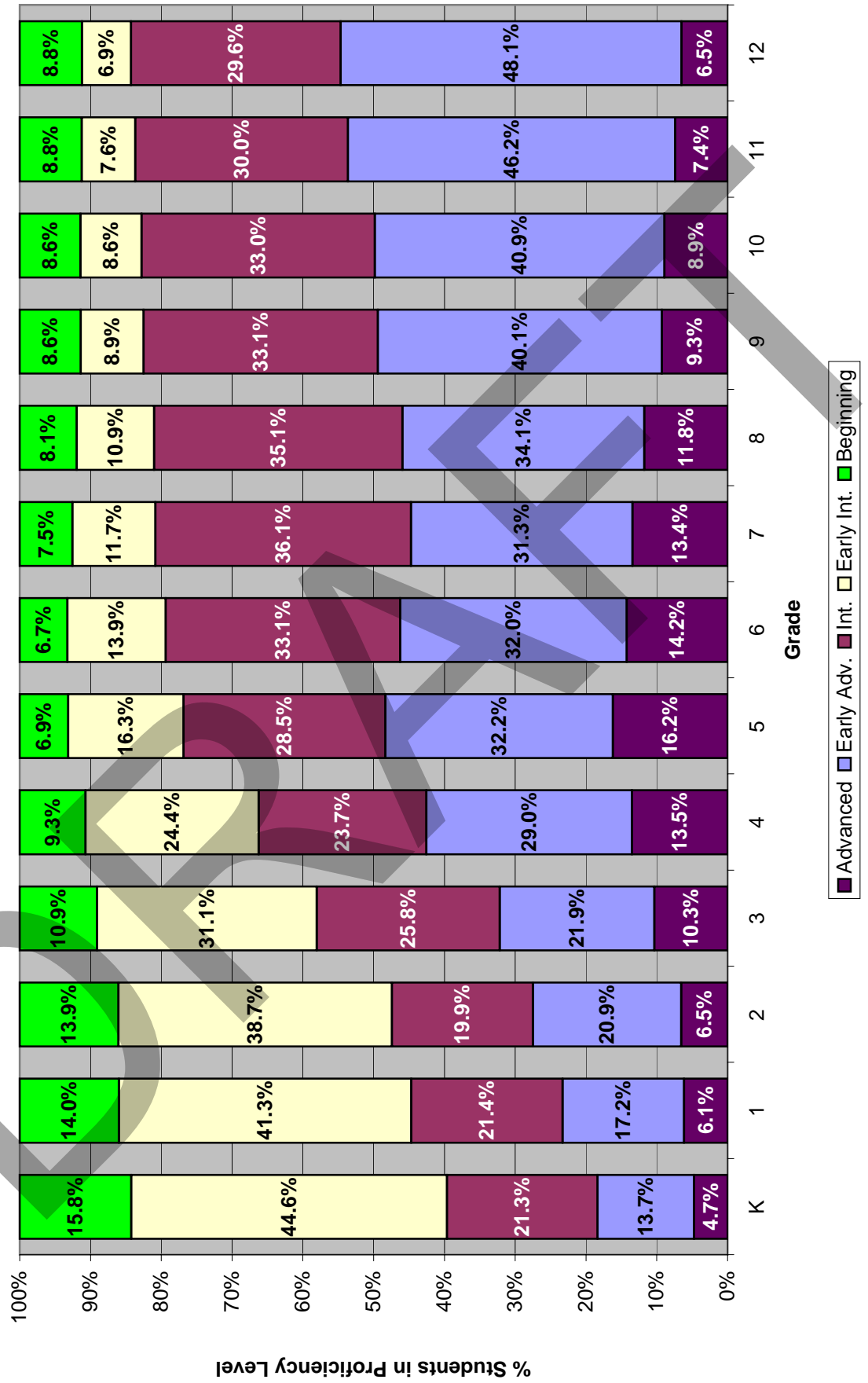
Charts of Impact Data from Articulation Discussion and Final Cut Scores

DRAFT

Oregon ELPA: Percent of Students by Proficiency Level, Articulation Discussion



Oregon ELPA: Percent of Students by Proficiency Level, Final Recommended Cut Scores



SECTION K

The Bookmark Standard Setting Procedure: Methodology and Recent Implementations

DRAFT

The Bookmark Standard Setting Procedure: Methodology and Recent Implementations

Daniel M. Lewis, Donald Ross Green, Howard C. Mitzel,

Katherine Baum, Richard J. Patz

CTB/McGraw-Hill

Paper presented at the 1998 Annual Meeting of the National Council on Measurement in Education

1. Introduction

Setting performance standards has become commonplace due to the standards-based education reform movement, Title 1 requirements, and public demands for educational accountability. However, standard setting—the determination of the cut scores for an assessment used to measure students’ progress towards performance standards—remains a controversial topic. Recent trends in standards and assessments have presented challenges for standard setting techniques. First, there is a need for a standard setting procedure that efficiently accommodates multiple cut scores. Title 1 requires the demonstration of growth through at least three performance levels—Partially Proficient, Proficient, and Advanced. Second, there is a need for a standard setting procedure that accommodates multiple item types—selected-response (SR) and constructed-response (CR). The development of new standard setting procedures has been driven in part because the widely used Angoff procedure (Angoff, 1971) does not accommodate these trends effectively and has been criticized as being seriously flawed (National Academy of Education, 1993; Mitzel, 1996).

The Bookmark Standard Setting Procedure (Lewis, Mitzel, and Green, 1996) is an item response theory-based item mapping procedure developed to address these trends in standards and assessment and to simplify the cognitive tasks required of the participants setting the cut scores. This paper presents the methodology used to conduct the Bookmark Procedure. Section 2 reviews item response theory (IRT) based standard setting procedures. Section 3 describes the Bookmark Procedure in detail. The results of recent implementations of the Bookmark Procedure are presented in Section 4. The paper closes with a discussion of these results in Section 5 and conclusions in Section 6.

2. Review of IRT-Based Item Mapping Procedures

Item mapping, sometimes referred to as “behavioral anchoring,” has been used for over a decade to help identify what students at various scale locations know and are able to do. NAEP (ETS, 1987) used scale anchoring to help interpret what students know and are able to do by mapping selected “anchor” points on the scale for the NAEP reading assessment. They selected items that discriminated well according to the criteria, “(a) eighty percent or more of the students at that [anchor] point could answer the item correctly; (b) less than 50 percent of the students at the next lower [anchor] point could answer the item correctly...” (ETS, 1987, p. 386). Item mapping, then, refers to the general approach of mapping items to locations on the IRT scale such that students with scale scores near the location of specific items can be inferred to hold the knowledge, skills, and abilities required to respond successfully to those items. NAEP continued to use scale anchoring to help interpret their results for later assessments, but the discrimination criteria applied to anchor items was modified.

The 1991 Maryland School Performance Assessment Program (MSPAP) used an item mapping procedure to set proficiency levels (CTB Macmillan/McGraw-Hill, 1992). For this purpose, score points for performance assessment items were mapped to the scale at the IRT maximum information location. The proficiency levels were set by identifying interpretable clusters of item locations on the scale and the items falling within each cluster were analyzed by content experts to interpret what students in each proficiency level knew and were able to do.

Both the NAEP anchor points and the 1991 MSPAP proficiency levels were intended to help interpret what students at various points on a scale knew and were able to do. Neither was a “true” standard setting procedure in the sense that no judgments were made concerning what students should know and be able to do; instead, both used item mapping as a means to interpret what students did know and could do at various scale locations.

NAEP conducted a bona fide standard setting for the 1992 math and reading assessments using a modified Angoff procedure (Angoff, 1971). An item mapping study was conducted as part of the review of the achievement level setting (National Academy of Education, 1993). Content experts evaluated the appropriateness of the cut scores and the quality of the achievement level descriptions. Item maps, in which items were located at the point where 80% of students in the appropriate grade could answer the items correctly (after allowing for guessing), were provided to facilitate the evaluation. Although the approach used was not intended as a new or alternative standard setting method, several positive features of the item mapping approach were noted and contrasted with the Angoff procedure that was used to set cut scores. For example, it was noted that participants using the item mapping approach had "...a more systematic understanding of the item pool as a whole than did participants using the Angoff approach (National Academy of Education, 1993, p. 110)."

One drawback of the method was also reported—the lack of clear guidelines for the probability level at which to map items to the scale. It was noted that the 80-percent-correct level possibly contributed to the experts setting very high cut scores for some of the achievement levels, and that different cut scores would possibly have resulted had a 65-percent-correct mapping criterion been used.

An "item matching" procedure was used to set proficiency levels for the 1993 MSPAP (Westat, 1994). Participants studied proficiency level descriptions and conceptualized what students at a higher level could do that students at the next lower level could not do. Initial cut scores were determined by having participants match items to the proficiency level descriptions. For example, to determine the level 2 cut score, participants examined items in order of scale location and identified the items as "clearly level 1," "clearly level 2," or "borderline." When participants identified a "run" of "clearly level 1" items followed by a "run" of "clearly level two" items, the scale locations of the items constituting the two runs were used to identify the level 2 cut score. Initial cut scores for higher levels were determined in an analogous manner, and final cut scores were determined after several rounds of discussion and consensus building.

Lewis and Mitzel (1995) developed an "IRT-Modified Angoff Procedure" for which SR items were mapped onto the IRT scale at the location at which a student would have a .5 probability of a correct response, with guessing factored out. Each positive CR item score point was mapped onto the same IRT scale at the location at which a student would have a .5 probability of obtaining at least the given score point. To determine a proficient cut score, participants conceptualized "just barely proficient" students, studied the test items in order of scale location, and classified each item according to whether a just barely proficient student should have greater than, less than, or equal to a .5 likelihood of success on the item. The cut score was determined by averaging the locations of items that participants classified at the "equal to .5" level.

Under both the Maryland 1993 standard setting procedure (Westat, 1994) and the Lewis and Mitzel (1995) procedure participants could, and did, classify items such that the participants' classifications were not consistent with the scale locations. Under the Maryland procedure, participants classified some items with higher scale locations as being associated with lower proficiency levels than other items with lower scale locations. Under the Lewis and Mitzel procedure, participants judged that Proficient students should have greater success on some items with higher scale locations than on other items with lower scale locations. This inconsistency might in part be explained by noting that the scaling of items is based on empirical student performance data, that is, what students do know and can do, and that participant judgments were based on expected student performance, that is, what students should know and be able to do. However, making judgments based on "what students should know and be able to do" without conditioning those judgments based on "what students do know and can do" can lead to serious problems in 1) interpreting the results of the assessments to which standards are applied and 2) assessing student growth relative to content standards. These problems are discussed by Lewis and Green (1997).

In 1995, the Bookmark Standard Setting Procedure was developed and used to set standards for CTB/McGraw-Hill's new standardized assessment TerraNova® and has been used to set standards in 18 states or districts from 1996 to 1998. The Bookmark Procedure evolved from Lewis and Mitzel's IRT-Modified Angoff Procedure and was designed to remove the inconsistency noted above between participants' item level judgments and the items' scale locations. This was accomplished by moving the level of judgment from the item level to the cut score level, that is, instead of making judgments about each item, participants considered all the items together to make judgments about each cut score.

Several aspects of the IRT-Modified Angoff Procedure that were particularly successful were retained in the Bookmark Procedure. Most notable are 1) the use of the ordered item booklet to help participants understand how items work together to measure student achievement relative to specified content standards and 2) the common framework for interpreting SR and CR items by mapping them to the same scale and at the same probability level. These two components were central to the primary goals of the Bookmark Procedure—to provide a standard setting procedure that treats SR and CR items in a unified manner and that is based on judgments that ease the cognitive load on participants by drawing primarily on the participants’ expertise, that is, their understanding of content standards, the curriculum, teaching practices, the assessment, and student performance. The fundamental tasks required of participants in the Bookmark Procedure are analyzing items to determine what they are measuring and specifying which items students in the various performance levels should be expected to respond to successfully. We next consider the Bookmark Procedure in detail, first providing information about basic assumptions underlying the structure of the procedure.

3. Basic Assumptions and Overview of The Bookmark Procedure

3.1 Mapping Items to the IRT Scale

Item response theory (IRT, Lord 1980) provides a framework that simultaneously characterizes the proficiency of examinees and the difficulty of test items. Each IRT-scaled item has an estimated item characteristic curve (ICC) that describes how the probability of success on the item depends on the proficiency or “scale score” of the examinee. Just as it is possible to order examinees by estimated proficiency, IRT enables items to be ordered by the proficiency needed to have a specified probability of success. The facility to order items on the IRT proficiency scale is fundamental to the Bookmark Procedure.

Selected-response (SR) items can be scaled under a variety of models, for example, the Rasch (1960) model, or the 2- and 3-parameter logistic models (Birnbaum, 1968). Constructed-response (CR) items can be scaled using polytomous models, for example, the 2-parameter or generalized partial credit model (Yen, 1993; Muraki, 1992). The 3-parameter logistic (3PL) model and the 2-parameter partial credit (2PPC) model are the default models used by CTB for SR and CR items, respectively.

Scaling SR and CR items together brings significant advantages to the standard setting process, most importantly, the ability to order the CR score points with the SR items. This joint scaling allows participants to consider all items on which the standard is to be set, regardless of item format, and to directly set a single cut score for each performance level. The joint scaling of CR and SR items can be accomplished using commercially available computer programs (e.g., PARDUX, Burket, 1996; PARSCALE, Muraki & Bock, 1991).

For the purpose of standard setting, SR and CR items are located on the IRT scale such that the location of each item type is directly interpretable and conceptually similar.

Selected-Response Items. The location of an SR item is defined as the point on the ability scale at which a student would have a .67 (2/3) probability of success, with guessing factored out. We remove consideration of guessing as a factor because participants are asked to make complex judgments about what students should know and be able to do, and the consideration of guessing unnecessarily complicates those judgments. We also note that this approach was used for the item mapping studies that followed the 1992 NAEP achievement level setting (National Academy of Education, 1993).

For the 3PL model, the probability that a student with trait or scale score θ will respond correctly to SR item j is given by

$$P_j(\theta) = c_j + (1 - c_j) / [1 + \exp(-1.7a_j(\theta - b_j))].$$

where a_j is the item discrimination, b_j is the item difficulty, and c_j is the probability of a correct response by a very low-scoring student. We estimate the probability, P_j^* , of a correct response with guessing removed using the formula

$$P_j^*(\theta) = (P_j(\theta) - c_j) / (1 - c_j).$$

The location of SR item j is θ , such that $P_j^*(\theta) = .67$.

Constructed-Response Items. Each CR score point has a unique location on the scale. The location of a given CR score point is defined as the position on the ability scale for which students have a .67 probability of achieving at least that score point, that is, that score point or higher. This criteria was selected so that the location of the CR score point could be interpreted in a manner similar to the location of a SR item and in a way that is conceptually useful to the participants in setting the cut score.

Using the 2PPC model for CR items, the probability that a student with trait or scale score θ will respond at score level k to CR item j is given by

$$P_{jk}(\theta) = \exp(z_{jk}) / \sum_{i=1}^{m_j} \exp(z_{ji}),$$

where $z_{jk} = (k-1)\alpha_j - \sum_{i=0}^{k-1} \gamma_{ji}$, α_j and γ_{ji} , $i = 1, 2, \dots, m_{j-1}$, are the parameters estimated during calibration,

$\gamma_{j0} = 0$ for all j , and m_j is the number of levels for item j .

For the purpose of standard setting, the location of score point k for constructed response item j , is the scale score θ , such that $P_{jk}^*(\theta) = .67$, where

$$P_{jk}^*(\theta) = \sum_{i=k}^{m_j} P_{ji}(\theta).$$

Although the selection of .67 as the probability level used to map items to the scale is somewhat arbitrary, this value was not selected capriciously. First, because the probability level must be considered by the participants when making their judgments, a familiar value was desired. That is, using a probability level of .5823 would not be useful, but values such as .5 (1/2), .67 (2/3), or .75 (3/4) would be. Second, other item mapping procedures and research have provided some precedent. Huynh (1998) showed that for the 3PL model, the item information function is maximized at θ for which $P(\theta) = (c + 2)/3$. This corresponds to the value of 2/3 when guessing is factored out. Thus, the choice of 2/3 for mapping SR items corresponds to the maximum information location. Huynh states that the maximum information location associated with a correct response "...might serve as a signal that an examinee located at this place would be 'expected' to have the skills underlying the item."

3.2 Bookmark Standard Setting Materials

Many of the materials used for Bookmark Standard Settings are commonly used within other standard setting procedures, such as operational test booklets, student exemplar papers, and scoring guides. The following materials are unique to Bookmark Standard Settings and other item mapping procedures.

Ordered Item Booklets. Ordered item booklets are typically assembled using all items on which the standards are to be based, in order of scale location. The ordered item booklet focuses the participants' attention on one item per page, with the "easiest" item (lowest scale location) first and the "hardest" item (highest scale location) last. The purpose of the ordered item booklets is to help participants' foster an integrated conceptualization of what the test measures, as well as to serve as a vehicle to make cut score judgments. Studying the items one by one, from easiest to hardest, discussing what each item measures and why each item is more difficult than items that precede it in the book, is intended to provide participants with an understanding of how the trait increases in complexity as the items ascend the scale, and of the knowledge, skills, and abilities students must hold in order to respond successfully to items.

The items used in the ordered item booklets can be items from single or multiple forms of an operational test or items on a common scale from an item pool that is representative in content and difficulty of a single form of the operational test. The use of items beyond those of a single operational form is recommended when possible, to increase the generalizability of the standards to other forms to which the standards may be applied in future years.

Item Map Rating Forms. The item map rating form is a guide to the ordered item booklet, and lists all items ascending by location, that is, in the same order in which they appear in the ordered item booklets. Associated item information is also included on the item map rating form, such as the items' scale location, item number in the operational or field test booklet, the standard or objective the item was written to measure, space for the participants to make notes about the items, and the cut score judgments they are considering for each round.

3.3 Determining Cut Scores Under the Bookmark Procedure

The cut score for a given performance level, for example, Proficient, can be identified by a bookmark placed between two items in the ordered item booklet such that from the judge's perspective, the items preceding the bookmark represent content that all proficient students should be expected to know and be able to do (with at least a 2/3 likelihood of knowing the correct response for SR items or of obtaining at least the given score point for CR item score points). By placing the bookmark at the furthest most item for which this is true, a location on the ability scale can be estimated as the cut score. This is computed as the scale location of the item that appears immediately prior to the bookmark. Judgments are made at the cut score level, that is, participants consider all the items when they place their bookmarks, but the bookmarks define cut scores.

To set two cut scores defining three performance levels, for example, Partially Proficient, Proficient, and Advanced, each judge considers the items in the ordered booklet and places two bookmarks that define the two cut scores. The items that precede the first bookmark should represent content that all proficient students are expected to know and be able to do. The items that precede the second bookmark should represent content that all advanced students are expected to know and be able to do.

When an item precedes a judge's bookmark, the judge is stating that all proficient students should have ability sufficient to have at least a 2/3 likelihood of responding correctly to the SR item or of obtaining at least that score point for a CR item score point. This probability level is held only by students with scale ability locations as high or higher than the scale location of the item. Thus, all proficient students must have ability level at least as high as the scale location of each item before the bookmark. On the other hand, when an item falls after the bookmark, the judge is stating that a student could be classified as proficient, yet have less than a 2/3 likelihood of success on the item. This means that a student could have ability lower than the location of the first item after the bookmark and still be classified as proficient. Thus, the proficient cut score is at least the location of the item immediately prior to the bookmark but less than the location of the item following the bookmark. The location of the item immediately prior to the bookmark is used as the operational cut score.

3.4 Writing Performance Level Descriptors

Performance level descriptors are intended to be valid descriptions of the knowledge, skills, and abilities held by students that place in the various performance levels. Performance level descriptors emerge as an outcome of setting cut scores under the Bookmark Procedure. For example, suppose two cut scores are set defining the three performance levels Partially Proficient, Proficient, and Advanced. Items prior to the Proficient bookmark reflect content that all Proficient students are expected to know and be able to do, and therefore, the knowledge, skills, and abilities required to respond successfully to these items are synthesized to form descriptors of the Proficient student. Similarly, the items following the Proficient bookmark and prior to the Advanced bookmark are used to yield descriptors of the additional knowledge, skills, and abilities a student must hold to be considered Advanced.

The estimated probability of a successful response for a student in a given performance level is at least .2/3 for the items used to write the performance level descriptors. Thus, descriptors written with this approach are valid to the degree that participants can communicate the knowledge, skills, and abilities required to successfully complete the items attributed to the respective performance levels. Of course, because they are based on probabilities, not every student will have mastered all the skills attributed to them by the descriptors. The validity of performance level descriptors written in this manner is discussed more fully by Lewis and Green (1997).

3.5 Bookmark Standard Setting Panel Composition and the Use of Multiple Panels

Operationally, the composition of a standard setting panel results from the sponsoring agency's selection criteria and availability of participants. We recommend at least 18 participants per panel. The panel of participants for a given grade and content area are typically divided into three small groups. One participant within each small group is predesignated to act as a small group facilitator for the process, and receives training prior to the standard setting. Small-group facilitators are selected from the pool of participants based on experience with the students, curriculum, instruction, assessment, and the ability to facilitate groups. The small-group facilitators are voting members of their small group. The sponsoring agency makes recommendations for the assignment of participants to small groups such that the three small groups are roughly balanced in terms of the educational background and geographic location of the participants. The use of small groups facilitates having all participants actively involved in the discussion of items and expectations for student performance. A Bookmark standard setting is typically facilitated by a single large group leader who is responsible for monitoring the process for a given grade and content area and the small group facilitators who monitor the process within their small groups.

The use of multiple small groups is integrated into the structure of the judgment process. Prior to the first round of judgments, participants study the ordered item booklets within their small groups, and discuss what each item measures and why each item is more difficult than the preceding items in the booklet. Following discussion, participants make individual and independent Round 1 judgments, that is they place bookmarks that indicate the items that reflect content they expect students in each performance level to know and be able to do.

In Round 2, each small group discusses the items for which there was not consensus according to the small group's Round 1 judgments. For a given performance level, these are the items in the ordered item booklet between the first and last of the small group participants' bookmarks. This appropriately narrows the discussion only to the items for which participants have differing opinions relative to expected student performance for a given performance level. Following discussion, Round 1 judgments may be modified with Round 2 judgments.

Prior to Round 3, a small-group judgment is computed for each small group as the median of the small group's bookmark placements. In Round 3, the large group is presented with each small group's Round 2 judgments and the estimated percent of students in each performance level based on the current large group median. The large group discusses the reasonableness of the impact data and the items for which there was not consensus among the small groups. Following discussion, Round 2 judgments may be modified with Round 3 judgments.

The Bookmark Procedure is structured so that each small group works independently of the other small groups until the third round. The standard error estimated from each small groups' independent Round 2 results provides a measure of the stability of the cut scores, as discussed in the next section.

3.6 Capturing and Communicating Degrees of Consensus

The Bookmark Standard Setting Procedure is a collaborative enterprise that fosters consensus among participants as to the standards to which we hold our students accountable. However, consensus is not forced. In the results discussed in Section 4, varying degrees of consensus were attained. It is important that the degree of consensus be measured and reported with the recommended cut scores to the governing bodies who make final cut score decisions.

The degree of consensus is quantified by calculating a standard error for each cut score arrived at through the multiple-group, three-round process. Because the small groups act independently through the first two rounds, an appropriate standard error can be calculated by treating individual Round 2 scores as if sampled from independent clusters. Formulas for the cluster sample standard error (Cochran, 1963, p. 210) are presented in Appendix 1.

Data arising in standard setting contexts have complex dependency structures and reflect many sources of error. It is important to appreciate this complexity and avoid making strong conclusions based on statistical procedures whose assumptions can not be satisfied. In Bookmark standard settings we use appropriately general statistics such as the cluster sample standard error, as well as graphics to help inform these judgments.

4. Recent Implementations of the Bookmark Procedure

4.1 Background

Table 1 summarizes the grades, content areas, test scales, test formats, and numbers of participants associated with four state and one district Bookmark standard settings facilitated by CTB in 1996 and 1997. A total of twenty panels set cut scores in grades ranging from 3 to 10 in Reading, Language Arts, and Mathematics.

For thirteen of the twenty grade/content areas, the ordered item booklets used to set cut scores included more items than were on the operational test forms. As Table 1 indicates, the operational test forms had an average of 67 score points and the ordered item booklets used to set cut scores had an average of 111 score points. The operational tests were all composed of a mixture of SR and CR items with an average of 76 percent SR items and 24 percent CR items. On average 59 percent of the total score points were from SR items and 41 percent were from CR items. The ordered item booklets used to set standards had an average of 73 percent SR items and 27 percent CR items. On average, 54 percent of the total score points in the ordered item booklets were from SR items and 46 percent were from CR items.

Table 1 also shows the number of cut scores, number of small groups, and total number of judges per grade/content area.

4.2 An Illustrative Example

Figures 1-4 illustrate the Bookmark Standard Setting Procedure for an example selected from the recent implementations. In this case, three cut scores were set for a Grade 8 Language Arts assessment. Figures 1, 2, and 3 show the individual participants' Proficient cut score ratings for Small Groups 1, 2, and 3, respectively. The vertical axes indicate the test scale referenced to a mean of 0 and standard deviation of 1. The horizontal axes indicate the round (1, 2, or 3).

Figure 1 shows the Proficient cut score ratings for the four participants in Small Group 1. Note that there is a reasonable amount of variability in the first round, with Group 1 participants' cut scores ranging from .05 to .44 on the scale. The observed variability reflects the fact that in the first round, participants make individual and independent judgments.

In the second round, the small group participants discuss and debate the rationale and perspective that lead to each of their Round 1 judgments. This tends to decrease the variability within each small group. In the case of Group 1 (Figure 1), a high degree of consensus has been reached in Round 2, with participants' cut scores ranging from .41 to .44 on the scale. Three of the four Group 1 participants raised their cut scores, apparently strongly influenced by the fourth participant's perspective.

In the third round, small-group cut scores are computed for each small group (based on small-group medians). Each small group presents the rationale and perspective that lead to their Round 2 judgments, and impact data is presented. In the example indicated in Figure 1, all participants in Group 1 maintained their Round 2 judgments in Round 3. This was probably due to the fact that Small Groups 2 and 3 both made Round 2 judgments that were very similar to those of Small Group 1, as can be observed in Figures 2 and 3.

Figures 2 and 3 illustrate the three rounds of judgments for Small Groups 2 and 3, respectively. Figure 2 indicates that Group 2 made judgments for each round that were very similar to those of Group 1. Figure 3 shows a different pattern of ratings for Small Group 3. There is a reasonable amount of variability in the Round 1 ratings for Small Group 3, with the five participants' cut scores ranging from .31 to .61. In the second round, we see the results of consensus building, however in this case, the participants tended toward the group's median cut score. The range of the participants' cut scores (.41 to .46) has decreased considerably from that of Round 1. In the third round, Small Group 3 reached consensus, with all five participants rating the Proficient cut score at .44.

Figure 4 illustrates the judgments for all participants, by round, for all three cut scores (Partially Proficient, Proficient, and Advanced). The middle set of lines indicate the Proficient judgments examined in Figures 1-3. It can easily be seen that in Round 2, each of the three groups independently arrived at the same median cut score (.44). However, this does not occur routinely. The reader need only look at the patterns for the Advanced and Partially Proficient cut scores to observe that although Round 2 does typically bring a degree of consensus, it is not as uniform for these cut scores as for the Proficient cut score.

Also depicted in Figure 4 are confidence bands centered at the Round 3 median cut score with a width of two Round 2 standard errors. The Round 3 median best captures the consensus cut score from the entire Bookmark Procedure. Round 2 standard errors are used to quantify the degree of consensus obtained across independent groups, as discussed in Section 3.6 Capturing and Communicating Degrees of Consensus. The type of information exemplified in Figure 4, is valuable to decision makers who must act on the recommendations of the standard setting panels. In the example depicted in Figure 4, the participants' recommended cut scores were adopted by the sponsoring agency.

4.3 Results

The results for the proficient cut score by round for each of the 20 examples are located in Table 2 (Summary data for all performance level cut scores are provided in Tables 3 and 4.). All statistics that are derived from the participants cut score judgments are presented in standardized units, that is, referenced to the standard deviation units of the scale. This allows statistics across scales to be compared.

The column labeled "Range (Cut)" indicates the magnitude of the range of the participants' scale score cut scores for each round and each cut score in scale standard deviation units (computed as the difference between the maximum and minimum of the participants' cut scores divided by the scale standard deviation). The column "SD (Cut)" indicates the standard deviation of the participants' scale score cut scores for each round in scale standard deviation units.

The columns labeled "Intra Class Corr" [Intraclass Correlations] and "Round 2 SE (Cut)" [standard errors] provide information about the replicability of the participants' judgments across groups. These are explained in detail in Appendix 1. The standard error is reported in scale standard deviation units.

Table 3 presents the mean SD of the participants' cut score judgments for each cut score and round (in standardized units), as well as the standard deviation, minimum, and maximum of these standard deviations. For the Advanced cut scores, the mean SDs decreased from .35 (Round 1) to .16 (Round 2) to .15 (Round 3). For the Proficient cut scores, the mean standard deviations decreased from .32 (Round 1) to .14 (Rounds 2 and 3). For the Partially Proficient cut scores, the mean standard deviations decreased from .27 (Round 1) to .16 (Round 2) to .13 (Round 3).

Table 3 also presents the mean Round 2 standard errors and intraclass correlations of the participants' cut score judgments for each cut score. The mean Round 2 standard errors are .07, .08, and .07, and the mean Round 2 intraclass correlations are .67, .69, and .70 for the Advanced, Proficient, and Partially Proficient cut scores, respectively.

Table 4 presents the mean difference in median cut scores between successive rounds, as well as the standard deviation, minimum, and maximum of the mean differences. The mean differences between the median Round 2 and Round 1 cut scores were .22, .16, and .10, for the Advanced, Proficient, and Partially proficient cut scores, respectively. The mean differences between the median Round 3 and Round 2 cut scores were .04, .00, and .04, for the Advanced, Proficient, and Partially Proficient cut scores, respectively.

5. Discussion

As would be expected in a consensus building process, the variability of participants' judgments tended to decrease in successive rounds for each cut score. The magnitude of the variability was similar for the three performance levels in each round. This is indicated by the mean standard deviations (Table 3) for the Advanced, Proficient, and Partially Proficient cut scores of .35, .32, and .27, respectively, in Round 1; .16, .14, and .16, respectively in Round 2; and .15, .14, and .13, respectively, in Round 3. This suggests a consistency in the degree to which participants are able to translate their qualitative conceptualizations of each performance level operationally into expected performance on test items. The ability for participants to be able to clearly conceptualize the knowledge, skills, and abilities of students within each performance level is fundamental to any standard setting process. These results indicate that participants seem to be able to do so to a similar degree for three performance levels. This may not hold when there are more than three performance levels.

A pattern of decreasing variability in participants' judgments from each round to the next is also consistent for the three performance levels. The mean standard deviations decreased from .35 (Round 1) to .16 (Round 2) to .15 (Round 3) for the Advanced performance level; from .32 to .14 to .14 for the Proficient performance level; and from .27 to .16 to .13 for the Partially Proficient performance level. A considerable reduction in variability occurs from

Round 1 to Round 2, but there is only a nominal reduction from Round 2 to Round 3. This indicates that the participants' perspectives change considerably from the interactions within their small groups during Round 2, but do not change as much from the interactions between the small groups or the consideration of impact data in Round 3. This is desirable from the perspective that participants should feel more confident of their judgments with each round, and therefore, should be less likely to modify their judgments in subsequent rounds. However, the results may not only reflect an increase in confidence in participants' judgments, but also the support of other members within the small group to maintain their judgments in spite of differences between the small groups.

The mean standard errors computed from Round 2 provide an estimate of the variability of the cut scores across panels. The mean standard errors of .07, .08, and .07 for the Advanced, Proficient, and Partially Proficient cut scores are of similar magnitude to those reported for Math and Reading in the NAEP 1992 standard setting (ACT, 1993). It is important to remember that these are estimated from the small groups' independent Round 2 results.

The mean Round 2 intraclass correlations of .67, .69, and .70 for the Advanced, Proficient, and Partially Proficient cut scores, respectively, indicate that an appropriate degree of within-group consensus occurred in Round 2, and that individual judgments should not be treated as independent once group discussions have taken place.

Several conclusions can be drawn from looking at the mean differences between the median of the participants' cut scores between Rounds 2 and 1 and between Rounds 3 and 2. The mean differences in medians between Rounds 2 and 1 of .22, .16, and .10, for the Advanced, Proficient, and Partially Proficient cut scores, respectively, indicate that participants' cut scores tend to rise considerably from Round 1 to Round 2. This is somewhat surprising, as one might expect participants' judgments to tend toward the median, but leave the median relatively unchanged. The rise may be attributable to social pressure for high standards. For example, suppose one participant enters Round 2 having placed his/her bookmark in the ordered item booklet at say, page 50, and a second participant has placed his/her bookmark on page 60. In Round 2, the participants discuss items 50-59 in terms of whether a student should be expected to master these items to be considered proficient. It may be that under these circumstances, a psychological advantage exists for "higher standards." It is interesting to note that the increase in median cut scores from Round 1 to Round 2 is greatest for the Advanced cut score, and the least for the Partially Proficient cut score. Thus, the increase is positively correlated with the performance level, suggesting that this social pressure is greatest when the standards are expected to be highest.

The mean differences between the median of the participants' cut scores between Round 3 and Round 2 are .04, .00, and .04, for the Advanced, Proficient, and Partially Proficient cut scores, respectively. Thus, the increase in median cut scores from Round 2 to Round 3 tends not to be large. This must be considered in light of the two new pieces of information that are provided to participants in the third round. First, the participants view and discuss the results from the other small groups. Second, the participants discuss impact data associated with the median cut score computed from all participants' bookmarks. The results indicate that although these factors can affect participants' judgments, they are not systematic. Again, it seems that by Round 3, participants are well grounded in their judgments.

6. Conclusions

In sum, the results indicate that the participants are making judgments as would be expected and desired, given the structure of the Bookmark Procedure. The patterns of variability are particularly encouraging. The highest variability occurs in the first round, when participants make independent ratings, and decreases significantly from Round 1 to Round 2, but does not decrease significantly from Round 2 to Round 3. This indicates that participants listen to each others' perspectives and in many cases find the arguments persuasive and therefore modify their judgments in Round 2. The stability of the small group median scores from Round 2 to Round 3 suggest that participants have developed a stable perspective by the third round. They do not react strongly to the new information provided in the third and final round as they did to that of the second round.

Setting standards is a complex process involving educational, psychological, statistical, and ultimately, political considerations. We have observed that the Bookmark Procedure facilitates the standard setting process by providing a framework through which informed educators come to understand how a particular test measures the skills the students are expected to master, and by providing a structure that fosters rational consensus building regarding expected student performance. Participants' judgments are based on well defined criteria—which items students be expected to respond successfully to be classified in the various performance levels.

Further studies are required to determine the degree to which cut scores arrived at through the Bookmark Procedure are consistent with other measures of student proficiency such as teacher judgment or cut scores set concurrently with other procedures. There is no “gold standard” for cut scores or standard setting procedures. Research has shown that different standard setting procedures will likely lead to somewhat different cut scores (National Academy of Education, 1993). However, several aspects of the Bookmark Procedure have lead CTB to make it their default standard setting method.

First, participants leave the Bookmark Standard Setting with a strong understanding of what their final cut scores mean in terms of expected student performance for each performance level, as measured by the assessment. This understanding is fostered by the use of the ordered item booklets and the structure provided by item mapping procedures in general. Observations during the item mapping studies that followed the 1992 NAEP standard setting have also been observed following each Bookmark standard setting:

“...the experts or judges using the item-mapping approach had a much more direct understanding of the continuum for which they were attempting to devise levels...by engaging in discussions and studying the item maps, participants had a more systematic understanding of the item pool as a whole than did participants using the Angoff approach.... (National Academy of Education, 1993, p. 110).”

Second, Bookmark Standard Setting participants are able to translate this “understanding” to communicate what students in each performance level know and are able to do by writing performance level descriptors based on empirical data. Teachers, parents, and students are able to use the performance level descriptors to understand the level of achievement required for students to place in each performance level. The sponsoring agency and the public can use the performance level descriptors and the percent of students in each performance level to better understand the current state of student achievement relative to the standards.

Third, Bookmark Standard Setting participants frequently comment on how instruction would improve if every teacher could go through a similar process. Their comments suggest that they have a unique awareness of how the assessment relates to the content standards, curriculum, and instruction. CTB is currently experimenting with methods of capturing the participants’ perspectives to provide information to the sponsoring agency that may improve the alignment of content standards, curriculum, instruction, and assessment. This topic is more fully discussed in Lewis and Green (1998).

TerraNova is a registered trademark of The McGraw-Hill Companies, Inc.

Send requests for information to: Daniel M. Lewis
Research Department
CTB/McGraw-Hill
Monterey, CA 93940

References

- ACT (1993). Setting achievement levels on the 1992 National Assessment of Educational Progress in Mathematics, Reading, and Writing: A technical report on reliability and validity.
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 508-600). Washington, DC: American Council on Education.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395-479). Reading, MA: Addison-Wesley.
- ETS. (1987). The NAEP 1983-84 Technical Report. Princeton, NJ: Educational Testing Service.
- Burket, G. R. (1996). PARDUX [Computer program]. Monterey, CA: CTB/McGraw-Hill.
- Cochran, W. G. (1963). *Sampling techniques*. New York: John Wiley & Sons.
- CTB Macmillan/McGraw-Hill. (1992). Final technical report: Maryland School Performance Assessment Program, 1991. (Available from the Maryland State Department of Education, Baltimore, MD)
- Huynh, H. (1998). On score locations of binary and partial credit items and their applications to item mapping and criterion-referenced interpretation. *Journal of Educational and Behavioral Statistics*, 23, 37-58.
- Lewis, D. M., & Mitzel, H. C. (September 1995). An item response theory based standard setting procedure. In D. R. Green (Chair), Some uses of item response theory in standard setting. Symposium presented at the annual meeting of the California Educational Research Association, Lake Tahoe, NV.
- Lewis, D. M., Mitzel, H. C., & Green, D. R. (June 1996). Standard setting: A Bookmark approach. In D. R. Green (Chair), IRT-based standard setting procedures utilizing behavioral anchoring. Symposium presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.
- Lewis, D. M., & Green, D. R. (June 1997). The validity of performance level descriptors. Paper presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.
- Lewis, D. M., & Green, D. R. (June 1998). Assessing the state of the standards: Linking content standard, curriculum & instruction, and assessment. Paper presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Colorado Spring, CO.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New York: Erlbaum.
- Mitzel, H. C. (1996). Standard setting as a judgment task. In D. R. Green (Chair), IRT-based standard setting procedures utilizing behavioral anchoring. Symposium presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.
- Muraki, E., & Bock, R. D. (1991). PARSCALE: Parameter scaling of rating data [Computer program]. Chicago: Scientific Software, Inc.
- National Academy of Education. (1993). Setting performance standards for student achievement. Stanford: Author.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.
- Westat. (1994). Establishing proficiency levels and descriptions for the 1993 Maryland School Performance Assessment Program (MSPAP). Technical Report. Rockville, MD.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local independence. *Journal of Educational Measurement*, 30, 187-213.

Appendix 1

Calculating a Meaningful Standard Error for the Bookmark Cut Score

In the Bookmark Standard Setting Procedure for a given grade and content area, participants are assigned to roughly equivalent small groups that work independently through Round 2. Thus, the set of Round 2 cut scores provide some information about the stability of consensus in Bookmark cut scores across independent small group replications. To quantify this degree of consensus, we calculate the cluster sample standard error (Cochran, 1963, p. 210) of the Round 2 mean cut score. Cluster sample standard errors are appropriate when, as may be reasonably assumed here, data are collected from groups and independence can be assumed between groups but not within groups.

For the Bookmark Procedure, the standard error of the Bookmark cut score (SE_{cut}) is given by the cluster sample standard error of the Round 2 mean cut score:

$$SE_{cut} = \sqrt{\frac{S^2}{N}[1 + (n-1)r]},$$

where S^2 is the sample variance of individual Round 2 cut scores, r is the Round 2 intraclass correlation, N is the number of participants, and n is the number of groups. To be precise, if Y_{ik} is the cut score from the i^{th} participant in the k^{th} group, \bar{Y}_k is the average cut score for group k , and $\bar{\bar{Y}}$ is the average of all Round 2 cut scores, then

$$r = \frac{Var(\bar{Y}_k)}{Var(\bar{Y}_k) + Var(Y_{ik} - \bar{Y}_k)} \quad \text{and} \quad S^2 = \frac{1}{N-1} \sum_{n,k} (Y_{nk} - \bar{\bar{Y}})^2$$

If we have only two groups ($n=2$) and perfect dependence (agreement) within groups ($r=1$), then the cluster sample standard error simplifies to $SE_{cut} = |\bar{Y}_1 - \bar{Y}_2|/2$, which is the standard error formula employed by NAEP for two independent replications of a modified Angoff procedure (ACT, 1983, pp. 4-8). If, on the other hand, individual participants acted independently of their groups ($r=0$), then the cluster sample standard error simplifies to the traditional standard error of the mean for independent observations, $SE_{cut} = \sqrt{S^2/N}$. In this manner, SE_{cut} provides a simple, flexible, and general way to quantify the amount of uncertainty associated with final Bookmark cut scores.

It is appropriate (if statistically imprecise) to say that repeated replications of this very standard setting procedure with different judges sampled from the same population of potential judges would result in a range of cut scores, most of which would fall in a band of width $4 * SE_{cut}$. In Figures 1-4 we depict such an interval centered at the median of the Round 3 cut score. The purpose of calculating statistics like SE_{cut} and producing graphs of the types displayed here is to effectively communicate the complex information that is gathered during a Bookmark Standard Setting Procedure.

Table 1. Background Information for Recent Implementations

Grade	Content Area	Operational Test			Ordered Item Booklet				Total # of Judges
		# of SR Items	# of CR Items	Total Score Points	# of SR Items	# of CR Items	Total Score Points	# of Small Groups	
3	Reading	30	13	51	74	18	104	2	4
	Language	20	4	31	56	10	83	1	6
	Math	30	8	48	96	16	132	2	4-5
6	Reading	31	9	55	98	23	159	2	4-5
	Language	24	5	43	67	12	114	1	6
	Math	31	11	58	95	33	176	1	8
8	Reading	34	6	51	67	12	98	2	6
	Language	21	5	41	43	12	88	3	4-5
	Math	31	10	54	63	20	116	2	5-6
4	Reading	46	21	95	46	21	95	3	6-7
4	Writing	32	13	59	42	34	112	3	6-7
4	Math	32	23	77	31	62	146	4	6-8
8	Math	31	20	73	31	56	145	4	5-6
10	Math	25	19	73	25	43	133	4	4-5
3	ELA	46	2	55	46	2	55	1	6
6	ELA	55	13	87	55	13	87	1	6-7
8	ELA	72	18	113	72	18	113	1	5-6
10	ELA	69	11	100	69	11	100	1	6
10	ELA	67	13	98	67	13	98	1	6-7
10	Math	59	22	74	59	22	74	1	6

Note. ELA = English/Language Arts, SR = selected response, CR = constructed response.

Table 2. Results

Grade	Content Area	Cut	Round	Range (Cut)*	SD (Cut)*	Intra Class Corr	Round 2 SE (Cut)*
3	Reading	Proficient	1	0.45	0.15	0.96	0.17
			2	0.53	0.25		
			3	0.31	0.11		
3	Language	Proficient	1	0.29	0.11	NA	NA
			2	0.19	0.07		
			3	0.00	0.00		
3	Math	Proficient	1	1.09	0.37	0.37	0.04
			2	0.24	0.08		
			3	0.00	0.00		
6	Reading	Proficient	1	0.72	0.26	0.50	0.01
			2	0.05	0.02		
			3	0.00	0.00		
6	Language	Proficient	1	0.41	0.16	NA	NA
			2	0.27	0.11		
			3	0.27	0.11		
6	Math	Proficient	1	1.32	0.36	NA	NA
			2	0.67	0.19		
			3	0.00	0.00		
8	Reading	Proficient	1	0.55	0.13	0.70	0.02
			2	0.11	0.03		
			3	0.00	0.00		
8	Language	Proficient	1	0.56	0.18	0.09	0.00
			2	0.05	0.01		
			3	0.05	0.01		
8	Math	Proficient	1	0.89	0.23	0.81	0.10
			2	0.38	0.15		
			3	0.28	0.13		
4	Reading	Proficient	1	0.97	0.25	0.72	0.06
			2	0.32	0.13		
			3	2.07	0.56		
4	Writing	Proficient	1	1.52	0.69	0.16	0.04
			2	0.51	0.12		
			3	2.13	0.55		
4	Math	Proficient	1	2.52	0.52	0.63	0.08
			2	1.07	0.25		
			3	1.05	0.20		
8	Math	Proficient	1	2.37	0.44	0.65	0.08
			2	1.32	0.24		
			3	1.32	0.24		
10	Math	Proficient	1	1.33	0.28	0.73	0.02
			2	0.29	0.08		
			3	0.42	0.10		
3	ELA**	Proficient	1	0.89	0.25	1.00	0.03
			2	0.12	0.06		
			3	0.10	0.02		
6	ELA	Proficient	1	1.53	0.29	1.00	0.05
			2	0.18	0.08		
			3	0.17	0.07		
8	ELA	Proficient	1	2.66	0.56	0.94	0.14
			2	0.59	0.23		
			3	0.09	0.02		
10	ELA	Proficient	1	1.45	0.43	0.98	0.25
			2	1.13	0.43		
			3	1.05	0.34		
10	ELA	Proficient	1	1.74	0.41	0.60	0.08
			2	1.06	0.19		
			3	1.04	0.18		
10	Math	Proficient	1	1.54	0.34	0.41	0.06
			2	0.60	0.17		
			3	0.58	0.17		

* Values are in scale standard deviation units.

** ELA = English/Language Arts.

Table 3. Summary Statistics: Measure of Variability in Participants' Cut Score Judgments

	Standardized Standard Deviation				Standardized Standard Error				Intra Class Correlation			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Advanced												
Round 1	0.35	0.16	0.17	0.73								
Round 2	0.16	0.12	0.02	0.46	0.07	0.05	0.02	0.15	0.67	0.20	0.37	0.99
Round 3	0.15	0.15	0.00	0.51								
Proficient												
Round 1	0.32	0.16	0.11	0.69								
Round 2	0.14	0.10	0.01	0.43	0.08	0.07	0.00	0.25	0.69	0.27	0.09	1.00
Round 3	0.14	0.17	0.00	0.56								
Partially Proficient												
Round 1	0.27	0.20	0.05	0.68								
Round 2	0.16	0.14	0.03	0.53	0.07	0.04	0.03	0.13	0.70	0.30	0.11	1.00
Round 3	0.13	0.10	0.00	0.28								

Table 4. Summary Statistics: Difference Between Successive Round Medians

	Round 2 - Round 1				Round 3 - Round 2			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Advanced	0.22	0.26	-0.16	0.78	0.04	0.15	-0.11	0.52
Proficient	0.16	0.23	-0.13	0.81	0.00	0.22	-0.73	0.24
Partially Proficient	0.10	0.20	-0.11	0.66	0.04	0.16	-0.14	0.55

Note. Standardized scale score units are used.

Figure 1. Group 1 Proficient Cutscores for Rounds 1, 2, and 3
(Vertical axis in scale standard deviation units: Mean = 0, SD = 1.)

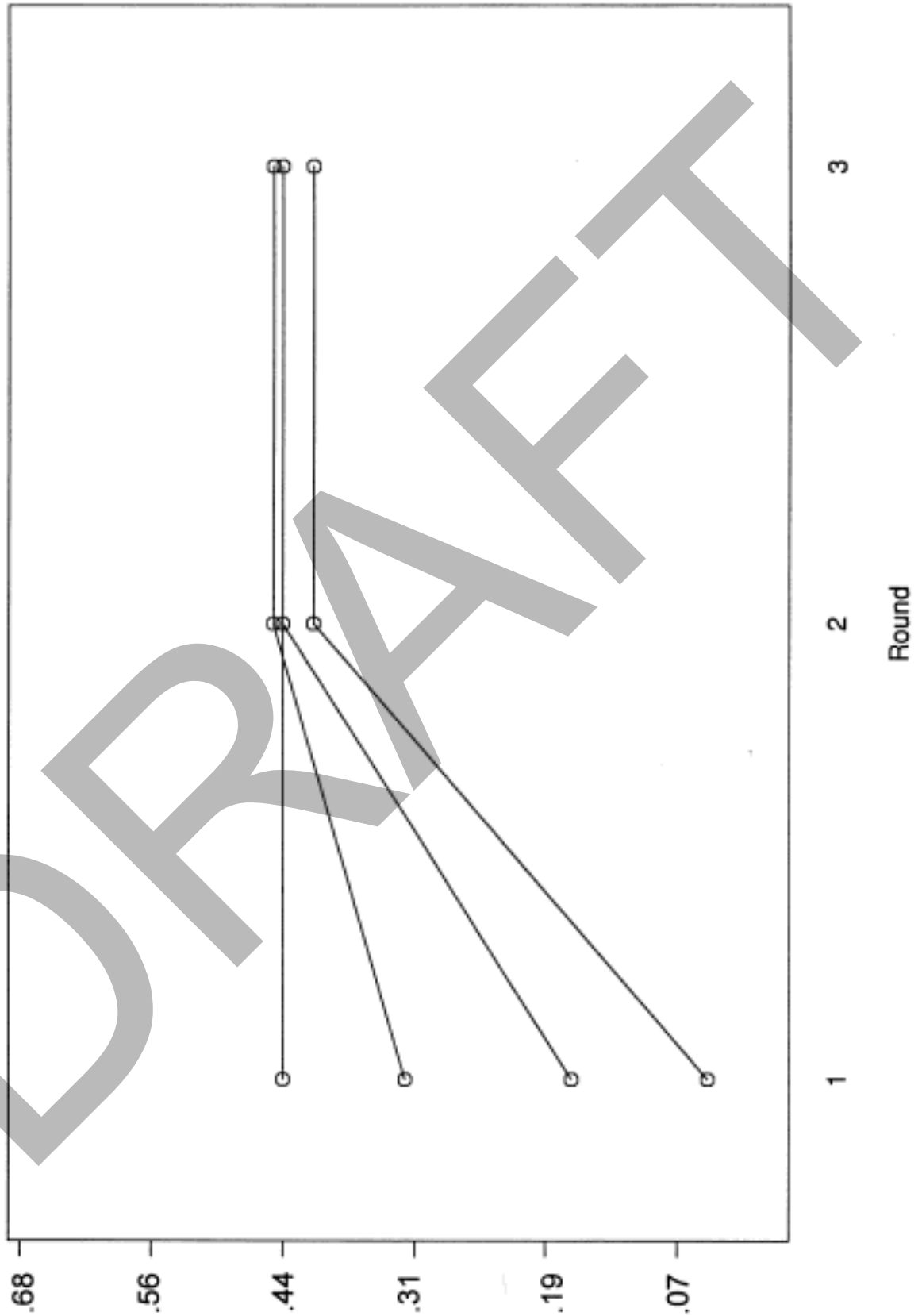


Figure 2. Group 2 Proficient Cutscores for Rounds 1, 2, and 3

(Vertical axis in scale standard deviation units: Mean = 0, SD = 1.)

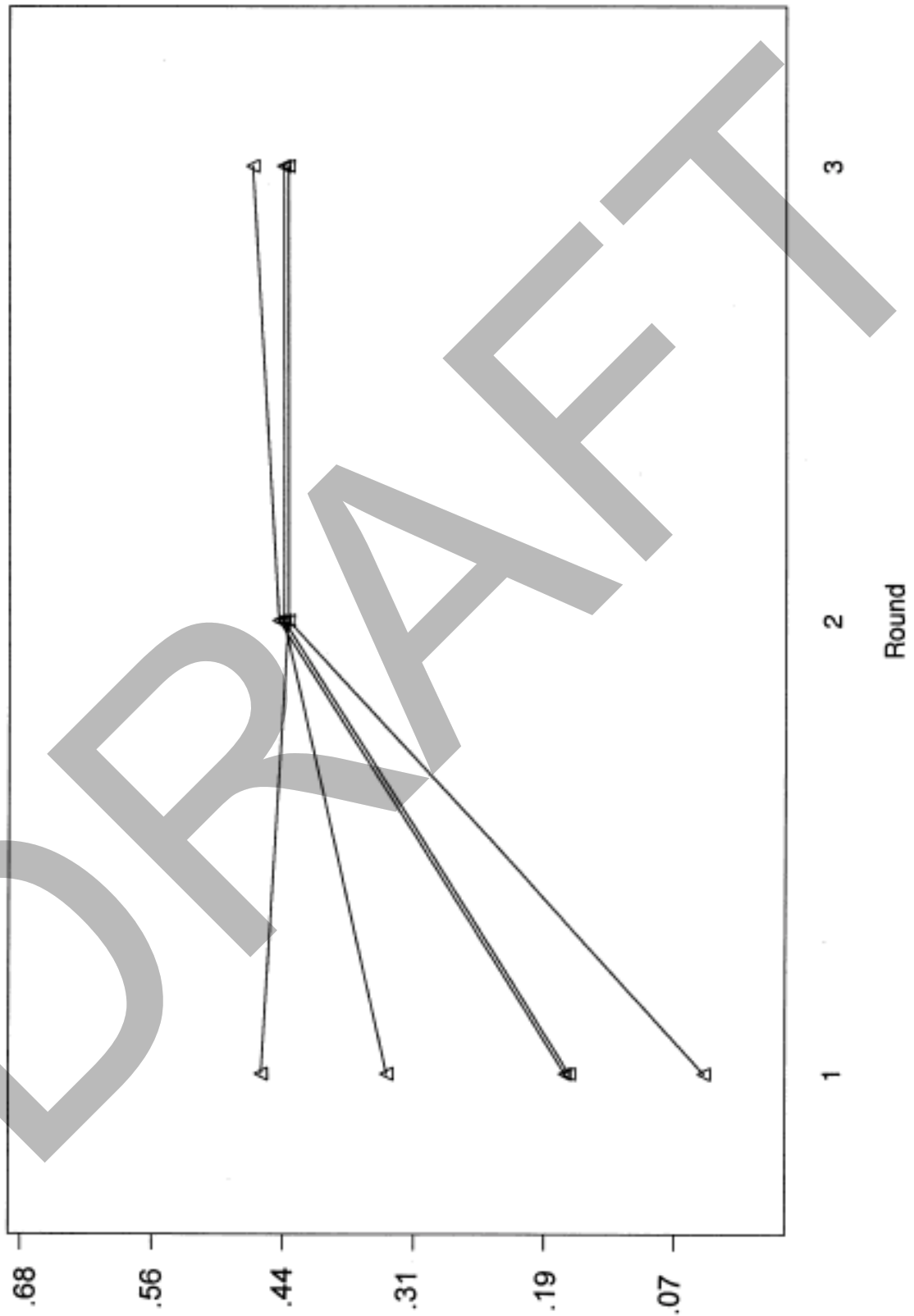


Figure 3. Group 3 Proficient Cutscores for Rounds 1, 2, and 3

(Vertical axis in scale standard deviation units: Mean = 0, SD = 1.)

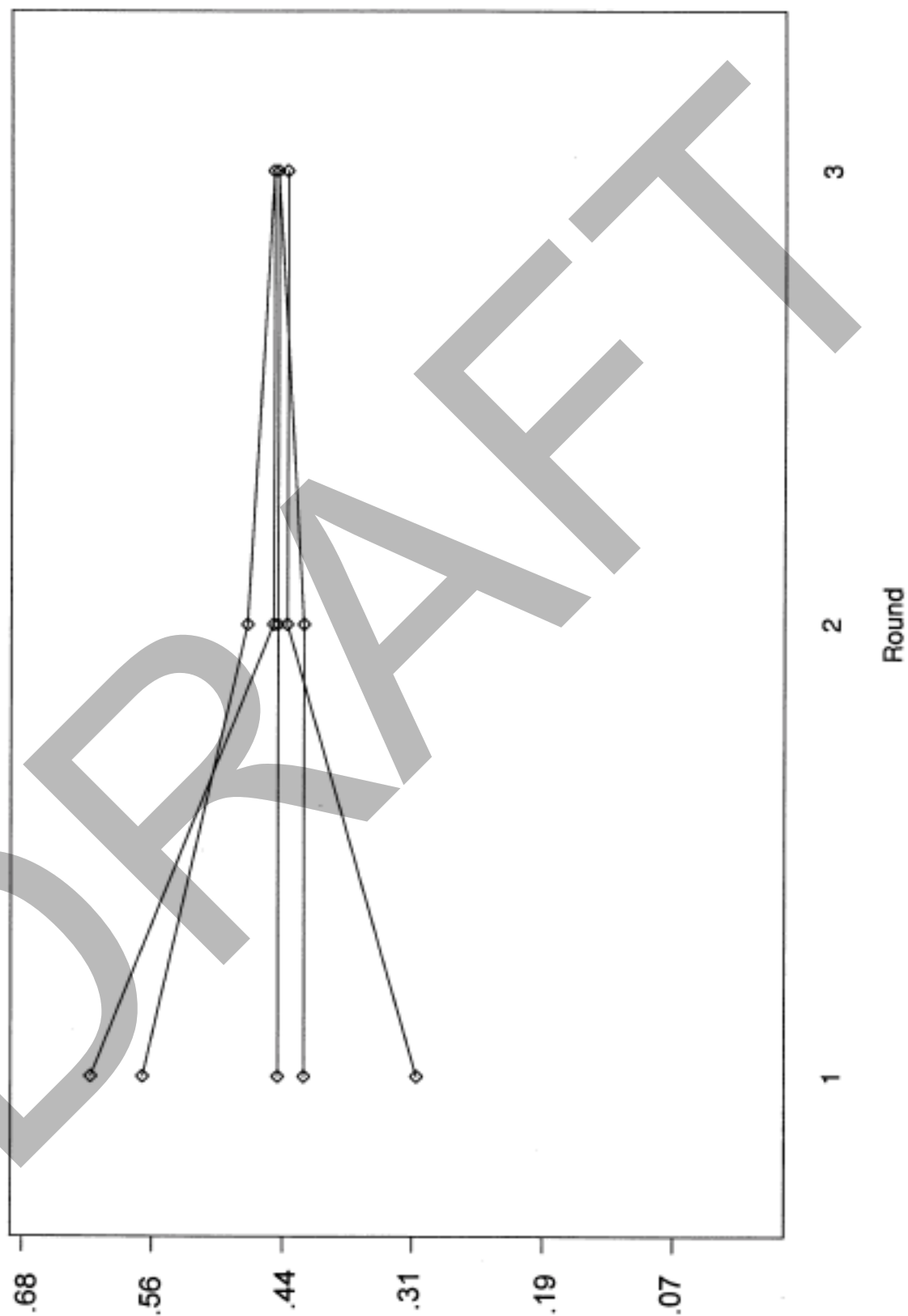


Figure 4. Advanced, Proficient, and Partially Proficient Cutscores of All Participants

