Standard Setting Technical Report

Kaiapuni Assessment of Educational Outcomes (KĀʻEO)

Language Arts and Mathematics

Grades 3 and 4

February 20, 2017

Prepared by Karla Egan

EdMetric LLC

EXECUTIVE SUMMARY

A Bookmark Standard Setting Procedure (BSSP) was conducted in Honolulu, Hawaii on July 23-25, 2016. Dr. Karla Egan, EdMetric, LLC, designed the standard setting workshop collaboratively with Dr. Pohai Shultz from the University of Hawaii, Mañoa. Panelists engaged in content-based discussions to recommend three cut scores that separated four achievement levels.

Personnel from the Hawaii Department of Education (HIDOE) provided policy guidance for the workshop. Staff for the University of Hawaii, Mañoa helped answer content-related questions during the workshop. Dr. Egan answered all process-related questions for the workshop.

Staff from the University of Hawaii, Mañoa recruited 25 panelists to recommend cut scores in either mathematics or Hawaiian language arts (HLA). Thirteen panelists recommended cut scores in Grades 3 and 4 HLA, and 12 panelists recommended cut scores in Grades 3 and 4 mathematics. Panelists completed three full days of work to arrive at recommended cut scores for the KĀʻEO. Day 1 of the standard setting event included an overview of the Hawaiian Immersion Assessment Project, detailed examination of the Grade 3 operational test form, and study of the grade 3 ordered item booklets (OIB). On the morning of Day 2, panelists reviewed threshold ALDs, completed Bookmark training, and engaged in two rounds of ratings for the Grade 3 assessments. Round 3 ratings for the Grade 3 assessments were completed right after lunch. Panelists then reviewed the Grade 4 operational form and the corresponding OIB, and made the first round of ratings on the Grade 4 assessments before closing for the day. On the final day of the standard setting event, panelists completed Round 2 and Round 3 ratings on the Grade 4 assessments. Table 1 (next page) shows the final recommended cut scores and impact data from the standard setting. The final column of the table shows the distribution of student scores on the Grade 3 and Grade 4 Smarter Balanced summative assessments, which were administered to the general Hawaiian student population in 2016. The percent of students scoring at or above Level 3 on the KAEO assessments in 2016 was considerably higher than the percent of students scoring at or above Level 3 on the Smarter Balanced summative assessments, which speaks to the standard setting panelists' emphasis on rigor.

In general, panelists' evaluations of the workshop were positive, with the vast majority of panelists (greater than 90 percent) either agreeing or strongly agreeing with most evaluation statements. Panelists unanimously indicated that the workshop was a valuable professional development experience that increased their overall understanding of the KĀʻEO and would positively impact their instructional practices.

Panelists felt the workshop, overall, was well organized and the facility was adequate for them to complete their work. More significantly, they indicated their work was valuable to them professionally, and the experience would benefit them and their students in the future. Panelists

shared the following statements regarding their overall perceptions of the KĀʻEO standard setting event:

> *"I now have a clear understanding of how the process works as far as the assessment creation and putting it into place. Also an understanding of how hard the OHE works in partnership with UHM and the kaiapuni teachers statewide."*
>
> *"It gave me a much better understanding of everything going on with the testing situation. Mahalo nui!"*
>
> *"Thank you for bringing us together! It was wonderful to hear and learn from other Hawaiian Immersion/Medium educators. It helps me measure how we as a group are performing."*
>
> *"Everything was well thought out and planned."*
>
> *"Learned a lot. I appreciated the opportunity to meet with fellow kumu from various schools."*
>
> *"Mahalo nui no kēia hālawai, ua aʻo au i nā mea he nui mai ia mau kumu kaiapuni ʻē aʻe. Mākaukau au e holomua a hoʻoikaika i ka ʻōlelo Hawaiʻi ma koʻu kula. I look forward to our next Standard Setting hālāwai!"*
>
> *"The workshop coordinators and presenters helped to facilitate the process very well. I truly appreciate the process we used and feel that I can leave the workshop with a greater understanding of the assessment, of the process, and of tasks ahead."*
>
> *"This workshop was an amazing experience and was at the perfect time too! I would love to do this again."*

The BSSP standard setting methodology was implemented for the KĀʻEO standard setting in accordance with best practices and industry standards, using processes and procedures that adhered to the American Educational Research Association/American Psychological Association/National Council on Measurement in Education (AERA/APA/NCME) *Standards for Educational and Psychological Testing*. Additionally, the standard setting was conducted with attention to the requirements of Peer Review Guidance as provided by the United States Department of Education.

TABLE 1: Final Recommended Cut Scores and Impact Data for Grade 3 and Grade 4 KĀʻEO HLA and Mathematics Assessments, and 2016 Smarter Balanced Impact Data

	Cut Scores			Impact Data (% of Students)					Smarter Balanced
KĀʻEO Assessment	**Level 2**	**Level 3**	**Level 4**	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Levels 3 & 4**	**Levels 3 & 4**
Grade 3 Language Arts	494	531	641	51.1%	21.1%	27.4%	0.4%	27.80%	49%
Grade 4 Language Arts	499	520	577	54.0%	14.3%	23.4%	8.3%	31.70%	50%
Grade 3 Mathematics	474	513	546	22.6%	45.4%	21.8%	10.2%	32.00%	53%
Grade 4 Mathematics	487	535	557	47.2%	33.9%	6.4%	12.5%	18.90%	47%

CONTENTS

CHAPTER 1. INTRODUCTION

In compliance with the federal Every Student Succeeds Act (ESSA) (and previously the reauthorization of Elementary and Secondary Education Act – or ESEA – known as No Child Left Behind), HIDOE annually administers Smarter Balanced English language arts/literacy and mathematics summative assessments to students in grades 3 through 8 and high school. Hawaii, however, is home to a unique subset of public schools designed specifically to preserve and promote Hawaiian language and culture. These schools, known as Ka Papahana Kaiapuni schools deliver instruction in the Hawaiian language medium until Grade 5, at which time one hour of each school day is devoted to the English language as a content area. Five of Hawaii's eight major islands provide a K-12 Hawaiian language immersion experience through Kaiapuni schools (either public or charter). Collectively, Hawaii's Kaiapuni schools instruct approximately 2,400 students. All families residing in Hawaii have the option of enrolling their children in a Kaiapuni school.

The Kaiapuni Assessment of Educational Outcomes (KĀʻEO) was administered operationally for the first time in Spring 2016. Using the operational data, a Bookmark Standard Setting Procedure (BSSP) was held University of Hawaii, Mañoa from July 28-30, 2016. Through the BSSP, educators from Kaiapuni schools recommended three cut scores that resulted in four achievement levels: Level 1 (Ho'omaka), Level 2 (Holomua), Level 3 (Mākaukau), or Level 4 (Kelakela). Achievement levels, along with specific descriptions of the knowledge, skills, and competencies a student at that performance demonstrate, will be reported for each individual student. In addition to providing information regarding individual students' performance in Hawaiian language arts and mathematics, KĀʻEO assessment results will be aggregated and will provide the basis for each Kaiapuni school's Strive HI scores.

The KĀʻEO standard setting is particularly notable because it marks the first time performance levels and content associated with state assessments are culturally and linguistically relevant for students in Hawaiian language immersion classrooms. The assessment and the resulting standards and performance levels will present a valid picture of students' knowledge, skills, and abilities that will be informative and meaningful to stakeholders.

PURPOSE AND ORGANIZATION OF THE REPORT

The purpose of this report is to detail all steps involved in the standard setting process for Kaiapuni Assessment of Educational Outcomes (KĀʻEO), Grades 3 and 4 language arts (HLA) and mathematics assessments. The first chapter of the report describes all steps leading up to the standard setting event, including the development of the standard setting plan; selection of the vendor to facilitate the standard setting event; identification of participants in the standard setting event; preparation of materials for standard setting; and selection and pre-training of table facilitators. The second chapter of the report provides details regarding the standard setting event,

including a description of the facility; an overview of the training process; and the results of each round of participant judgments. The third chapter of the report provides a description and summary of results of participant evaluations completed during the standard setting event. The fourth chapter of the report summarizes the steps completed by the Hawaii Department of Education (HIDOE) to finalize recommended cut scores for the KĀʻEO HLA and mathematics assessments. The final chapter of the report addresses the contribution of standard setting to the overall validity argument for the KĀʻEO HLA and mathematics assessments, including evidence that the standard setting was completed with fidelity to the AERA/APA/NCME Standards and adhered to recognized best practices.

LIST OF ACRONYMS

The following acronyms are found throughout the text of this report. The first time an acronym is used, it will be preceded by the term spelled out in its entirety. Each subsequent reference will include only the acronym. This list provides a quick-reference for the reader.

AERA/APA/NCME – American Educational Research Association/American Psychological Association/National Council on Measurement in Education
ALD – Achievement Level Descriptor
BSSP – Bookmark Standard Setting Procedure
ELA – English language arts
ELL – English Language Learner
ESSA – Every Student Succeeds Act
HIDOE – Hawaii Department of Education
IRT – Item Response Theory
KĀʻEO – Kaiapuni Assessment of Educational Outcomes
HLA –Language Arts
NDA – Non-Disclosure Agreement
OIB – Ordered Item Booklet
PLD – Policy Level Definition
RFP – Request for Proposal
SEM – Standard Error of Measurement
SWD – Students with Disabilities
TAC – Technical Advisory Committee
UHM – University of Hawaii – Mānoa
USED – United States Department of Education

RELEVANT DEFINITIONS

The following definitions will apply throughout this report:

Achievement Level Descriptor (ALD) – the knowledge, skills and abilities students at each identified performance level are able to demonstrate.

Bookmark – A physical or virtual marker placed by a standard setting panelist within an ordered item booklet to designate the point at which a target student should demonstrate mastery of all preceding items.

Content Standards – The specific knowledge, skills, and abilities students are expected to demonstrate within a content area and grade level or grade range.

Cut Score – A specific score point that separates two achievement levels.

Every Student Succeeds Act – The reauthorization of the federal Elementary and Secondary Education Act (ESEA) signed into law on December 10, 2015.

Hawaii Common Core Standards – Hawaii's content standards defining the knowledge, skills, and abilities students are expected to demonstrate in English language arts/literacy and mathematics.

Impact Data – The percentage of student scores within each achievement level based on recommended cut scores.

Item Map – A table showing each item in an Ordered Item Booklet, along with the item identification number, the item's page number in the Ordered Item Booklet, the location of the item on the score scale, the score point associated with the item, the item type, the answer key, and the content standard with which the item is associated. During standard setting, panelists add qualitative information regarding what the item or score point measures, and what makes the item more difficult than those that precede it.

Ordered Item Booklet – A group of items representing the constructs measured by an assessment, in ascending order according to item difficulty. Typically, an ordered item booklet consists of items from one or two test forms that are ordered by item difficulty with the easiest item first and the most difficult item last.

Policy Descriptor – Broad descriptions of the policy or program impacts for students within a given achievement level.

Smarter Balanced Assessment Consortium (Smarter Balanced) – A consortium of 15 states, one territory, and the Bureau of Indian Education, originally funded through a 2010 Race to the Top grant, working collaboratively to develop and implement large-scale assessments aligned to the Common Core State Standards in English language arts/literacy and mathematics.

Smarter Balanced Summative Assessments – Large-scale assessments in English language arts/literacy and mathematics developed by the Smarter Balanced Assessment Consortium for students in grades three through eight and high school, typically used for state and federal accountability purposes.

Strive HI – Hawaii's statewide accountability program.

Target Student Descriptor – The characteristics of a student just at the entry of each performance level.

Table Leader or Table Facilitator – A standard setting panelist who serves as a leader at his/her table during a standard setting event to ensure that all standard setting processes are carried out with fidelity and within the given time constraints.

Vertical (or Cross-Grade) Articulation – A review of cut scores across grade levels for an assessment to ensure reasonable progression from one grade to the next.

CHAPTER 2. PREPARATION FOR STANDARD SETTING

EdMetric LLC developed a standard setting plan and standard setting materials in advance of the standard setting workshop. The staff with the UHM Hawaiian Immersion Assessment Project selected and assigned panelists for the workshop. Staff from the UHM Hawaiian Immersion Assessment Project also developed threshold achievement level descriptors (ALDs) prior to the workshop itself.

STANDARD SETTING PLAN

EdMetric LLC worked collaboratively with staff at the UHM Hawaiian Immersion Assessment Project to design the standard setting. The plan was reviewed by Dr. Kerry Englert, Dr. Pohai Shultz, and members of the Hawaiian Immersion Assessment Project Technical Advisory Committee.

The standard setting plan provided a detailed implementation of the Bookmark Standard Setting Procedure (BSSP). The BSSP was recommended based on the technical characteristics of the KĀʻEO and its intended uses. This was also the standard setting methodology employed by the Smarter Balanced Assessment Consortium.

The original standard setting plan provided two alternatives for the implementation of the standard setting workshop. The first design proposed a cut score review while the second design called for a full standard setting. The cut score review would have used the cut scores from the general Hawaii assessment as a starting point for the work of the standard setting panel. The second design allowed panelists to recommend cut scores without reference to the cut scores on the general Hawaii assessments. Instead, this design relied on social moderation using common achievement level descriptors (ALDs). Social moderation is one way of linking test scores (Sireci, 2010). With social moderation, comparability between two sets of cut scores from different tests relies on panelists using a common tool(s). In this case, the tools are the common achievement level descriptors and the common standard setting method. The ALDs were transadapted from the Smarter Balanced ALDs so that the Kaiapuni standard setting had a similar starting point as the Smarter Balanced standard setting.

The standard setting plan for the KĀʻEO called for empaneling two groups of content experts (one for HLA, and one for Mathematics) to complete three rounds of judgments pertaining to assessment items and content. Each content panel would review both Grade 3 and Grade 4, with each content group divided into two tables of 6-7 panelists. Because each content team would consider both grade levels, no cross-grade articulation discussion would be necessary. Table level discussion would be facilitated by a designated table leader. To maintain continuity with the Smarter Balanced summative assessments, panelists would rely upon the Policy Level Definitions resulting

from Smarter Balanced's standard setting work as the foundation for their content discussions. Smarter Balanced Threshold Achievement Level descriptors translated into Hawaiian would provide additional definition of the knowledge, skills, and content students falling within each achievement level would be likely to demonstrate. Panelists would recommend three cut scores delineating four levels of performance: Level 1 (Ho'omaka), Level 2 (Holomua), Level 3 (Mākaukau), or Level 4 (Kelakela). The standard setting plan may be found in Appendix A.

DEVELOPMENT OF THRESHOLD ACHIEVEMENT LEVEL DESCRIPTORS

Prior to the standard setting workshop, staff from UHM developed threshold ALDs. These ALDs are based on the Kaiapuni Standards, and they define the specific knowledge, skills, and abilities students should have at the beginning of an achievement level. To maintain consistency with the general assessment, UHM staff developed the threshold KĀʻEO ALDs by transadapting the Smarter Balanced threshold ALDs.

The Smarter Balanced Threshold ALDs for ELA/literacy and mathematics for Grades 3 and 4 were then translated into Hawaiian and adapted for the Kaiapuni Standards and assessments to provide a preliminary content benchmark to guide panelists in characterizing the knowledge, skills, and abilities of a student entering each achievement level for a given target. These

PANELIST SELECTION

Per the standard setting plan for the KĀʻEO, UHM staff selected 13 panelists for HLA and 12 panelists for mathematics. Each panel completed work for both Grade 3 and Grade 4 for the content area. The KĀʻEO assessment was developed specifically for students in Kaiapuni schools, so panelists were selected from the small pool of teachers currently teaching in Kaiapuni schools. Because the pool of teachers was limited, and the number of schools from which they would be selected was small, it was not necessary to use a formal survey or application process to select participants. Staff from the Hawaiian Immersion Assessment Project met with the group of Kaiapuni principals to emphasize the importance of the standard setting event and to encourage them to recommend teachers to participate as panelists. Following this meeting, an email was sent to Kaiapuni principals across the state to recruit panelists.

Initially, principals were encouraged to submit teachers from grades 3 and 4 only. Because the pool of teachers was so small, it became necessary to include teachers from Grades 1 through 7. Inclusion of teachers from this range reflects the composition of Kaiapuni classrooms and schools. Due to limited resources, and the small Kaiapuni student population, many Kaiapuni teachers work with students in combined classes, or teach different classes across the entire grade range. Including teachers from across the grade range also supported the proposed model of having one panel per

content area complete standard setting for both grade 3 and grade 4. The majority of panelists in each content area (9 in language arts and 10 in mathematics) were classroom teachers in grade 3 or grade 4, or a grade range including grade 3 or grade 4. Two panelists (one in each content area) also served as resource teachers.

The primary diversity consideration for selecting panelists was representation from a wide range of Kaiapuni schools. Table 2.1 shows the number of panelists per content area, disaggregated by island and school. It should be noted that Oahu is Hawaii's most populated island, as well as the island with the greatest number of Hawaiian language immersion schools; therefore, a larger proportion of panelists are from the island of Oahu. Twelve of Hawaii's 14 language immersion schools were represented among the panelists.

Table 2.1. Panelists Per Content Area, by Island and School

Island and School	Language Arts	Mathematics	Total
Hawaii	**2**	**2**	**4**
Ke Kula ʻO Nāwahīokalaniʻōpuʻu Iki LCPS	1		1
Ka 'Umeke Kā'eo	1		1
Ke kula ʻo ʻEhunuikaimalino		1	1
Ka ʻUmeke Kāʻeo PCS		1	1
Kauai	**1**		**1**
Kawaikini	1		1
Maui	**3**	**2**	**5**
Paia		1	1
Ke Kula Kaiapuni ʻO Nāhiʻenaʻena	1		1
Pāʻia School	1		1
Kula Kaiapuni ʻo Nāhiʻenaʻena	1		1
Pāʻia		1	1
Molokai	**1**	**2**	**3**
Kualapuu		2	2
Kula Kaiapuni o Kualapuʻu	1		1
Oahu	**6**	**6**	**12**
Anuenue		1	1
Hauula		1	1
Pūʻōhala	2	1	3
Hauʻula	1		1
Ke Kula Kaiapuni ʻo Waiau	1	1	2
Ke Kula Kaiapuni ʻo Nānākuli	1		1
Kula Kaiapuni ʻo Ānuenue	1		1
Ke Kula Kaiapuni o Pūʻōhala		1	1

Ke Kula Kaiapuni o Hauʻula		1	1
Total	**13**	**12**	**25**

Panelists in each content area were assigned to one of two tables based on the geographic location of their school and the grade level(s) they taught, with a goal of having equal representation of islands, schools, and grade levels at each table.

Table 2.2 shows the distribution of panelists at each table for each content area by school, island, and grade level(s) currently teaching.

Table 2.2. Distribution of Panelists Per Table by School, Island, and Grade Level(s)

School	Island	Grade(s)
Language Arts Table A		
Ke Kula ʻO Nāwahīokalaniʻōpuʻu Iki	Hawaii	2
Ke Kula Kaiapuni o Pūʻōhala	Oahu	3
Pāʻia	Maui	4
Ke Kula Kaiapuni ʻo Nānākuli	Oahu	4-6
Ke Kula Kaiapuni o Hauʻula	Oahu	1-3/resource teacher
Kula Kaiapuni ʻo Nāhiʻenaʻena	Maui	1
Kula Kaiapuni o Kualapuʻu	Molokai	5
Language Arts Table B		
Kawaikini	Kauai	4
Kula Kaiapuni ʻo Ānuenue	Oahu	3
Ka 'Umeke Kā'eo	Hawaii	4
Ke Kula Kaiapuni ʻo Waiau	Oahu	3
Ke Kula Kaiapuni o Pūʻōhala	Oahu	2
Kula Kaiapuni ʻo Nāhiʻenaʻena	Maui	4-5
Mathematics Table A		
Kula Kaiapuni ʻo Ānuenue	Oahu	3/resource teacher
Pāʻia	Maui	4
Ke Kula Kaiapuni o Pūʻōhala	Oahu	4
Ke Kula Kaiapuni o Hauʻula	Oahu	5-6
Kula Kaiapuni o Kualapuʻu	Molokai	3-6
Ka 'Umeke Kā'eo	Hawaii	3
Mathematics Table B		
Pāʻia	Maui	3
Ke kula ʻo ʻEhunuikaimalino	Hawaii	2
Ke Kula Kaiapuni o Hauʻula	Oahu	3-4
Kula Kaiapuni o Kualapuʻu	Molokai	3-7
Ke Kula Kaiapuni o Pūʻōhala	Oahu	1-4
Ke Kula Kaiapuni ʻo Waiau	Oahu	4

After panelists were divided into two table groups per content area, a table facilitator for each table was identified. Like panelists in general, table facilitators were selected to ensure broad representation from the islands. Additionally, all table facilitators were classroom teachers who had participated in the KĀʻEO development process as item writers, standards developers, and/or alignment study panelists.

Special Education. During recruitment, special efforts were made to identify and recruit special education teachers. One panelist in each group had special education experience.

Teacher Experience. On average, the panelists in the HLA group had 7.5 years (standard deviation=4.8 years) of teaching experience. The mathematics panelists had an average of 12 years (standard deviation =7.6 years) of experience.

All panelists were classroom teachers. Instructional coaches were invited to participate, but none were able to attend the workshop.

PREPARATION OF MATERIALS

Prior to the standard setting event, Dr. Egan developed presentations and materials for use prior to and during standard setting. These materials included: agendas; PowerPoint slides for the opening session and Bookmark training; OIBs and item maps; and surveys for panelist feedback. All materials and presentations were submitted to UHM for review prior to being finalized.

Presentations. PowerPoint slides were created for onsite training during the standard setting event. The Bookmark Training presentation, developed by Dr. Egan, provided a detailed overview of the BSSP. Appendix B includes the presentation slide deck.

Agenda. A high-level agenda was created for the standard setting event. This agenda is included in Appendix B.

Surveys and Evaluations. EdMetric completed surveys to administer following the orientation, the training, and each round of bookmark rating. A final workshop evaluation was administered at the end of the workshop. These surveys are presented in Appendix B.

OIBs. EdMetric LLC prepared the OIBs and item maps for each grade level and content area. Since each content/grade-level assessment is a single fixed form, it was reasonable to include all items from each assessment in the OIBs. Within an OIB, each item was presented on a single page, and items were ordered in ascending order of difficulty. Items with multiple score points were presented multiple times in the OIB, once for each score point. To order the items, it was necessary to find each item's location on the test scale where students had a 50/50 chance of answering each item correctly. In other words, the item's difficulty estimate was passed on a .50 response probability. The item location was estimated using Spring 2016 operational data. OIBs were created in hard-copy

format and were color-coded by grade level and content area. Stimuli for the items were printed in a separate stimulus booklet for each grade level/content area assessment.

Item Maps. EdMetric LLC also created item maps based on the OIBs. Item maps presented the items in table format in the same order as their presentation in the OIB, along with their location on the score scale; the number of score points associated with the item; the item number on the operational assessment form; the item identification number; the stimulus to which the item is connected; the content standard(s) to which the item maps; and the item type. The final two columns of the item maps were left for panelists to complete by answering the questions, "What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?" and, "Why is this item or score point more difficult than the items that precede it?" Item maps were prepared in electronic format as Excel spreadsheets.

Figure 2.1 shows a portion of a sample item map for the KĀʻEO.

	A	B	C	D	E	F	G	H	I	J	K
4	Difficulty (OIB Page Number)	Location	Score Point	Test Item #	Problem ID	Story/Stimulus	Code	Content Classification	Item Type	What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?	Why is this item or score point more difficult than the items that precede it?
5	1	382	1	11	90	Nā Hoapili	1.2	3/4.O.7	SC		
6	2	403	1	9	85	Nā Hoapili	1.1	3/4.O.2	SC		
7	3	405	1	12	91	Nā Hoapili	1.2	3/4.O.6	SC		
8	4	434	1	1	93	Hū Aku Nō!	2.1	3/4.MO.1	SC		
9	5	444	1	4	94	Hū Aku Nō!	2.1	3/4.MO.1	STV		
10	6	447	1	45	145	Mālama ʻĀina	4.2	3.H.1	ER		
11	7	449	1	43	150	Mālama ʻĀina	4.2	3.H.7	SC		
12	8	455	1	21	162	No ka ʻOhi ʻŌpala	3.3	3K12	SC		
13	9	465	1	20	161	No ka ʻOhi ʻŌpala	3.2	3K9	SC		
14	10	470	1	33	125	Hoʻokahi Leo	4.1	3.H.4	SC		
15	11	470	1	7	89	Nā Hoapili	1.2	3/4.O.6	SC		
16	12	473	1	48	141	A ʻIke ʻIa He Manu	4.2	3.H.7	SC		
17	13	476	1	46	152	A ʻIke ʻIa He Manu	4.3	3.H.13	SC		
18	14	479	1	24	117	No ka ʻOhi ʻŌpala	3.2	3.K.10	SC		
19	15	483	1	42	156	Mālama ʻĀina	4.3	3.H.14	SC		
20	16	492	1	18	107	Hawaiʻi Pono	2.2	3/4.MO.2	SC		
21	17	494	1	8	84	Nā Hoapili	1.1	3/4.O.5	FTG		
22	18	498	1	5	99	Hū Aku Nō!	2.1	3/4.MO.1	SA		
23	19	501	1	36	123	Hoʻokahi Leo	4.1	3.H.5	SC		
24	20	502	1	37	137	Hoʻokahi Leo	4.1	3.H.6	SC		
25	21	506	1	47	122	A ʻIke ʻIa He Manu	4.1	3.H.6	SC		
26	22	511	1	22	110	No ka ʻOhi ʻŌpala	3.1	3.K.4	SC		
27	23	511	1	14	92	Nā Hoapili	1.2	3/4.O.7	SA		
28	24	512	1	16	97	Hawaiʻi Pono	2.1	3/4.MO.1	FTG		
29	25	519	1	6	83	Nā Hoapili	1.1	3/4.O.5	SC		
30	26	520	1 of 2	44.1	138	Mālama ʻĀina	3.3	3.H.6	ER		
31	27	520	1	52	136	A ʻIke ʻIa He Manu	4.1	3.H.3	SC		
32	28	523	1	41	158	Hoʻokahi Leo	4.3	3.H.12	SC		
33	29	525	1	50	159	A ʻIke ʻIa He Manu	4.3	3.H.10	SC		
34	30	528	1	15	95	Hawaiʻi Pono	2.1	3/4.MO.1	SC		
35	31	531	1	28	114	Pilialoha	3.1	3.K.6	SC		
36	32	531	1	30	119	Pilialoha	3.2	3.K.7	SC		
37	33	532	1 of 4	31.1	126	Pilialoha	3.1/3.2/3.3	3.K.10/3.K	PT		

Figure 2.1. Sample KA'EO Item Map

PANELIST DATA ENTRY

Prior to the meeting, an online spreadsheet tool was created to capture all panelist data entry. Panelists used the online tool throughout the standard setting. This tool allowed panelists to enter their bookmark ratings directly into the spreadsheet. The tool automatically checked all bookmark ratings to ensure that the ratings associated with Level 2 was lower than the rating for Level 3, etc… Ratings were also flagged if an entry was left blank. The room facilitator monitored panelist ratings in real time. If the ratings were flagged for possible incorrect entry or incomplete entry, then the room facilitator asked the panelist to confirm or correct their entry. Panelists also completed all surveys and evaluations using the online tool.

To complement the online tool, each content area received a Control Panel. This Control Panel was a PDF with links to the online tool where panelists would find the data entry forms for ratings or for evaluations.

Figure 2.2 shows a portion of the Control Panel for the HLA group.

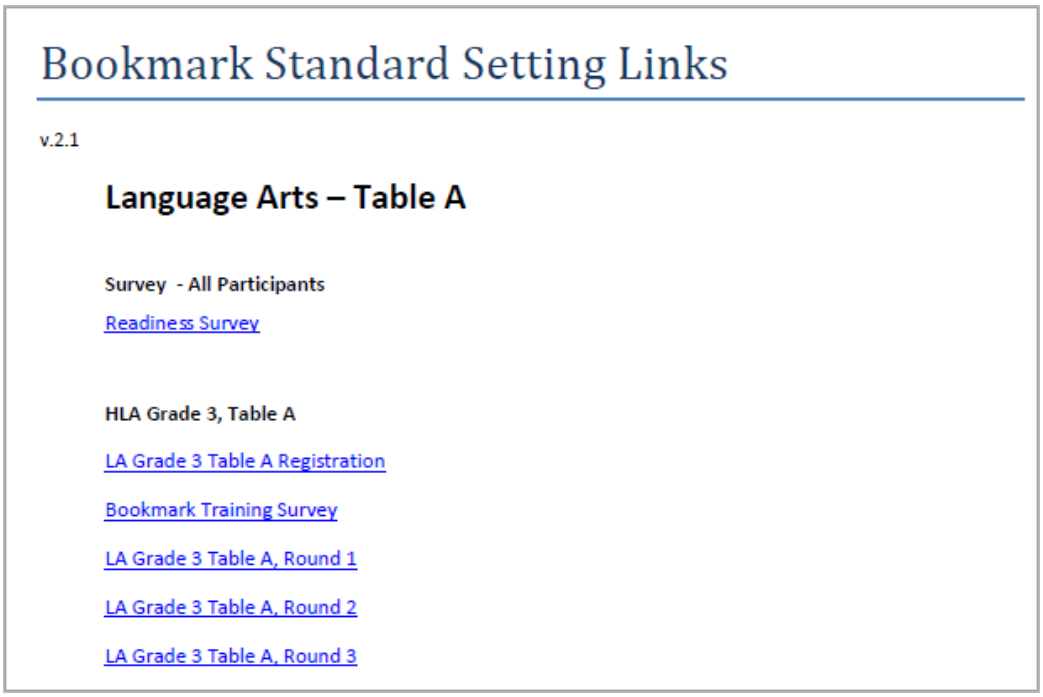


Figure 2.2. Portion of Standard Setting Control Panel PDF

DATA TOOLS

Once panelists completed data entry, the data were immediately downloaded for use in an offline Excel spreadsheet. The data from the online tool was automatically input into the offline Excel spreadsheet. Formulas, tables, and graphics were created prior to the workshop so they would be efficiently computed and populated during the workshop.

TABLE FACILITATOR ORIENTATION AND MATERIALS

Since most table facilitators were new to the standard setting process and unfamiliar with their roles, Dr. Egan provided them with a document that briefly summarized the steps involved in the BSSP, explained the Ordered Item Booklet (OIB) and item maps that would be used in the process, and described their roles and responsibilities as table facilitators. This document was accompanied by a detailed agenda with specific annotations for table facilitators. To allow table facilitators to develop a more thorough understanding of the standard setting process, Dr. Egan developed a Table

Facilitator Training presentation which all Table Facilitators accessed via recorded webinar prior to the beginning of the standard setting event. Table facilitator preparation materials are included in Appendix C.

CHAPTER 3: STANDARD SETTING IMPLEMENTATION

OVERVIEW OF WORKSHOP

The KĀʻEO standard setting event was held on July 23-25, 2016, at the University of Hawaii, Mañoa. As shown in the High Level Agenda (Appendix B), panelists arrived at 9:00 a.m. on July 23rd, and completed three full days of work to arrive at recommended cut scores for the KĀʻEO. Table facilitators arrived 30 minutes prior to the beginning of the standard setting event to meet with workshop facilitators. Day 1 of the standard setting event included an overview of the Hawaiian Immersion Assessment Project, detailed examination of the Grade 3 operational test form, and study of the grade 3 OIB. On the morning of Day 2, panelists reviewed threshold ALDs, completed Bookmark training, and engaged in two rounds of ratings for the Grade 3 assessments. Round 3 ratings for the Grade 3 assessments were completed right after lunch. Panelists then reviewed the Grade 4 operational form and the corresponding OIB, and made the first round of ratings on the Grade 4 assessments before closing for the day. On the final day of the standard setting event, panelists completed Round 2 and Round 3 ratings on the Grade 4 assessments.

FACILITIES AND SECURITY OF MATERIALS

The KĀʻEO standard setting event was held at the University of Hawaii, Mañoa. Three rooms were reserved for use during the event – one large room for the opening session and large group discussions, and a smaller breakout room for each content area. Each breakout room included two tables placed sufficiently far apart for participants to complete their work without disrupting one another.

Participants were asked to provide a personal laptop to access online test forms and standard setting tools. All work was completed through a secure, cloud-based location (Google Drive) which participants accessed via links in a PDF provided on a secure thumb drive. Prior to checking out secure materials, all participants were required to sign a non-disclosure form that included acknowledgment that they would not download any materials from the cloud onto their personal laptops. All secure materials were numbered and color-coded. Secure materials were checked out just prior to beginning work with assessment items, and collected by table facilitators and returned to the secure operations room between working days.

ROLES AND RESPONSIBILITIES

The BSSP standard setting has several roles with differing responsibilities, including: lead facilitator, room facilitator, content facilitator, table leader, and panelists.

Lead Facilitator. Dr. Egan served as the lead facilitator during the workshop. The lead facilitator is charged with the overall implementation of the workshop, which includes providing orientation, providing training, answering questions, distributing materials, and attending to other needs as they arise.

Group Facilitator. Dr. Egan served as the group facilitator for HLA in addition to her role as the overall facilitator. Dr. Kerry Englert (Seneca Consulting) served as a group facilitator for mathematics. Group facilitators ensure each breakout room is running smoothly. Group facilitators provide additional training, clarify points of confusion, ensure panelists follow the agenda, and guide group-level discussions. Group facilitators are responsible for answering all questions related to the test data and standard setting process.

Content Facilitator. Pono Fernandez served as the content facilitator for the standard setting. The content facilitator is responsible for answering all questions related to test items and the test itself.

Table Facilitator. One panelist at each table was identified as a table facilitator. Prior to the standard-setting event, table facilitators were responsible for completing table facilitator training provided by Dr. Egan via recorded webinar. During the standard-setting event, table facilitators were responsible for ensuring item security by overseeing secure materials check-out and return and monitoring the use of electronic devices during standard setting activities. Facilitators were also responsible for leading discussions with integrity and objectivity at their tables, and ensuring that panelists stayed on task according to the agenda.

Panelists. Expert panelists (classroom teachers and resource teachers from Kaiapuni schools) were responsible for reviewing the content and assessment items, providing thoughtful and objective discussion of the assessment items, setting Bookmarks within the OIBs, discussing impact data, and recommending final cut scores for the assessments.

State Staff. Also present during the standard setting event were HIDOE Office of Hawaiian Education Director, Kau`i Sang, and Dr. Kalehua Krug, Director of the Hawaiian Immersion Assessment Project, and Dr. Pohai Shultz, Principal Investigator of the Hawaiian Immersion Assessment Program. Their role was to set the stage for standard setting by providing the context for development of the KĀʻEO and to facilitate a discussion of Target Student Descriptors and ALDs. Both UH and HIDOE staff served as content and historical test development resources throughout the standard setting event.

Each of these roles must be fulfilled by trained and knowledgeable staff in order to successfully conduct a standard setting workshop. Table 3.1 shows the roles, the person who fulfilled it, and the qualifications of each person.

Table 3.1. Qualification of BSSP Lead Staff

Role	Person	Qualifications
Lead Facilitator/ HLA Facilitator	Dr. Karla Egan	Dr. Egan has designed and lead over 40 standard setting workshops. She has implemented all major standard setting methods, including BSSP, Body of Work, and Modified Angoff.
Math Facilitator	Dr. Kerry Englert	Dr. Englert has worked as a psychometrician both as a consultant and for the Colorado Department of Education. She has worked on developing assessments, leading educator panels, scaling and analyzing assessment data for over 15 years.
Content Facilitator	Pono Fernandez	Ms. Fernandez has a Master of Arts degree in Hawaiian language and is a fluent Native Hawaiian speaker. She led many of the KĀʻEO content development efforts.

TABLE FACILITATOR TRAINING

Table facilitators' training occurred the week prior to the standard setting event via recorded webinar (see Appendix B for the slide presentation). During the recorded webinar, table facilitators were trained regarding security procedures. They were instructed that they would be responsible for collecting participants' signed non-disclosure agreements and that participants were not to have access to cell phones or other electronic devices during standard setting. Dr. Egan also explained that all secure materials would be color coded and table leaders would be responsible for asking participants to put their names on secure materials, and for accounting for secure materials check-out and return. Next, table leaders were provided with a high-level overview of the standard setting process, who is involved, and why it is important to set standards. The overview emphasized the importance of referencing achievement levels to content standards, and explained that three cut scores and four levels of performance would be identified for the KĀʻEO. Next, Dr. Egan provided table facilitators with a description of the BSSP, including an explanation of the OIB and item maps. She walked table facilitators through the process of studying the OIB and asking the questions, "What do you know about a student who responds successfully to this item; that is, what skills must a student have in order to know the correct answer?" and "What makes this item more difficult than preceding items?" to guide small group discussions.

STANDARD SETTING EVENT DAY 1

TABLE FACILITATORS' MEETING

Prior to the opening session on Day 1 of the standard setting event, Drs. Egan and Englert met with table facilitators to review the annotated agenda (Appendix C), to explain the process for secure materials sign-out and auditing, and to ensure all table facilitators fully understood their roles and responsibilities.

OPENING SESSION

The KĀʻEO standard setting event began on Day 1 with a general session for all participants. Kau`i Sang, Director of the Office of Hawaiian Education, and Dr. Kalehua Krug, head of the Hawaiian Immersion Assessment Project, provided opening remarks for the standard setting event, emphasizing the importance of the work for HIDOE and for native groups across the country who recognize the importance of assessing students in their native language. Dr. Pohai Kukea Shultz, Principal Investigator of the Hawaiian Immersion Assessment Program, provided an overview of the developmental history of the KĀʻEO to provide context for participants who had not been involved in the test development process.

Following Dr. Shultz's presentation, Dr. Egan provided an orientation to the Bookmark process and to the Bookmark materials. Dr. Egan provided panelists with an overview of how the Bookmark process would be implemented during the three-day workshop.

She led the panelists through a practice session with items from the National Assessment of Educational Progress (NAEP). With the NAEP items, Dr. Egan introduced the OIB and the item maps. She led the panelists through a discussion of the NAEP items in the practice OIB. She modeled the types of discussions panelists should have once the returned to their breakout room. She also modeled how they should complete the item maps that would be provided to them.

GRADE 3 ROUND 1

Following the opening session, participants moved to their assigned tables in each of the two content-area breakout rooms to begin Round 1. The Round 1 activities included: studying the content standards, Grade 3 operational assessment, and Grade 3 OIB and item map; reviewing threshold ALDs; participating in Bookmark training; and placing Round 1 Bookmarks.

At the beginning of Round 1, table facilitators gave panelists an opportunity to introduce themselves; panelists signed non-disclosure agreements, and signed out their secure materials. Table facilitators assigned numbered packets of secure materials including OIBs, stimulus booklets, and item maps for each grade for the content area to each panelist in consecutive order. Additionally, table leaders identified a scribe to take notes for the group during table discussions.

After all questions were answered, all panelists completed a survey to assess their perceived understanding of the purpose of the standard-setting workshop and their readiness to study the OIBs. Surveys were completed and submitted online, with panelists in each content area room indicating their table number so that workshop facilitators could address questions individually as necessary. Table 3.2 shows the results of the readiness survey completed by HLA and mathematics panelists. The results in Table 3.2 show the panelists did not want additional training nor did they have additional questions regarding the standard setting process.

Table 3.2. Results from Standard-Setting Readiness Survey (% who agree or Strongly Agree)

	HLA (n=13)	Mathematics (n=12)
1. The orientation session provided a clear overview of the standard setting process	100%	100%
2. I understand the goals of the standard setting workshop.	100%	100%
3. I understand my role in the standard setting workshop.	92.3%	100%
4. The orientation session provided a clear explanation of the development of the Kaiapuni assessment	100%	100%
5. I understand the result of the standard setting will be used to support the reporting of the Kaiapuni assessment results.	100%	100%
6. I understand how to study the items in the ordered item booklet.	92.3%	100%
The percentage of panelists who answered "yes" are reported in the following questions.		
7. I would like additional training on studying the ordered item booklet.	0.0%	0.0%
8. I have additional questions on materials presented during the opening session.	0.0%	0.0%

Study of the Kaiapuni Standards

Workshop facilitators and UHM staff in each room then directed panelists to the Kaiapuni Standards for the applicable content area and opened the floor for any questions panelists might have. All panelists were asked to review the content standards prior to the workshop, so only a small amount of time was allotted for addressing questions pertaining to the standards.

Study of the Grade 3 Operational Form

All panelists accessed the operational form in the same online format experienced by students during the operational administration of the assessment. Panelists used their personal laptops to access the assessments, using credentials provided by the workshop facilitators and UHM staff.

Workshop facilitators walked participants step-by-step through the process of logging into the appropriate assessment and instructed them to move through each item as a student would.

Study of the OIB and Item Map

Table facilitators then facilitated a discussion of the Grade 3 OIB, beginning with the first (easiest) item in the OIB and progressing through the last (most difficult). Table facilitators led the panelists at the table through a discussion of two questions for each item:

- What does this item measure? That is, what do you know about a student who responds successfully to this item?
- Why is this item more difficult than the preceding items?

Items with multiple score points were discussed at each score point within the OIB. As panelists discussed each item, the scribe for the table noted the table's collective response to the questions on the item map. Scribes accessed the item maps via the cloud and recorded notes electronically. Panelists had access to OIBs and stimuli in electronic format via the cloud, and also had hard copies of the OIBs and stimulus booklets. Throughout the review process, table facilitators monitored time and ensured the discussion continued to progress at a reasonable rate within the allotted time. Workshop facilitators were available to respond to questions about the review process and to collect any questions regarding policy for response by UHM and/or HIDOE staff.

Collection of Secure Materials and Daily Debrief

Following completion of Day 1 activities, table facilitators collected secure materials from panelists and met briefly with workshop facilitators to discuss any challenges that occurred during the day, to share what went well and what could be improved, and to provide feedback regarding the quality of the workshop.

STANDARD SETTING EVENT DAY 2

GRADE 3 ROUND 1, CONTINUED

TARGET STUDENT DESCRIPTORS

After completing their review of the Grade 3 OIB, panelists in each content area participated in a discussion of Target Student Descriptors for their content area, led by UH-Mañoa staff. The goal of the target student discussion was to encourage panelists to consider the characteristics of the student who demonstrates just enough content knowledge to be considered proficient.

BOOKMARK TRAINING

Following the target student discussion, Dr. Egan provided Bookmark training to panelists in both content areas. Bookmark Training began with a reminder of the definition of the target student for Level 3 on the KĀʻEO and a refresher about the organization of the OIB. Dr. Egan then explained the meaning of "placing a bookmark" for Level 3, as indicated in Figure 3.1.

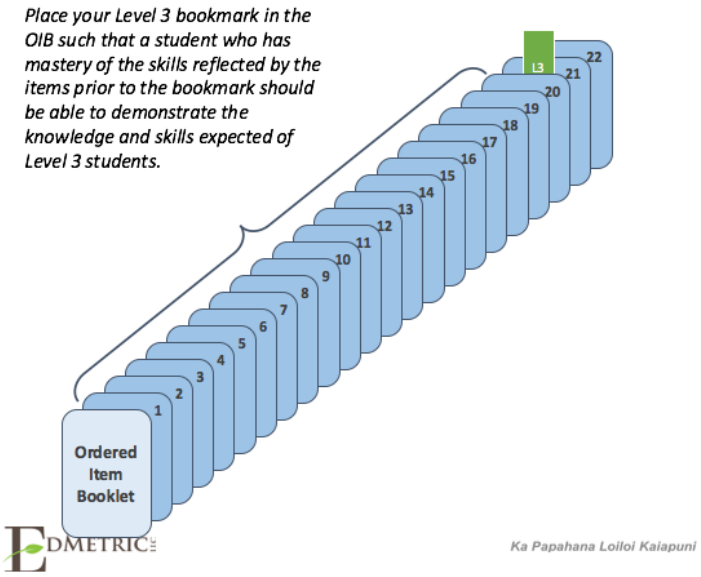


Figure 3.1. Level 3 Bookmark Placement

Training then addressed the connection between Bookmark placement and item location, establishing the item location as the scale score necessary for a student to have a 50/50 chance of answering the item correctly (as directed by the response probability identified in the standard setting plan). Workshop facilitators then demonstrated the process by which panelists would electronically "set" their Bookmarks.

BSSP Readiness Surveys

Prior to beginning Bookmark placement, all participants completed an evaluation to indicate their perceived readiness to begin the Bookmark task. Participants accessed and submitted their responses to the BSSP Readiness Survey via a link on their secure thumb drive. Table 3.3 shows the results of the online BSSP Readiness Survey. All panelists indicated that they understood how to place Bookmarks and that they were ready to place their Round 1 Bookmarks.

Table 3.3. Results from BSSP Readiness Survey (% Who Agree or Strongly Agree)

	HLA (n=13)	Mathematics (n=12)
1. I reviewed and was provided the opportunity to ask questions about and discuss the Target Student descriptors.	100%	100%
2. I participated in bookmark training and had the opportunity to ask questions and discuss the meaning of the bookmarks.	100%	100%
3. I understand how to place my bookmarks.	100%	100%
4. I understand I will have opportunities to change my bookmark in Rounds 2 and 3.	100%	100%
The percentage of panelists who answered "yes" are reported in the following questions.		
5. I would like additional training on placing my bookmarks for Round 1.	0.0%	0.0%
6. I have additional questions that I would like to ask before placing my Round 1 bookmarks.	0.0%	0.0%

Round 1 Bookmark Placement, Grade 3

When all participants felt comfortable with their understanding of the BSSP and how to place their Bookmarks, they began with Round 1 Judgments for Grade 3 assessments. For the KĀʻEO standard setting event, panelists input their Bookmarks into the online data entry tool. Each panelist was required to complete a registration, as shown in the example in Figure 3.2, prior to accessing or using the bookmark rating system.

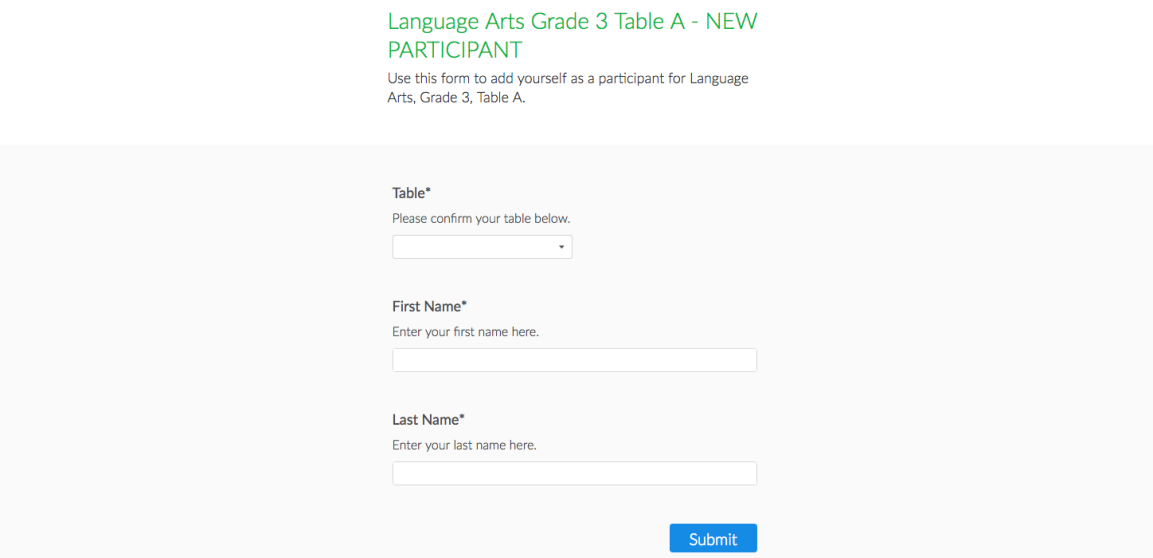


Figure 3.2. Online Panelist Registration

Panelist registration was specific to Grade Level, Content Area, and Table, in order to allow median Bookmark placements and impact data for specific tables to be easily determined. Following registration, each panelist accessed the system via the appropriate links in their Control Panel (a PDF form provided on their secure thumb drive, shown in Figure 2.2) for each round's activities. An example of the system's Round 1 Bookmarks form is shown in Figure 3.3.

Panelists were instructed to first place their Level 3 Bookmarks, followed by Levels 2 and 4. Panelists were asked to write their Bookmark placements for each level on a paper rating form (Appendix B) before entering it into the online system. All ratings were completed independently, and without discussion. Finally, as shown in Figure 3.3, panelists accessed the Bookmark rating forms using the appropriate link in the PDF form and entered their Round 1 Bookmarks in the online system by indicating the item number in the OIB after which they wished to place their Bookmark for Level 2, Level 3, and Level 4.

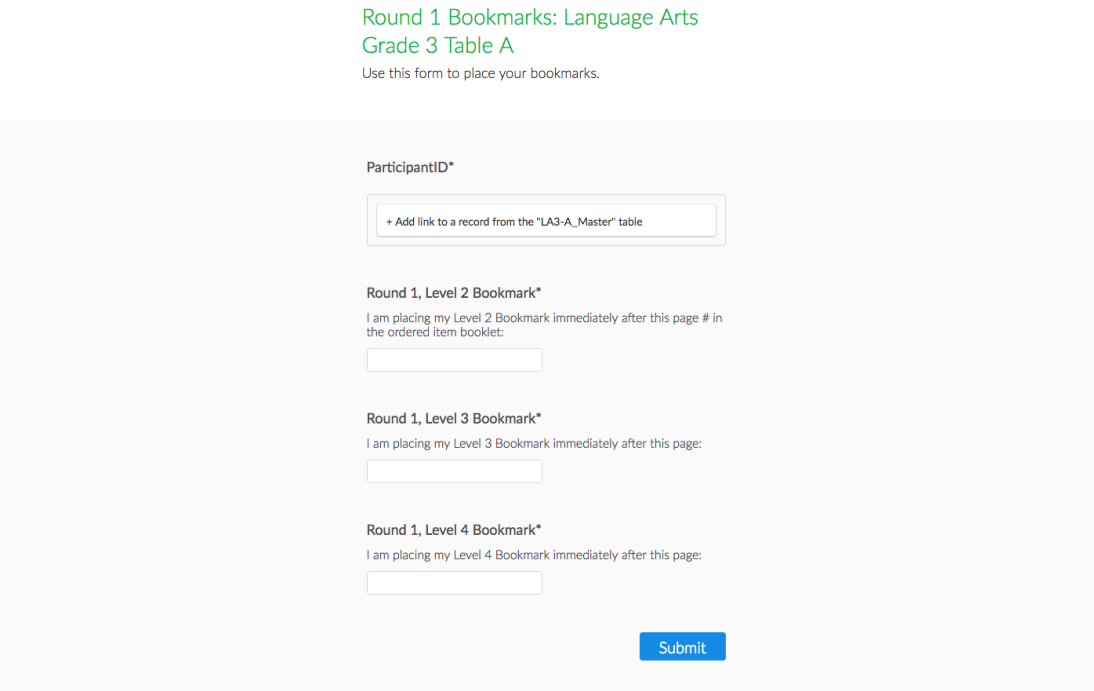


Figure 3.3. Online Bookmark Placement Form

A unique online Bookmark placement form was provided for each grade level, content area, and table for each round of judgments.

Following Round 1 Bookmark placements, workshop facilitators imported panelists' Bookmarks into the Bookmark Processor system to analyze the data and determine the cut scores associated with the initial Bookmark placements. The Bookmark Processor is an electronic system by which each panelist's Bookmark rating can be imported for each judgment round and is identifiable by panelist identification number, table, grade level, and content area. Scale scores can then be determined for each Bookmark placement according to the established criteria, and results can be aggregated by round in a variety of configurations and presented graphically to panelists.

Table 3.4 shows the median Grade 3 cut scores for each content area associated with Round 1 judgments. Detailed judgments may be found in Appendix E.

Table 3.4. Grade 3 Round 1 Cut Scores

Content Area	Level	Round 1 Cut Score
Language Arts	Level 2	473
	Level 3	520
	Level 4	641
Mathematics	Level 2	478
	Level 3	517
	Level 4	556

Grade 3 Round 2

Following Round 1 judgments, workshop facilitators provided a brief orientation to Round 2 ratings. Table facilitators then led a discussion at their tables regarding panelists' Round 1 Bookmark placements. Discussion was based on content and panelists' rationale for placing their Bookmarks as they did. Impact data were calculated, but not provided to panelists, following Round 1, and panelists were reminded that they did not need to reach consensus on Bookmark placement. Following the table discussion, panelists had the opportunity to reset their Bookmark placements for each level. As in Round 1, panelists entered their Bookmark placements on the paper form (Appendix B), and then in the online system via the Control Panel link for Round 2 activities.

Table 3.5 shows the median Grade 3 cut scores for each content area associated with Round 2 judgments. Detailed judgments may be found in Appendix D.

Table 3.5. Grade 3 Round 2 Cut Scores

Content Area	Level	Round 2 Cut Score
Language Arts	Level 2	476
	Level 3	531
	Level 4	641
Mathematics	Level 2	482
	Level 3	510
	Level 4	555

Grade 3 Round 3

Following Round 2 judgments, the two tables of panelists in each content area joined together for the remaining Grade 3 standard setting activities. Each content group was provided median Bookmarks for each table, the median Bookmark for the grade level reflecting the judgments of all panelists, and the impact data based on the median Bookmark for the grade level. Figure 3.4 shows the impact data, or the percent of students in each level, for Round 2 based on the median Grade 3 Bookmark for each content area.

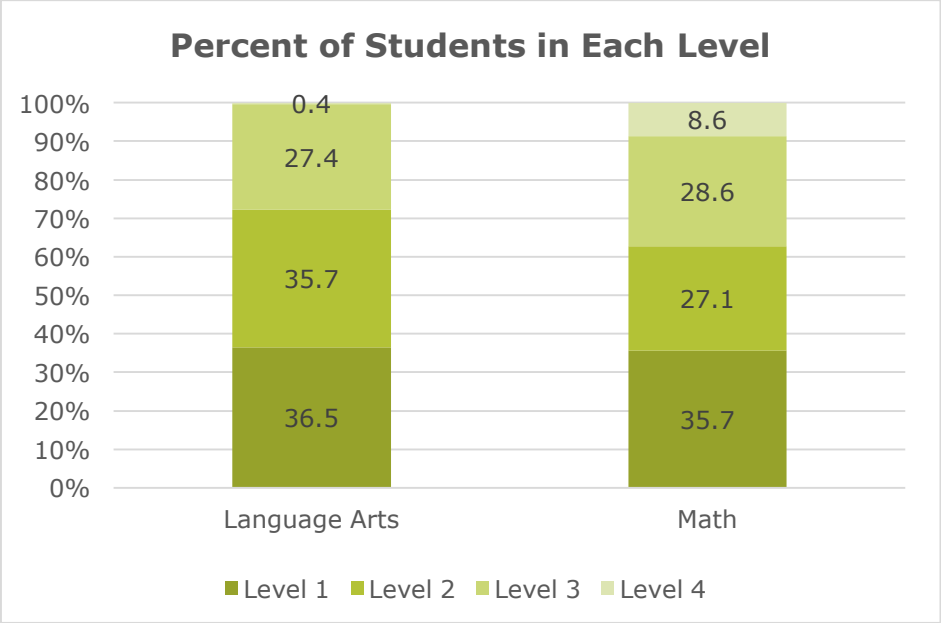
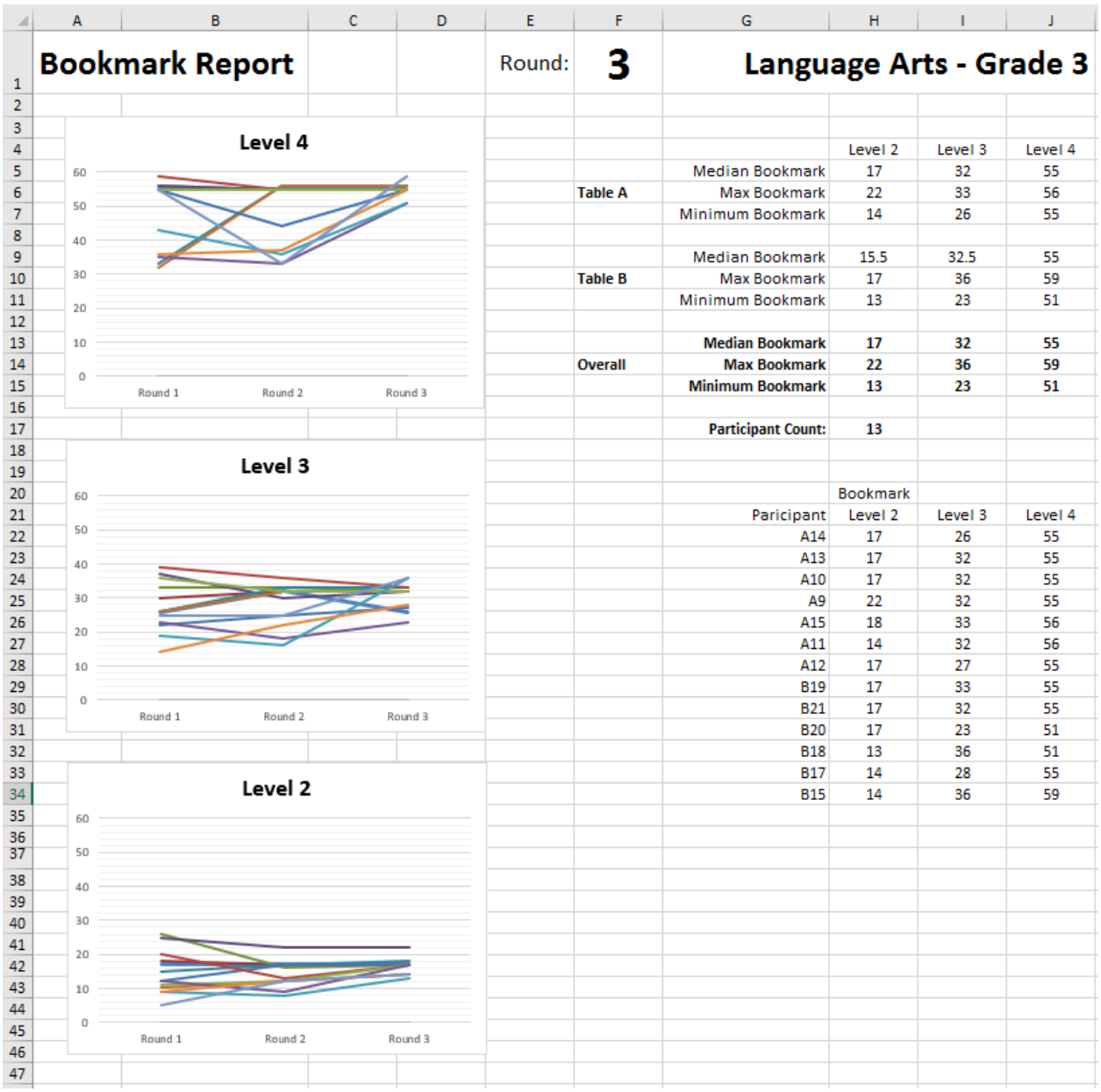


Figure 3.4. Grade 3 Round 2 Impact Data

Panelists within each content area worked together to discuss the differences between the tables' median Bookmarks for each level, beginning with Level 3, the anchor level, and to consider the impact data resulting from Round 2 judgments. Workshop facilitators were present in each content area room to answer any questions panelists and/or table facilitators might have. Panelists then entered their Round 3 ratings on their paper rating forms and in the online system using the link for Round 3 activities. Following completion of Round 3 judgments, table facilitators collected all Grade 3 secure materials. Workshop facilitators then reviewed Round 3 cut scores and impact data with panelists for each content area and presented a Bookmark Report to show detailed results of Round 3 for each content area, including the convergence of bookmark ratings between rounds. A sample Bookmark Report is shown Figure 3.5.



		Level 2	Level 3	Level 4
	Median Bookmark	17	32	55
Table A	Max Bookmark	22	33	56
	Minimum Bookmark	14	26	55
	Median Bookmark	15.5	32.5	55
Table B	Max Bookmark	17	36	59
	Minimum Bookmark	13	23	51
	Median Bookmark	**17**	**32**	**55**
Overall	**Max Bookmark**	**22**	**36**	**59**
	Minimum Bookmark	**13**	**23**	**51**
	Participant Count:	**13**		

	Bookmark		
Paricipant	Level 2	Level 3	Level 4
A14	17	26	55
A13	17	32	55
A10	17	32	55
A9	22	32	55
A15	18	33	56
A11	14	32	56
A12	17	27	55
B19	17	33	55
B21	17	32	55
B20	17	23	51
B18	13	36	51
B17	14	28	55
B15	14	36	59

Figure 3.5. Bookmark Report

Table 3.6 shows the median Round 3 cut scores for Grade 3 for each content area. Detailed judgments for Grade 3 Round 3 are included in Appendix D.

Figure 3.6 shows the final impact data for Grade 3 based on Round 3 median cut scores. Scores changed less than 4% from round 2 to round 3.

Table 3.6. Grade 3 Round 3 Cut Scores

Content Area	Level	Round 3 Cut Score	Cut Score Change from Round 2
Language Arts	Level 2	494	(+18)
	Level 3	531	No change
	Level 4	641	No change
Mathematics	Level 2	474	(-8)
	Level 3	513	(+3)
	Level 4	546	(-9)

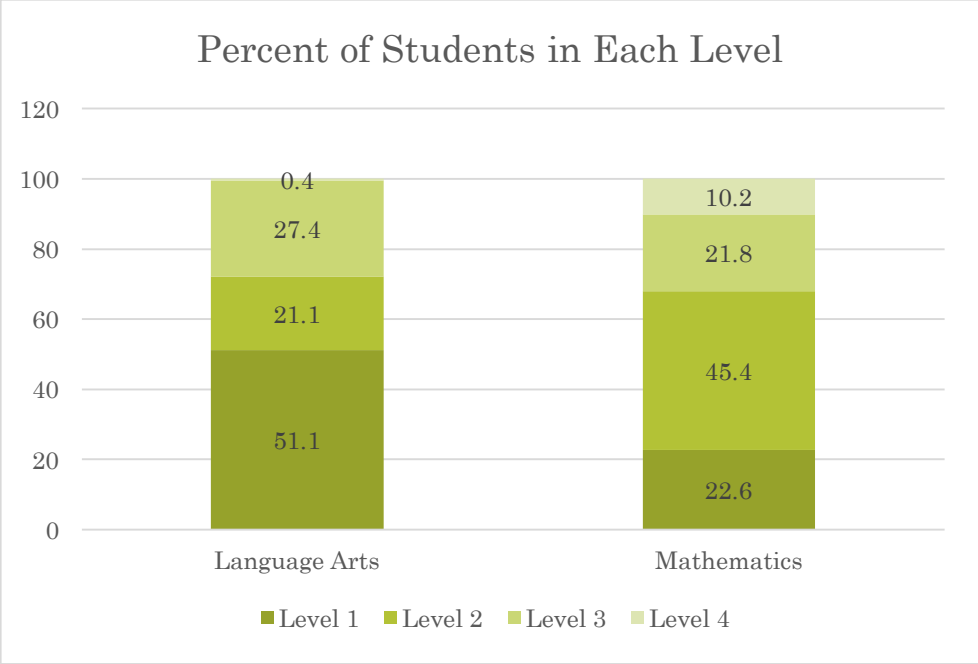


Figure 3.6. Grade 3 Round 3 Impact Data

Grade 4 Round 1

Following completion of all standard setting activities for Grade 3, table facilitators in each content area room distributed secure Grade 4 materials to panelists, using the same accounting procedures employed for distribution and collection of Grade 3 secure materials. Panelists in each content area began a review of the Grade 4 operational assessments and OIBs. Once again, panelists received paper copies of the OIB, stimulus booklet, and item map. All processes for accessing electronic copies of the materials were identical to those used for accessing Grade 3 materials. Table facilitators walked panelists through the process of logging into the Grade 4 operational assessment for the appropriate content area, and then facilitated the discussion of the Grade 4 OIB for the content area, again using the guiding questions, "What does this item measure? That is, what do you know about the knowledge and skills of a student who responds successfully to this item?" and "Why is this item more difficult than the preceding items." The scribe at each table recorded notes in the item map. Since Bookmark training had already been completed, panelists were able to move directly to Round 1 ratings on Grade 4 assessments for their content areas after completing study of the OIB using the same procedures used to place Bookmarks for the Grade 3 assessments. Once again, Round 1 ratings were completed independently, and without discussion.

Median cut scores for Grade 4 for Round 1 judgments are shown in Table 3.7. Detailed judgments for Grade 4 Round 1 are included in Appendix D.

Table 3.7. Grade 4 Round 1 Cut Scores

Content Area	Level	Round 1 Cut Score
Language Arts	Level 2	478
	Level 3	520
	Level 4	577
Mathematics	Level 2	488
	Level 3	542
	Level 4	563

Following Round 1 judgments for Grade 4 assessments, table facilitators concluded Day 2 activities by collecting secure materials and participating in the daily debrief with workshop facilitators.

STANDARD SETTING EVENT DAY 3

GRADE 4 ROUND 2

To begin Day 3 of the KĀʻEO standard setting event, table facilitators distributed Grade 4 secure materials to panelists and led a discussion of the Round 1 ratings, within their individual tables, focusing on content and panelists' rationale for Bookmark placement. Once again, impact data were calculated, but not presented to panelists following Round 1. Panelists were then given an opportunity to reset their Bookmarks in a second round of judgments. Round 2 median cut scores are shown in Table 3.8.

Table 3.8. Grade 4 Round 2 Cut Scores

Content Area	Level	Round 2 Cut Score
Language Arts	Level 2	499
	Level 3	525
	Level 4	577
Mathematics	Level 2	494
	Level 3	542
	Level 4	558

GRADE 4 ROUND 3

Following Round 2 judgments for Grade 4, workshop facilitators presented the impact data and Bookmark Reports from Round 2, and table facilitators led a cross-table discussion of the impact data and the differences in judgments between the two tables within the content area. Figure 3.7 shows the impact data, or the percent of students in each level, for Round 2 based on the median Grade 4 Bookmark for each content area.

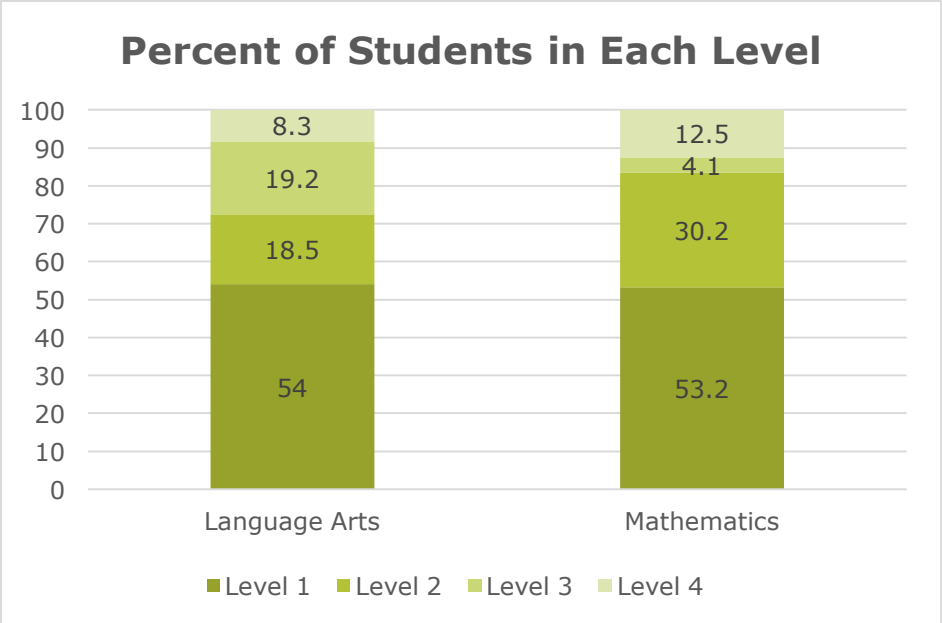


Figure 3.7. Grade 4 Round 2 Impact Data

Following review of the Bookmark Report and Round 2 impact data, with all panelists in the content area participating, panelists then entered their Round 3 judgments using both the paper rating forms and the online system. As a final step in the process, workshop facilitators shared the resulting median cut scores for the content area, along with the final impact data based on Round 3 judgments, with panelists.

Table 3.9 shows the Round 3 median cut scores for each content area for the Grade 4 assessments; detailed Round 3 judgments for Grade 4 included in Appendix D.

Figure 3.8 shows final impact data based on Round 3 judgments for Grade 4. Changes in cut scores from Round 2 to Round 3 were small (less than 3%).

Table 3.9. Grade 4 Round 3 Cut Scores

Content Area	Level	Round 2 Cut Score	Cut Score Change from Round 1
Language Arts	Level 2	499	No Change
	Level 3	520	(-5)
	Level 4	577	No change
Mathematics	Level 2	487	(-7)
	Level 3	535	(-7)
	Level 4	557	(-1)

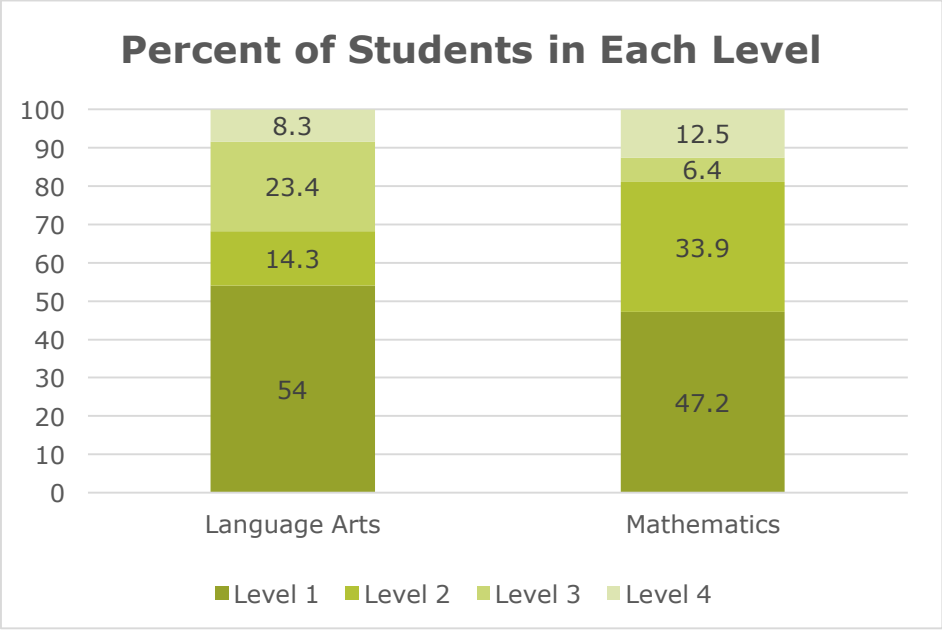


Figure 3.8. Grade 4 Round 3 Impact Data

CHAPTER 4. PARTICIPANT EVALUATION OF KĀʻEO STANDARD SETTING

Participant feedback was obtained in several ways during the KĀʻEO standard setting. First, an ongoing feedback loop allowed table leaders and panelists to communicate with workshop facilitators, UH staff, and HIDOE staff throughout the workshop. Second, all panelists completed standardized evaluations after the opening session of the standard setting event (Table 3.2), and prior to beginning Round 1 judgments of the BSSP to determine their readiness to participate in standard setting (Table 3.3). Finally, panelists completed evaluation forms following the completion of Grade 3 standard setting activities, and again following the completion of Grade 4 standard setting activities to provide feedback to workshop organizers and KĀʻEO developers about how well the standard setting process was implemented and how confident panelists felt in the result of their work. Panelists submitted their evaluations online, using links provided on their secure thumb drives. Each statement was followed by dropdown options to allow panelists to indicate their level of agreement with the statement. Figure 4.1 shows the online presentation of a portion of the evaluation.

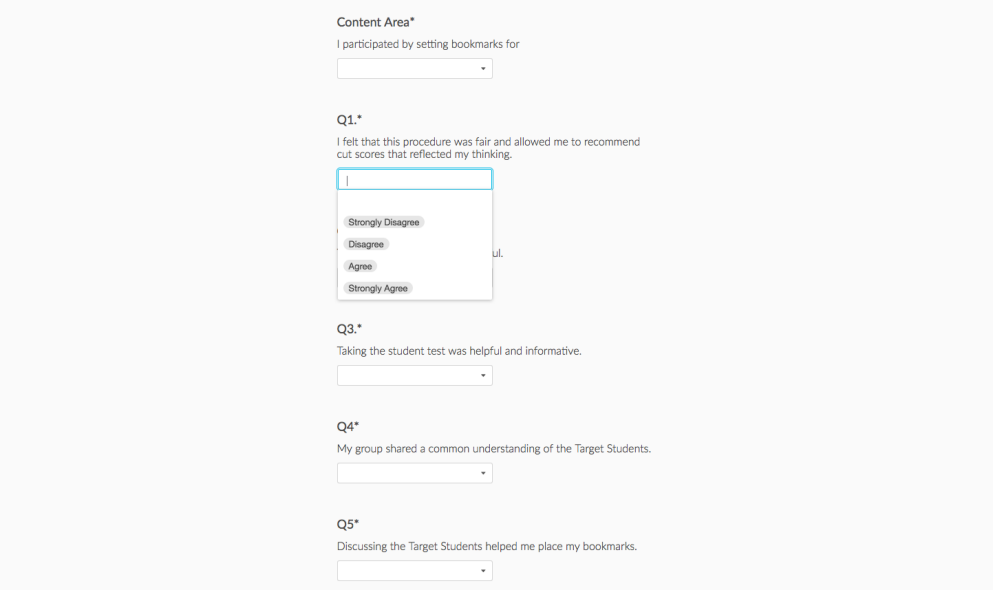


Figure 4.1. Panelist Evaluation Sample

Table 4.1 shows the complete list of statements to which panelists responded. Following the final round of Grade 3 judgments, panelists responded to Questions 1 through 18, providing information specific to Grade 3 activities. Following the final round of Grade 4 judgments, panelists again responded to Questions 1 through 18, providing information specific to Grade 4 activities. They also responded to Questions 19 through 26 in consideration of the standard setting workshop as a whole.

Table 4.1. Panelist Evaluation Questions

Grade/Content Specific Evaluation Questions	
Q1	I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking.
Q2	The training materials were helpful.
Q3	Taking the student test was helpful and informative.
Q4	My group shared a common understanding of the Target Students.
Q5	Discussing the Target Students helped me place my bookmarks.
Q6	During Round 1, I placed my bookmarks independently.
Q7	I considered the Kaiapuni Standards when I placed my bookmarks.
Q8	The policy definitions were clearly communicated.
Q9	I understood how to place my bookmarks.
Q10	I had enough time to consider my bookmark placement.
Q11	I feel the recommended standards that resulted from this process are reasonable.
Q12	The impact data helped me evaluate my group's final bookmarks.
Q13	I understood how to interpret the impact data.
Q14	The impact data influenced where I placed my final bookmarks.
Q14a	I would defend the panel's recommended Level 4 cut scores against criticism that they are too high.
Q14b	I would defend the panel's recommended Level 4 cut scores against criticism that they are too low.
Q15	I would defend the panel's recommended Level 3 cut scores against criticism that they are too high.
Q16	I would defend the panel's recommended Level 3 cut scores against criticism that they are too low.
Q17	I would defend the panel's recommended Level 2 cut scores against criticism that they are too high.
Q18	I would defend the panel's recommended Level 2 cut scores against criticism that they are too low.
Evaluation Questions Pertaining to Entire Workshop	
Q19	I feel that my grade group as a whole is credible.
Q20	Overall, I believe that my opinions were considered and valued by my group.
Q21	Overall, I valued the workshop as a professional development experience.
Q22	This experience will help me target instruction for the students in my classroom.
Q23	The food and service at the facility met my expectations.
Q24	The workspace had accommodations appropriate to facilitate our work.
Q25	Participating in the workshop increased my understanding of the Kaiapuni assessment.
Q26	The workshop was well organized.

In general, panelists' evaluations of the workshop were positive, with the vast majority of panelists (greater than 90 percent) either agreeing or strongly agreeing with most evaluation statements. Panelists unanimously indicated that the workshop was a valuable professional development experience that increased their overall understanding of the KĀʻEO and would positively impact their instructional practices. Specific evaluation components will be discussed with regard to Standard Setting Preparation and Training; Review of Content Standards, Target Student Discussion, and Policy Level Definitions; Bookmark Standard Setting Implementation; Confidence in Cut Scores; and Overall Impressions. Unless otherwise indicated, results are based on the responses of 13 Language Arts panelists and 12 Mathematics panelists. The tables that follow show the percent of panelists that indicated they "Strongly Agree" or "Agree" with the evaluation statement. Complete evaluation results for each grade level and content area, along with panelists' comments, are included in Appendix D.

STANDARD SETTING PREPARATION AND TRAINING

Table 4.2 shows the evaluation questions and panelist responses pertaining to Standard Setting Preparation and Training.

Table 4.2. Evaluation – Standard Setting Preparation and Training

	Language Arts		Mathematics	
	Grade 3	Grade 4	Grade 3	Grade 4
The training materials were helpful.	100.0	100.0	92.3	100.0
Taking the student test was helpful and informative.	100.0	100.0	100.0	100.0
I considered the Kaiapuni Standards when I placed my bookmarks.	92.3	100.0	100.0	100.0

With the exception of the Mathematics panelists following the Grade 3 activities, 100% of panelists indicated they either Agreed or Strongly Agreed with evaluation statements regarding training materials. Two major components of training – the BSSP training materials and review of the operational assessment – were addressed in evaluation. With the exception of Grade 3 LA, panelists rated the use of standards during BSSP as 100%. Based on responses to these statements, panelists overwhelmingly affirmed the effectiveness of the training process.

REVIEW OF CONTENT STANDARDS, TARGET STUDENT DISCUSSION, AND POLICY LEVEL DEFINITIONS

Table 4.3 shows panelists' responses to evaluation statements pertaining to content standards, the target student, and Policy Level Definitions.

Table 4.3. Evaluation – Content Standards, Target Student, and Policy Level Definitions

	Language Arts		Mathematics	
	Grade 3	Grade 4	Grade 3	Grade 4
My group shared a common understanding of the Target Students.	100.0	100.0	92.3	100.0
Discussing the Target Students helped me place my bookmarks.	100.0	100.0	92.3	91.7
I considered the Kaiapuni Standards when I placed my bookmarks.	92.3	100.0	100.0	100.0
The policy definitions were clearly communicated.	100.0	100.0	92.3	100.0

Panelists' overwhelmingly positive responses to evaluation items pertaining to the target student, content standards, and policy definitions indicate that workshop facilitators, UHM staff, and HIDOE were effective in communicating the underlying content framework for the standard setting process. Greater than 90 percent of panelists indicated an appropriate understanding of, and emphasis on, the knowledge, skills, and abilities that students should be expected to demonstrate using the Kaiapuni Standards as a reference.

BOOKMARK STANDARD SETTING IMPLEMENTATION

Table 4.4 shows panelists' responses to evaluation items pertaining to implementation of the BSSP.

Table 4.4. Evaluation – BSSP Implementation

	Language Arts		Mathematics	
	Grade 3	Grade 4	Grade 3	Grade 4
During Round 1, I placed my bookmarks independently.	100.0	100.0	100.0	100.0
The impact data helped me evaluate final bookmarks.	100.0	100.0	100.0	100.0
I understood how to interpret the impact data.	100.0	100.0	100.0	100.0
The impact data influenced where I placed my final bookmarks.	76.9	69.2	84.6	83.3

Panelists' responses to evaluation questions indicated that the BSSP was implemented with fidelity, beginning with panelists' independent placement of Bookmarks. Panelists unanimously acknowledged their understanding of how to interpret impact data and to consider impact data in their placement of final Bookmarks. The degree to which panelists indicated that impact influenced where they placed their final bookmarks ranged from 69.2 percent (Grade 3 Language Arts) to 84.6 percent (Grade 3 Mathematics). This indicates that the conversation among panelists was focused more on content and characteristics of the target student than on the anticipated distribution of students across the achievement levels based on recommended cut scores.

CONFIDENCE IN CUT SCORES

Table 4.5 shows panelists' responses to evaluation statements pertaining to their confidence in the recommended cut scores, as well as their belief in the credibility of the standard setting process.

Table 4.5. Evaluation – Confidence in Cut Scores

	Language Arts		Mathematics	
	Grade 3	Grade 4	Grade 3	Grade 4
I would defend the recommended Level 3 cut scores against criticism that they are too high.	84.6	92.3	100.0	100.0
I would defend the recommended Level 3 cut scores against criticism that they are too low.	69.2	92.3	100.0	100.0
I would defend the recommended Level 2 cut scores against criticism that they are too high.	69.2	84.6	100.0	100.0
I would defend the recommended Level 2 cut scores against criticism that they are too low.	76.9	100.0	100.0	100.0
I would defend the recommended Level 4 cut scores against criticism that they are too high.	83.3*	76.9	100.0	100.0
I would defend the recommended Level 4 cut scores against criticism that they are too low.	100.0*	92.3	100.0	100.0
I feel that my grade group as a whole is credible.**		100.0		100.0
Overall, I believe that my opinions were considered and valued by my group.**		100.0		100.0

* Based on the responses of 12 Language Arts panelists.

** Based on the responses of all panelists in both content areas.

Panelists in both content areas were in unanimous agreement (100 percent stating that they "Strongly Agreed" or "Agreed") that their groups as a whole were credible, acknowledging that the collective expertise of the panels was a valuable component of the standard setting process. Furthermore, they unanimously indicated that they felt their opinions were valued by their groups,

demonstrating that all panelists felt they contributed adequately to the standard setting process. Mathematics panelists were also unanimous in their agreement that they would defend cut scores at all levels against criticisms that they are too high or too low. Language Arts panelists were slightly less enthusiastic about their willingness to defend cut scores against criticisms that they are too high or too low, although the majority indicated that they would do so for all levels in both grade levels. For Grade 4, the percent of panelists that "Strongly Agreed" or "Agreed" that they would defend the cut scores was slightly higher, ranging from 76.9 percent of panelists indicating that they would defend Level 4 cut scores against criticisms that they are too high, to 100 percent of panelists indicating that they would defend Level 2 cut scores against criticisms that they are too low. For Grade 3, Language Arts panelists were less uniform, and slightly less positive, about their willingness to defend the cut scores. Panelists felt most strongly that they would defend the Level 4 cut scores against criticism that they are too high or too low. A particular point at which panelists appeared to have the least confidence in Grade 3 Language Arts cut scores emerged in the transition from Level 2 to Level 3, as 69.2 percent of panelists indicated that they would defend Level 3 cut scores against criticism that they are too low, and an equal percentage of panelists indicated that they would defend Level 2 cut scores against criticism that they are too high. Overall, however, evaluations indicated that panelists, in general, supported the final cut scores.

OVERALL IMPRESSIONS

Panelists' overall impressions of the standard setting event were overwhelmingly positive, as shown in Table 4.6, and reflected in panelists' comments on the evaluation.

Table 4.6. Evaluation – Overall Impressions

	Language Arts	Mathematics
Overall, I valued the workshop as a professional development experience.**	100.0	100.0
This experience will help me target instruction in my classroom.	100.0	100.0
The food and service at the facility met my expectations.**	100.0	100.0
The workspace had accommodations appropriate to facilitate our work.**	100.0	100.0
Participating in the workshop increased my understanding of the Kaiapuni assessment.**	100.0	100.0
The workshop was well organized.**	100.0	100.0

* Based on the responses of 12 Language Arts panelists.

** Based on the responses of all panelists in both content areas.

Overall, panelists felt the workshop was well organized and the facility was adequate for them to complete their work. More significantly, they indicated that their work was valuable to them professionally, and the experience would benefit them and their students in the future. Panelists shared the following statements regarding their overall perceptions of the KĀʻEO standard setting event:

- *"I now have a clear understanding of how the process works as far as the assessment creation and putting it into place. Also an understanding of how hard the OHE works in partnership with UHM and the kaiapuni teachers statewide."*
- *"It gave me a much better understanding of everything going on with the testing situation. Mahalo nui!"*
- *"Thank you for bringing us together! It was wonderful to hear and learn from other Hawaiian Immersion/Medium educators. It helps me measure how we as a group are performing."*
- *"Everything was well thought out and planned."*
- *"Learned a lot. I appreciated the opportunity to meet with fellow kumu from various schools."*
- *"Mahalo nui no kēia hālawai, ua aʻo au i nā mea he nui mai ia mau kumu kaiapuni ʻē aʻe. Mākaukau au e holomua a hoʻoikaika i ka ʻōlelo Hawaiʻi ma koʻu kula. I look forward to our next Standard Setting hālāwai!"*
- *"The workshop coordinators and presenters helped to facilitate the process very well. I truly appreciate the process we used and feel that I can leave the workshop with a greater understanding of the assessment, of the process, and of tasks ahead."*
- *"This workshop was an amazing experience and was at the perfect time too! I would love to do this again."*

CHAPTER 5. CONTRIBUTIONS TO THE VALIDITY ARGUMENT

Technically sound standard setting procedures are a critical piece in establishing the validity of an assessment. As such, the standard setting plan and methodology, the standard setting workshop itself, the recommended cut scores and corresponding impact data, and participants' evaluation responses must be considered together to create comprehensive evidence that the standard setting contributes to the overall validity argument for the assessment. The standard setting methodology must be well established psychometrically and well-suited to the characteristics of the assessment; the standard setting workshop must be carried out with fidelity to the plan; and qualified panelists must be confident that the cut scores they recommend are valid and defensible. Standard setting processes may be considered in terms of their adherence to generally agreed upon best practices, as well as their adherence to AERA/APA/NCME *Standards for Educational and Psychological Testing* (2014).

ADHERENCE OF THE KĀʻEO STANDARD SETTING TO BEST PRACTICES

As content-based standard setting has become common among large scale assessment programs, experts in the have begun to agree upon a core set of best practices (Hambleton & Pitoniak, 2006; Hambleton, Pitoniak, & Copella, 2012; Kane, 1994; Mehrens, 1995). Generally, best practices are considered in terms of internal criteria; external criteria; and procedural criteria, including panelists, method, and implementation.

INTERNAL CRITERIA

During a standard setting workshop, it is expected that agreement among panelists will increase; in other words, there should be increased agreement within the group. One way to examine evidence of convergence is to plot it across rounds. Figures 5.1 through 5.4 show the convergence plots for grades 3 and 4 HLA and mathematics. In general, there was greater agreement in Round 3 compared to Round 1 in all grade/content areas and for all cut scores, except for Level 2 in Grade 4 mathematics. Figure 5.4 shows there was little movement in panelist judgment between Round 1 and Round 3.

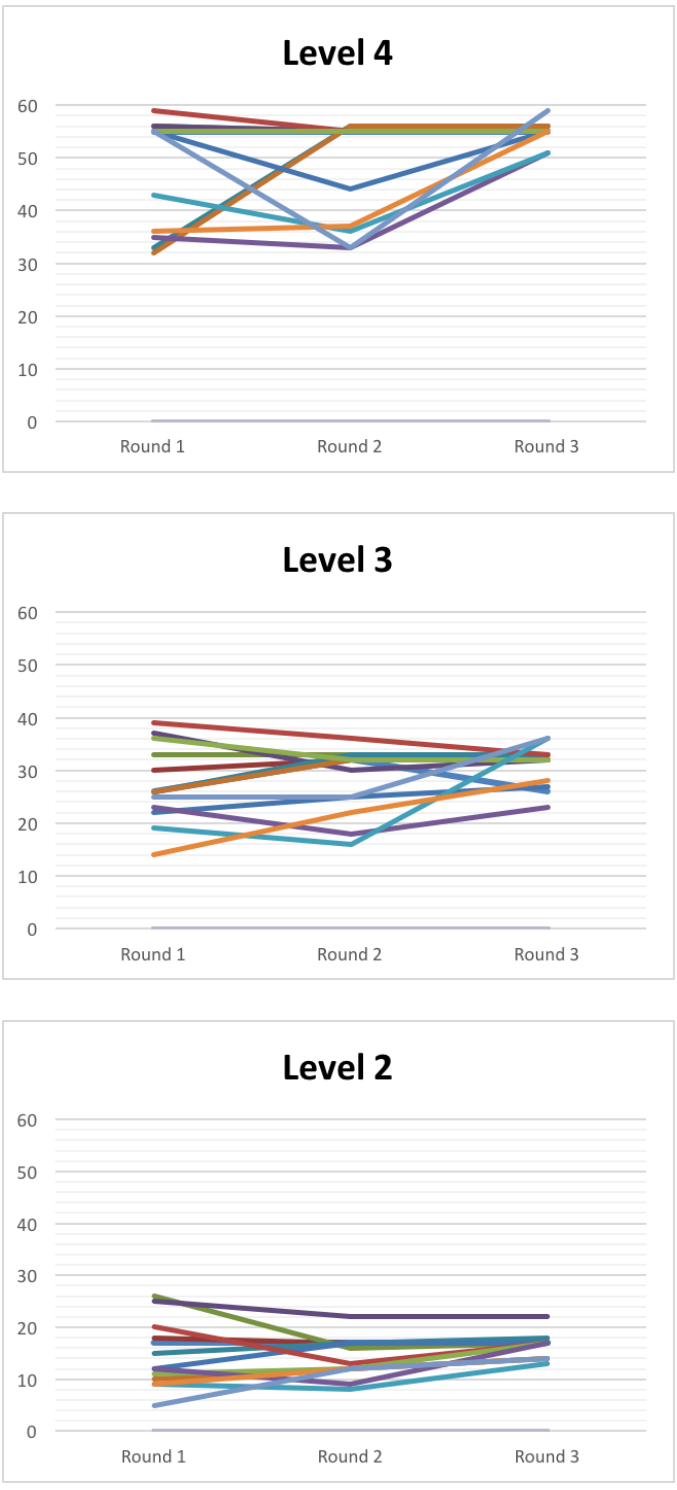


Figure 5.1. Internal convergence plots, HLA Grade 3

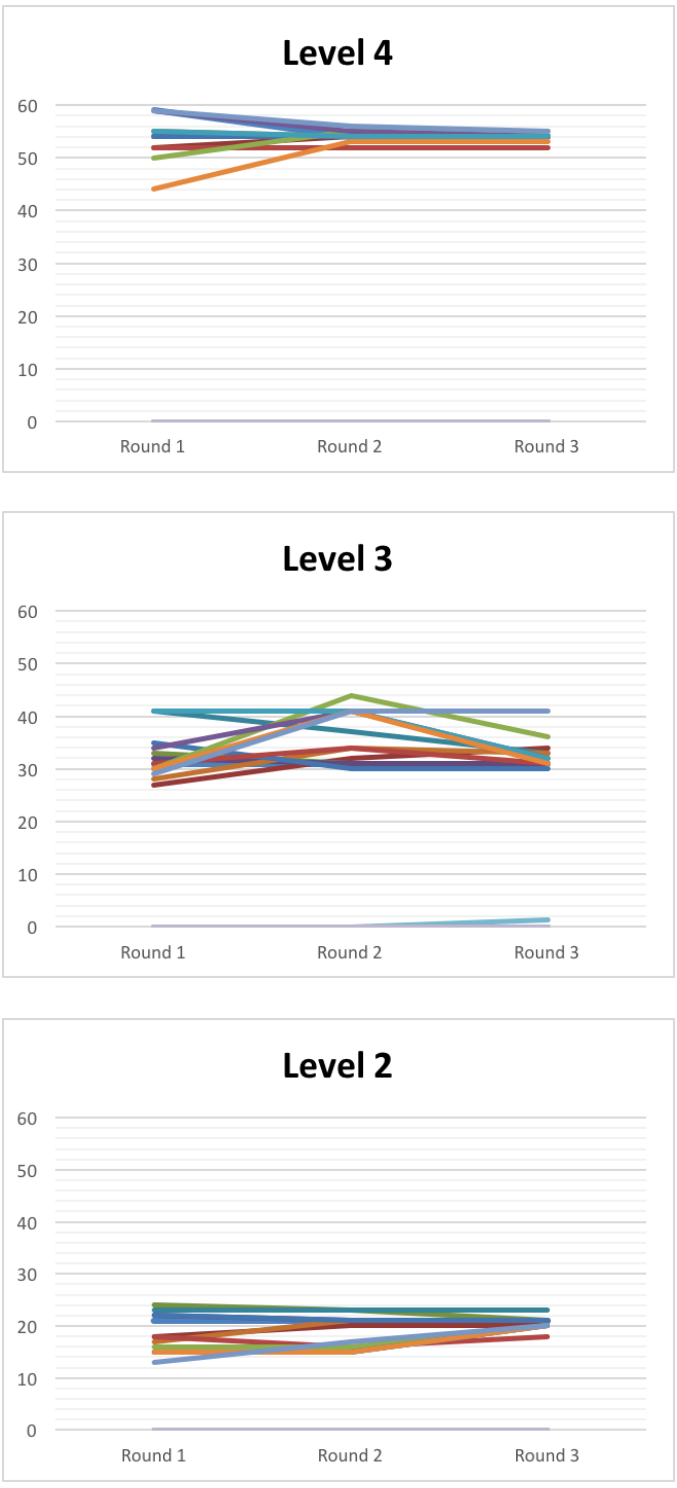
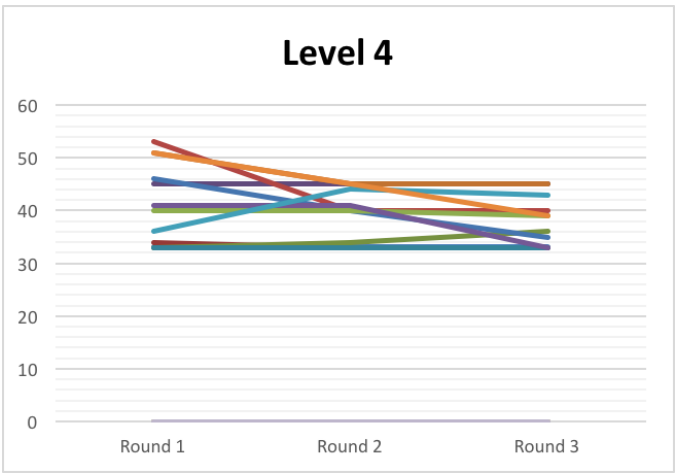
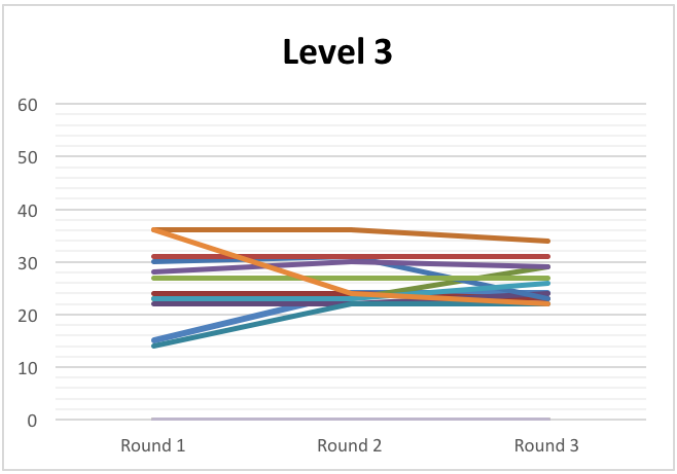


Figure 5.2. Internal covergence plots, HLA Grade 4





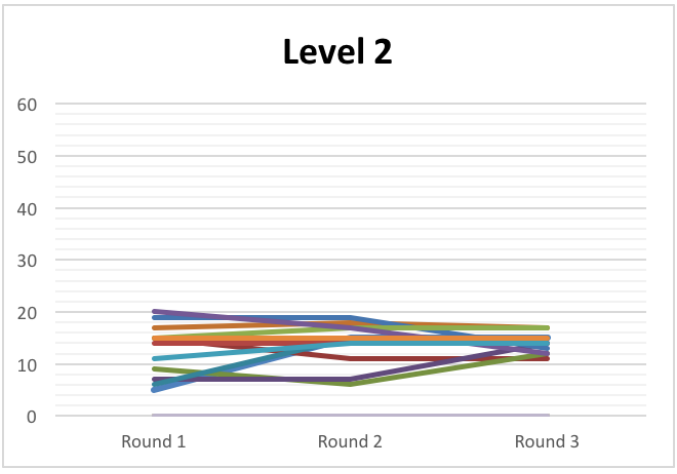
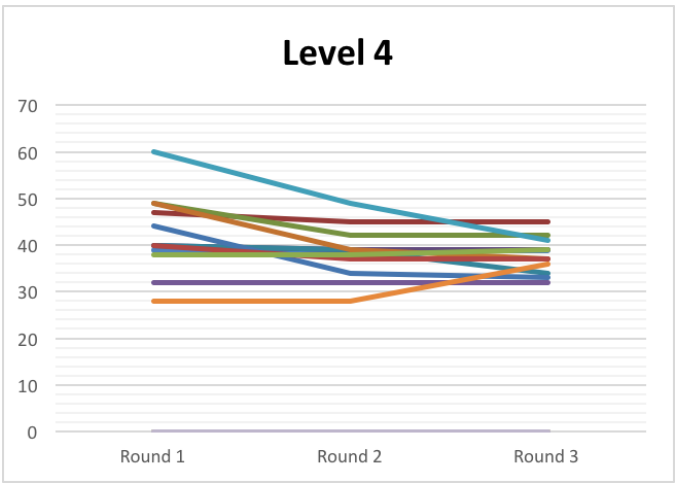
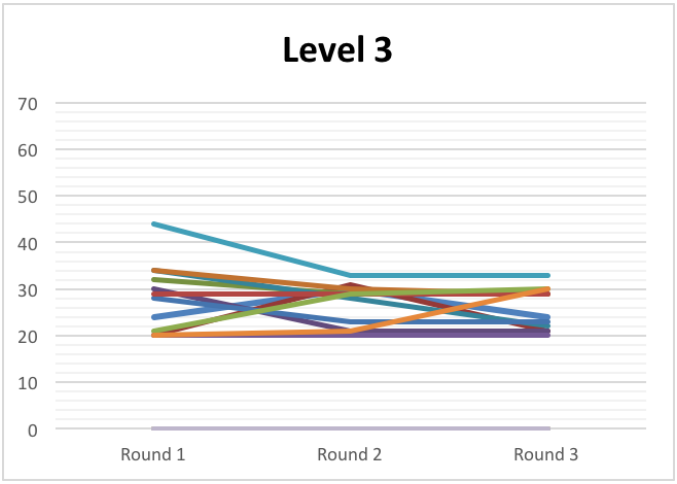


Figure 5.3. Internal convergence plots, Mathematics Grade 3





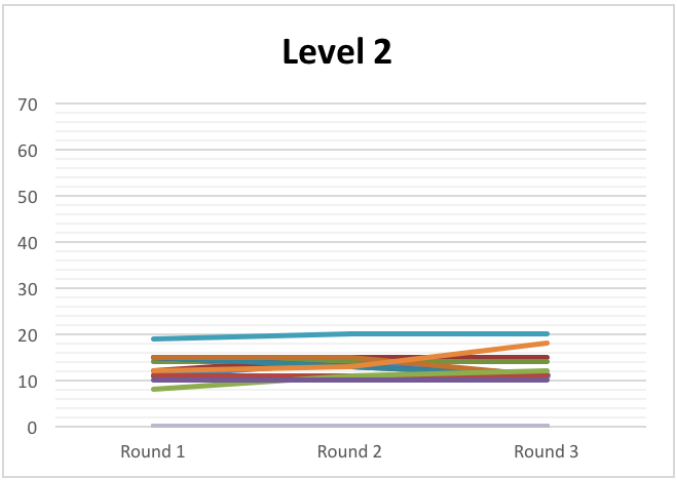


Figure 5.4. Internal convergence plots, Mathematics Grade 4

External Criteria

External criteria refers to the reasonableness of the performance levels. The panelists were asked if they would defend their cut scores against criticism that they were too high or too low (see Chapter 4). The majority of panelists agreed that they would defend the cut scores against criticism that they were too high (i.e., too stringent) or too low (i.e., too easy). Even so, this type of evidence is best collected outside of the standard setting workshop and is beyond the scope of this report.

Procedural Evidence: Panelists

There are several best practices related to panelists. The panel should be *representative* of the important demographic groups in the state, *suitable* to the task at hand, of sufficient size. In addition, multiple panels are often used as a check on generalizability.

Representativeness

Because standards are an expression of values, the most important contributors to their credibility are the number and nature of the panelists. The composition of the panel was described in Chapter 2. The UH staff recruited panelists from 12 of Hawaii's 14 language immersion schools, ensuring broad representation from the target audience.

Suitability

Suitable panelists understand the content being assessed as well as the students who are being tested. The panel for this standard setting consisted of very experienced educators. All worked in education, and all were classroom teachers. Additionally, the groups had panelists who worked with special education students and with LEP students. Overall, the group was qualified to recommend standards on the tests.

Size

In a large-scale assessment with high stakes, a large enough group of panelists are needed to ensure the incorporation of a variety of perspectives to produce reliable results. Raymond and Reid (2001) recommend the use of 15 panelists for recommending cut scores for operational tests. Because the assessment is only administered in two grade levels, and teachers often teach combined grade levels, it was appropriate to select a single panel to complete the standard setting for both grade levels within a content area. The Language Arts panel included 13 participants; the Mathematics panel included 12 participants. All panelists from the content area panel completed standard setting for each grade level, resulting in sufficient data points to ensure reliable results.

Multiple Panels

Multiple subpanels are often formed from the single panel in order to estimate the generalizability of the recommended cut scores. Hambleton, Pitoniak, and Coppella (2012) indicated it is highly

desirable, but optional, to use multiple panels. Within each content area, the panelists were split into two panels of 6 to 7 participants each.

Procedural Evidence: Standard Setting Method

The standard setting method is evaluated based on its appropriateness for the type of test administered and the understandability of the judgment task.

Appropriateness

The Kaipauni assessments consist of selected-response items and multi-point items (e.g., writing prompts). The Bookmark procedure was designed for use with assessments comprised of multiple item types, and, as such, it is appropriate for setting performance standards on the Kaiapuni assessments. The Bookmark method has been used in a majority of states for establishing cut scores on K-12 tests (Karantonis & Sireci, 2006).

Understandability of Judgment Task

The Bookmark method requires panelists to place Bookmarks in OIBs that separate the content needed to be, say, Proficient, from the content that is more than enough to just get into the Proficient category. The content in front of the Bookmark tells the story of what the Proficient student is able to do. The content that comes after the Bookmark is not expected of the borderline Proficient student. This concept works exactly like a regular Bookmark where a person places a Bookmark after the pages s/he has read. From the perspective of those asked to make judgments about cut scores, it presents a relatively simple task to panelists, and one with which, at a conceptual level, they are already familiar (Lewis, Mitzel, Mercado, & Schulz, 2012).

Panelists understood their rating tasks (see Table 3.2). In addition, all panelists indicated they were ready to make a rating (i.e., place a Bookmark) following the review of Bookmark training (see Table 3.3).

Implementation of the Bookmark Method

There are various aspects of implementation that must be considered when evaluating a standard setting. These include: (a) training, (b) using of PLDs, (c) taking the test, (d) using an iterative process, (e) providing opportunity for discussion, (f) and presenting impact data. In addition, the method should be efficient, allow transparency in the computation of cut scores, and provide time for evaluations.

Training

The standard setting process is not a familiar activity for panelists and training should be carefully prepared so that panelists are competent in completing the required tasks. Training should cover the following components (Raymond & Reid, 2001): (1) the overall process; (2) context for standard

setting within the process of test development, purpose of the test, and consequences of the test; (3) expectations for performance (the PLDs); and (4) the specifics of how to place a Bookmark.

As explained in Chapter 3, the overall process was introduced during the general training. Staff from the HI DOE explained the purpose of the test, provided context for the standard setting within the framework of the Kaiapuni testing program. Dr. Egan introduced the Bookmark process and provided training on the first tasks the panelists would complete. Almost all panelists indicated the opening session provided an adequate background on the Kaiapuni program, purpose of the meeting, and understood their role at the standard setting event (see Table 3.2).

Once panelists had studied their OIBs, Dr. Egan overviewed the purpose of the target students and the specifics of Bookmark placement. Panelists completed a practice exercise during this session. When panelists went to their breakout rooms, they engaged in further discussion about the target students. The majority of panelists indicated that discussions about the target students were helpful, (see Table 4.3) with all panelists indicating readiness to place their Bookmarks after the training (see Table 3.3).

Use of ALDs

The ALDs are used to guide the panelists when setting their cut scores. They allow the panelists to have a common frame of reference when recommending cut scores (see Egan, Schneider, and Ferrara, 2012). Throughout the process, the Dr. Egan and Dr. Englert reminded panelists to place Bookmarks based on the threshold ALDs (i.e., the target students). The majority of panelists indicated that the target students helped them place their bookmarks (see Table 4.3).

Taking the Test

Panelists should spend time taking the test. This allows them to experience the assessment in a similar manner to the students and understand the frame of mind of a student experiencing each item, rather than a knowledgeable practitioner with years of experience teaching the content. Panelists spent time going through the test. As indicated in Table 4.2, all panelists agreed that taking the test was helpful and informative.

Iterative Process

Panelists should provide ratings more than once. This allows the panelists to gain familiarity with the process and the expectations of the ALDs. During the Kaiapuni standard setting, panelists participated in three rounds of discussion and Bookmark placements.

Discussion

Discussions are used to increase consistency in the results and to provide panelists time to discuss and reflect on diverging viewpoints. Panelists participated in a small-group discussion in Round 2 and a large-group discussion in Round 3. Table 4.5 shows that panelists believed their opinions were considered and valued by their groups.

Impact Data

Impact data provide panelists with information on the consequences of their decisions. It allows panelists to see how their recommendations will play out in the real world. Impact data was presented after Round 2, and very few students were in Level 3 or Level 4. Table 4.4 shows that panelists used the impact data to evaluate the final bookmarks, and it shows the majority of panelists were not influenced in their Bookmark placements.

Efficiency of Implementation

In an efficient standard setting, the facilitators will be qualified, the materials will be useful to panelists when they are making their ratings, and the activities will be carried out in a timely fashion. Dr. Egan led the standard setting, and she has deep experience in this area. She has designed and led over 40 standard setting workshops. The standard setting was designed to occur over a 2.5-day period, with groups completing two grade levels within a content area. The workshop was completed within this timeframe.

Panelists entered their own data using the online tool, and results were computed within minutes of the final panelist entering their data.

Since both Dr. Egan & Dr. Englert would be implementing the process, a detailed agenda was created with step-by-step instructions which described how the implementation would proceed. This agenda promoted consistency for both the table leaders and the facilitators.

Transparency of Cut Scores

The means of computing cut scores from panelist data should be clearly described. Dr. Egan led the panelists through an hour-long training session on how to place a Bookmark. As part of this training, she described the process she would use to translate the panelists' recommended Bookmark into a scale score.

Evaluations

In accordance with best practices, panelists were provided opportunities to evaluate the process. The results of the evaluations are presented in detail in Chapter 4.

ADHERENCE OF THE KĀʻEO STANDARD SETTING TO AERA/APA/NCME STANDARDS

AERA/APA/NCME *Standards for Educational and Psychological Testing* (2014) include three standards that are of particular relevance to standard setting. The KĀʻEO standard setting plan and its implementation adhered to those standards.

> *Standard 5.21 – When proposed score interpretation involves one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly.*

The KĀʻEO project at UHM documented its standard setting plan and design in the scope of work that guided the standard setting event (Appendix A). The rationale for the BSSP methodology and processes involved were clearly explained to panelists during training. Each step completed prior to, during, and after the standard setting event is clearly and thoroughly documented in this report.

> *Standard 5.22 – When cut scores defining pass-fail or proficiency levels are based on direct judgments about the adequacy of item or test performances, the judgmental process should be designed so that the participants providing the judgments can bring their knowledge and experience to bear in a reasonable way.*

Panelists for the KĀʻEO standard setting were selected primarily based on their experience and expertise in working with students in Hawaiian Immersion Assessment Projects. They were uniquely qualified to provide relevant expertise to the standard setting process. Use of the BSSP allowed panelists to use their knowledge of the Kaiapuni standards and the Hawaiian language and culture to make reasonable and intuitive judgments about achievement levels.

> *Standard 5.23 – When feasible and appropriate, cut scores defining categories and distinct substantive interpretations should be informed by sound empirical data concerning the relation of test performance to the relevant criteria.*

Empirical data (impact data) based on the Spring 2016 operational administration of the assessments was presented to panelists following their Round 2 judgments and again after their Round 3 judgments.

REFERENCES

American Educational Research Association, American Psychological Association, & American Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.

Egan, K. L., Schneider, M. C., & Ferrara, S. (2012). Performance level descriptors: History, practice, and a proposed framework. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 79-106). New York: Routledge.

Hambleton, R. K., & Pitoniak, M. (2006). *Setting performance standards*. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 433-470). Westport, CT: Praeger.

Hambleton, R. K., Pitoniak, M.J. & Copella (2012). Essential steps in setting performance standards on educational tests and strategies for assessing the reliability of results. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 47 – 76). New York: Routledge.

Kane, M. T. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research, 64*(3), 425–461.

Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, E. M. (2012). The Bookmark Standard Setting Procedure. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 225–253). New York, NY: Routledge.

Mehrens, W. (1995). Methodological issues in standard setting for educational exams. In the Proceedings of the Joint Conference on standard setting for Large-Scale Assessment, (Vol. 2, pp. 221-263). Washington, D.C.: National Assessment Governing Board (NAGB) and National Center for Education Statistics (NCES).

Sireci, S. (2010). Chapter 8: Validity Issues and Empirical Research on Translating Educational Achievement Tests. In P. C. Winter (Ed.), *Evaluating the comparability of scores from achievement test variations.* (pp. 153-183). Washington, DC: CCSSO.

APPENDIX A. STANDARD SETTING PLAN

STANDARD SETTING PROPOSAL FOR THE ELA AND MATHEMATICS ASSESSMENTS IN NATIVE HAWAIIAN

The University of Hawai'i (UH) is developing the Hawaiian Immersion Assessment Project (HIAP). In this project, Native Hawaiian assessments are being developed for Grade 3 and 4 Language Arts and Mathematics. These tests were field tested in Spring 2015 and will be operational in Spring 2016. As part of the test development cycle, UH will hold a standard setting workshop where educators will make recommendations for the cut scores necessary to be in Level 1, 2, 3, or 4 on the HIAP.

In creating the NH assessments, UH has, to the degree possible, followed the test development methods and processes used by the Smarter Balanced Assessment Consortium. For the standard setting workshop, UH wishes to implement a method similar to the one used by Smarter Balanced.

Smarter Balanced implemented a multi-phase standard setting process, in which they (1) implemented an online standard setting where thousands of teachers and other interested parties recommended the Level 3 cut score; (2) held an in-person standard setting workshop where teams of educators used the Bookmark standard setting procedure to recommend Levels 2, 3, and 4 cut scores; (3) invited a subset of educators from the in-person panels to ensure that the achievement levels were appropriately aligned across grades. Smarter Balanced set their cut scores following the field test implementation of their assessments.

It is possible, but unnecessary, to implement the same process for the HIAP. Instead, it is recommended that UH implement a cut-score validation that builds on the standard setting work completed by Smarter Balanced. The proposed cut score validation would start with the Smarter Balanced achievement levels and cut scores. In addition, it will utilize the Bookmark standard setting method.

The cut score validation would then be implemented following the operational administration of the assessment. This is a departure from the process used by Smarter Balanced, who implemented their ALS following the field test administration.

DESIGN 1: CUT SCORE VALIDATION

A cut score validation asks panelists to confirm existing cut scores in the context of an assessment. This is in stark contrast to a typical standard setting workshop where panelists are allowed to recommend cut scores without constraints. Because UH desires that the HIAP achievement levels are linked to the Smarter Balanced achievement levels, a process that allows the panelists to see the relationship of the Smarter Balanced cut scores and achievement levels to the HIAP is proposed.

Bookmark Standard Setting Procedure

The Bookmark Standard Setting Procedure will be used to validate Smarter Balanced cut scores on the HIAP. The Bookmark Procedure engages panelists in discussions with their peers. These teams of experts will discuss the validity of the Smarter Balanced cut score in the context of the HIAP using the information from four sets of materials: ordered item booklets, item maps, achievement level descriptions, and preliminary cut scores.

Ordered Item Booklets

The ordered item book (OIB) is comprised of the items from the operational form of the HIAP. These items are ordered in terms of difficulty. The ordering is straightforward, with easier items placed earlier in the booklet and harder items following.

The test data used to create an OIB will be based on the responses of Hawaiian students who were administered the HIAP. To order the items by difficulty for the OIB, the items will be located on the Rasch scale location where a student has a two-thirds likelihood of answering each given item correctly.

Item Maps

The item maps summarize the materials in the OIB. The item map specifies the order of difficulty, the scale location, the item number on the operational test, the scoring key, and the content standard that the item measures. Panelist discussions are guided by two questions found on the item maps:

- What does this item measure? That is, what do you know about a student who can respond successfully to this item (or score point)?
- Why is this item more difficult that the items preceding it?

In responding to these questions, standard setting panelists gain a thorough understanding of the knowledge, skills, and processes needed to respond correctly to the items on the test. Panelists will use this knowledge to better inform their recommendations about whether to adjust or to keep the Smarter Balanced cut scores.

Achievement Level Descriptors

Achievement level descriptors (ALDs) summarize the knowledge, skills, and abilities of students in each achievement level. ALDs represent important policy-based decisions and are an important part of the overall system of performance standards in any testing program.

Smarter Balanced developed three types of ALDs prior to the standard setting. Dr. Egan designed and led the process used by Smarter Balanced to design these ALDs. Using Dr. Egan's framework, Smarter Balanced developed three types of ALDs:

- Policy ALDs that articulate the policymakers' vision of the goals and rigor for the final performance standards.
- Range ALDs are grade/content specific descriptors that may be used by item writers and describe the cognitive and content rigor that is encompassed within particular achievement levels.
- Threshold ALDs are used by standard setting panelists and are a subset of the range ALDs. These ALDs represent the minimal knowledge, skills, and abilities that a student needs in order to enter a particular achievement level.

Preliminary Cut Scores

Equipercentile linking will be used to link the HIAP Scale to the Smarter Balanced Scale. This will allow us to determine the HIAP scale score that is equivalent to each Smarter Balanced cut score. Since all Smarter Balanced data may not be available or accessible, this linking may need to be done using only the results of Hawaiian students on the Smarter Balanced assessment.

The UH should decide prior to the workshop whether or not panelists will be able to adjust the cut scores. The UH may decide that the panelists may adjust the cut scores within a restricted range. Alternatively, the UH may decide that panelists are free to adjust the cut score however they see fit.

Workshop Staff

The workshop will be designed and led by Dr. Karla Egan. Dr. Egan will serve as the lead facilitator. She will provide orientation for panelists as well as all training needed during the workshop.

Facilitators

Dr. Kerry Englert will co-facilitate the workshop with Dr. Egan. Together with Dr. Egan, she will help manage the major portions of the standard setting workshop, including: security, data management, and time management. They will communicate standard setting results to UH.

Content Specialists

Dr. Egan recommends that content specialists be available during the workshop in order to answer content-specific questions that will arise during the test.

Materials Production and Data Entry

EdMetric LLC will oversee materials production, organize and distribute workshop materials, and provide operational support during the workshop itself.

Workshop Participants

There are two levels of participation that will occur at the standard setting workshop: small-group leaders and panelists.

Small Group Leaders

Small Group Leaders will be assigned to groups of panelists within each grade/content area. Small Group Leaders facilitate discussion and keep the process on track within their small groups. Small Group Leaders are full participants, and it is recommended that they be educators of notable status. They will be identified by UH from among the participants. Their primary role is to monitor the group discourse, keep the group focused on the task, and watch the clock for the group. Often they will have to moderate discussion, find a diplomatic middle ground between participants, or call for assistance. Small Group Leaders need appropriate skills for group facilitation and should be very familiar with the content measured by the test, as well as the population tested.

Panelists

Six panelists should be invited for each content area (see Table 1). The number of panelists may be increased, depending on the number of Native Hawaiian teachers available. The numbers in Table 1 recognize that there is a limited pool of candidates with the requisite knowledge to participate in the standard setting. Panelists should be experienced educators who reflect the diverse backgrounds and needs of Hawaiian students. The final committee for each content area should represent a sample of expert panelists from a pool of all such qualified experts. These educators should have experience in language immersion programs. Table 1 shows the number of panelists that should be recruited for each grade and content area.

Table 1: Proposed Number of Panelists, Design 1

Grade/Content Area	Number of Panelists
Grade 3/4 Language Arts	6
Grade 3/4 Mathematics	6
Total	12

Workshop Overview

In the proposed modification of the Bookmark Procedure to be used for HIAP, participants will be trained at the beginning of the standard setting and participate in two rounds of discussion and decision-making. Table 2 shows a high-level overview of a proposed agenda.

Table 2: High-level Agenda for Cut Score Validation

	Standard Setting Task
Day 1 Morning	Orientation and Training
Day 1 Afternoon	Grade 3, Threshold Discussions & Round 1
Day 2 Morning	Grade 3, Round 2
Day 2 Afternoon	Grade 4, Threshold Discussions & Round 1
Day 3 Morning	Grade 4, Round 2
Day 3 Afternoon	Reporting PLDs (if desired)

Threshold Student Discussions

At the beginning of the procedure, the panelists from both grade groups will discuss the Smarter Balanced Threshold ALDs for their content area. The panelists will be instructed that these ALDs were used at the Smarter Balanced standard setting, and they should provide direction to panelists if they decide to adjust the Smarter Balanced cut scores. Following this discussion, participants break into their table-based groups and begin Round 1 of the Bookmark Procedure.

Round 1

To begin Round 1, participants take the operational form of the HIAP and study their OIBs. Within their small groups, participants discuss what each item measures and why each item is more difficult than the preceding items in the booklet. The panelists will be shown the placement of the Smarter Balanced cut scores within the HIAP OIBs. They will also be advised of the impact data given the Smarter Balanced cut scores. Impact data are the projected percentage of students in each achievement level. Once participants have studied the OIB completes, they will be asked to make their first recommendation on whether or not to adjust the Smarter Balanced cut scores.

Round 2

During Round 2, each small group will be shown their group median bookmark placements as well as impact data based on the current group median bookmark placements. The group will discuss the items for which there was not consensus according to their Round 1 judgments. For a given achievement level, these are the

items in the OIB between the first and last of the bookmarks placed by the participants at each table. Following discussion, each participant will independently make recommendations on adjusting the Smarter Balanced cut scores. The final cut scores are established by finding the median of the Round 2 results.

Reporting ALDs

If desired by UH and DOE, the panelists can engage in a process to refine the Target ALDs to Reporting ALDs. These descriptors explain the knowledge, skills, and abilities of the students in each achievement level. Participants will use the information gathered from their study of the test, content standards, and understanding of the target student to add clarity and conciseness to the Reporting ALDs.

Workshop Evaluation

At the conclusion of the workshop, participants will complete an evaluation of the standard setting. As part of this evaluation, participants will indicate how satisfied they were with the workshop and with the recommended performance standards.

Workshop Logistics

Non-disclosure Agreements

UH should create non-disclosure agreements that panelists and workshop staff will sign in order to participate in the workshop.

Meeting Rooms

UH will need to arrange meeting rooms for the workshop itself. At a minimum, four rooms will be needed for the workshop. Room 1 will serve as the operations room. This is where workshop staff will store materials, enter data, and create materials during the workshop.

Room 2 will be used for the workshop itself. This room should accommodate two round tables of 6 panelists. The rooms should be large enough that each small group can hold discussions without interrupting the other group. If more than 6 panelists can be recruited, then an additional room will be needed as the content areas should be split into two breakout rooms.

Room 2 should have a projection screen and an LCD projector so that results can be shared with panelists. All training and sessions will be conducted in this room.

If desired, an additional room should be available where the panelists can eat lunch (if provided).

Hotel Rooms

UH will need to secure a block of hotel rooms for panelists who cannot go home each night.

Design 2: Achievement-Level Setting

If UH prefers, Dr. Egan can design a standard setting workshop where panelists set their own cut scores instead of adjusting the Smarter Balanced cut scores. The workshop would use the Bookmark standard setting procedure. This section lists key differences from Design 1.

Workshop Materials

The OIBs and item maps would be identical to those used in Design 1. We recommend that the Smarter Balanced Target ALDs be used in order to establish a link between the HIAP and the Smarter Balanced assessments.

Workshop Participants

Using Design 2, 12 panelists should be invited for each grade/content area (see Table 3).

Table 3: Proposed Number of Panelists, Design 2

Grade/Content Area	Number of Panelists
Grade 3/4 Language Arts	12
Grade 3/4 Mathematics	12
Total	24

Workshop Overview

In this design, the Bookmark Procedure is also used. Participants will be trained at the beginning of the standard setting and participate in three rounds of discussion and decision-making. Table 4 shows a high-level overview of a proposed agenda.

Table 4: High-level Agenda for Standard Setting Workshop

	Standard Setting Task
Day 1 Morning	Orientation and Training
Day 1 Afternoon	Grade 3, Threshold Discussions & Round 1
Day 2 Morning	Grade 3, Round 2
Day 2 Afternoon	Grade 3, Round 3 Grade 4, Begin Threshold Discussions & Round 1
Day 3 Morning	Grade 4, Complete Round 1 Grade 4, Round 2
Day 3 Afternoon	Grade 4, Round 3 Cross-grade alignment/Reporting PLDs

Threshold Student Discussions

The procedure is the same as Design 1.

Round 1

To begin Round 1, participants take the operational form of the HIAP and study their OIBs. Within their small groups, participants discuss what each item measures and why each item is more difficult than the preceding items in the booklet. Once study of the OIB completes, participants will be asked to make their first recommendation on where to set their cut scores. Unlike Design 1, the panelists will not be shown the Smarter Balanced cut scores.

Round 2

During Round 2, each small group will be shown their median bookmark placements as well as the median bookmark placement for the large group. The group will discuss the items for which there was not consensus according to their Round 1 judgments. Following discussion, each participant will independently make recommendations on where to place cut scores.

Round 3

In Round 3, the large group will be presented their median bookmark placements as well as impact data based on the median bookmark placements. The group will discuss the reasonableness of the impact data and the items for which there was not consensus among the small groups. Following discussion, each participant

independently makes his or her Round 3 judgment. The final cut scores are established by finding the median of the Round 3 results.

Cross-Grade Alignment

Following the conclusion of Round 3, the Table Leaders will convene to study alignment of cut scores across grades. They will study the cross-grade impact data to ensure consistency of results across both Grades 3 and 4. This group may recommend changes to the cut scores in order to bring better alignment to the impact data.

Reporting ALDs

While the Table Leaders participate in the discussion of cross-grade alignment, the panelists can engage in a process to refine the Target ALDs to Reporting ALDs. These descriptors explain the knowledge, skills, and abilities of the students in each achievement level. Participants will use the information gathered from their study of the test, content standards, and understanding of the target student to add clarity and conciseness to the Reporting ALDs.

Workshop Logistics

Using this design, four meeting rooms will need to be secured.

- Room 1: Operations Room
- Room 2: Breakout room, mathematics
- Room 3: Breakout room, language arts
- Room 4: Large room for orientation, training, and lunch (if provided)

Standard Setting Documentation

For either design that UH chooses, Dr. Egan will provide UH with a comprehensive design document that describes the method by which the key materials of the workshop will be created. The design document will also include a detailed agenda of the workshop. If desired by UH and DOE, Dr. Egan is happy to present this design document to the technical advisory committee (TAC). Dr. Egan recognizes that the TAC must approve all standard setting procedures used in conjunction with the HIAP program.

After the standard setting, Dr. Egan will document the standard setting process in a comprehensive technical report. The report is designed to assist UH in evaluating the performance standards recommended by standard setting participants, and to promote clear understanding of the process by stakeholders.

Final Standard Setting Technical Report

Dr. Egan will provide UH with a Final Standard Setting Technical Report detailing the process and results of the standard setting. This Report will contain detailed information about judgments made by participants in each grade and content area combination; information about standard errors of measurement and of the cut score; graphical representations of participants' judgments; detailed summaries of participants' evaluations; and copies of the handouts and overheads used during the standard setting workshop. This technical report will be created to promote ease of understanding by stakeholders, including a narrative description of the events of the standard setting.

Workshop Security

In this section, we detail suggested procedures for maintaining the security of the tests. First, UH should require participants to sign a nondisclosure agreement. This agreement should specify that participants will not remove any secure materials and will not disclose the content of test items after the workshop.

Second, all secure standard setting materials (i.e., the test items) will be printed on colored paper. This creates a visual cue for panelists that the items are secure and should not leave the breakout room. These materials are sequentially numbered and assigned to participants and staff by name. Participants are continually reminded that test security is needed to ensure test validity.

Third, secure materials are not permitted outside the breakout rooms where panelists confer. After each day, Small Group Leaders will follow an auditing procedure in order to account for all secure materials.

Fourth, when the workshop is not in session, all materials will be stored in a centralized room where access is limited to workshop staff.

Finally, all materials are inventoried at the conclusion of the workshop. Any missing documents can be tracked to the participant or staff member who used them. We suggest that all materials are securely destroyed using a local vendor.

APPENDIX B. WORKSHOP PRESENTATIONS AND HANDOUTS

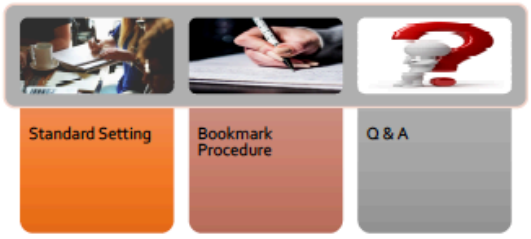
Kaiapuni Assessment of Educational Outcomes (KĀ'EO) Standard Setting

Training

Housekeeping

- Cellphones
- Laptops/devices
- Color-coded secure materials
- Non-disclosure agreement
- Questions about logistics

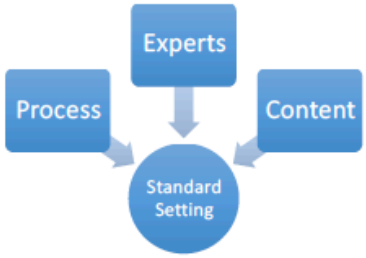
Overview of Table Facilitator Training



Standard Setting | Bookmark Procedure | Q & A

Understanding the Standard Setting Process

What is Standard Setting?



How do we do this?

You provide:

- Knowledge of student performance
- Knowledge of content

We provide:

- Content Standards
- Policy level descriptors
- Target students descriptors
- Process for setting cut scores

Content Standards

- Describe expected performance in HLA and mathematics

Policy Level Descriptors

Level 1	Demonstrates minimal understanding of and ability to apply the knowledge and skills associated with college content-readiness.
Level 2	Demonstrates partial understanding of and ability to apply the knowledge and skills associated with college content-readiness.
Level 3*	Demonstrates adequate understanding of and ability to apply the knowledge and skills associated with college content-readiness.
Level 4	Demonstrates thorough understanding of and ability to apply the knowledge and skills associated with college content-readiness.

* Level 3 is used for federal NCLB purposes.

KĀʻEO Cut Scores

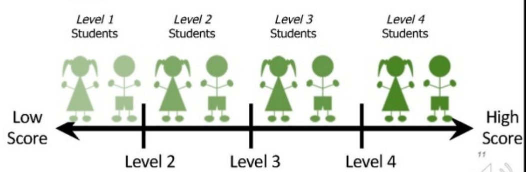
- Three cut scores
  - Level 2, Level 3, Level 4
- Four achievement levels
  - Level 1, Level 2, Level 3, Level 4

How do we set cut scores?

- Pe
  - trary
  - Test cific
  - Conten t considered
- Content
  - Use pre-established content standards
  - Considers educational objectives
- Bookmark Standard Setting Procedure

Purpose of the Standard Setting

- Allows cut scores to be set on the test scale
  - Test scale represents range of student scores
  - Cut scores will separate students into achievement levels



Purpose of the Standard Setting

Establish Three Cut Scores

Recommended by experts	On the test scale	Decisions based on content

Students placed in Achievement Levels

Students who meet or exceed the cut score have enough knowledge, skills, and abilities to be classified as Level 3

What is Bookmark Standard Setting?

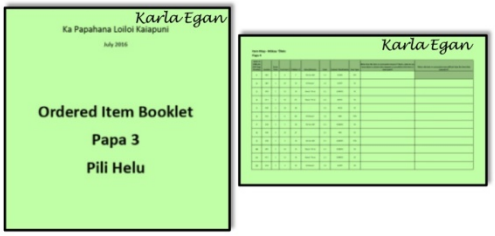
- Bookmark Standard Setting sets up a partnership between SMEs (you) and testing experts (us)
  - SMEs indicate the KSPs students should have to be in each achievement level
  - We (testing experts) use standardized methods, in this case, the Bookmark method, to identify cut scores on the test scale such that students that meet or exceed those cut scores tend to demonstrate those KSPs and students below those cut scores do not.

Developing Expertise

- You know students and content
- Activities to enhance knowledge:
  1. Study content standards
  2. Study test items
     - Take test
     - Study ordered item booklets using item maps
  3. Study target student descriptors

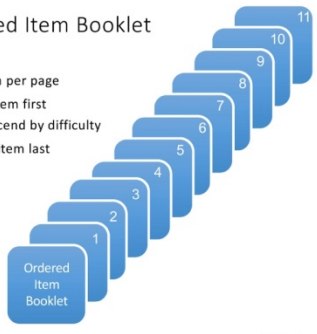
Bookmark Standard Setting Process

Bookmark Materials



Ordered Item Booklet

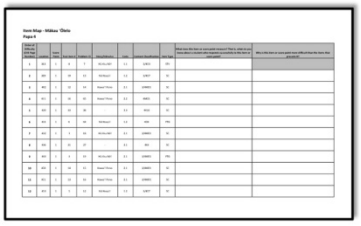
- One item per page
- Easiest item first
- Items ascend by difficulty
- Hardest item last



Item Map: Key Questions

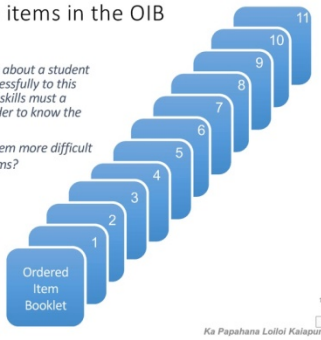
(1) What does this item measure?

(2) Why is this item more difficult than the items that precede it?



Studying the items in the OIB

1. *What do you know about a student who responds successfully to this item; that is, what skills must a student have in order to know the correct answer?*
2. *What makes this item more difficult than preceding items?*



Order of Difficulty (OIB Page Number)	Location	Score Point	Item ID	Content Classification	Item Type	What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?	Why is this item or score point more difficult than the items that precede it?
1	160	1	2013-4M7 #1 M15	Measurement	MC		
2	236	1	2013-4M7 #2 M17	Data analysis, Statistics, & Probal	MC		
3	251	1 of 2	2013-4Mt #14 M13	Data analysis, Statistics, & Probal	CR		
4	257	1	2013-4M7 #8 M15	Measurement	MC		
5	260	1	[illegible]	Algebra	MC		
6	282	2of 2	2013-4Mt #14 M13	Data analysis, Statistics, & Probal	CR		
7	295	1	2013-4M7 #12 M16	Number properties and operations	MC		

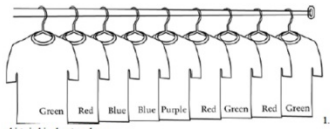
1. *What do you know about a student who responds successfully to this item; that is, what skills must a student have in order to know the correct answer?*
2. *What makes this item more difficult than preceding items?*

1

1. How much do these apples weigh?

A. 2 cups
B. 2 feet
C. 2 pounds
D. 2 quarts

2



1. Mark has nine shirts in his closet as shown.

If Mark picks a shirt out of the closet without looking, which two colors have the greatest chance of being picked?

A. Blue and purple
B. Green and blue
C. Red and blue
D. Red and green

3
1 of 2 score points

1. Al, Bev, and Carmen are going on a ride at the park. Only 2 people can go on the ride at a time. They can pair up 3 different ways, as shown below.

Al and Bev

Al and Carmen

Bev and Carmen

Derek decides to join the group. How many different ways can the 4 students pair up?

Answer: _____

Show your work or explain how you got your answer.

You will have rubrics for constructed response items

4

1. Which statement is true about all four shapes shown?

A. Each shape is a rectangle.
B. Each shape is a quadrilateral.
C. Each shape has two pairs of parallel sides.
D. Each shape has one or more right angles.

7
2 of 2 score points

1. Al, Bev, and Carmen are going on a ride at the park. Only 2 people can go on the ride at a time. They can pair up 3 different ways, as shown below.

Al and Bev

Al and Carmen

Bev and Carmen

Derek decides to join the group. How many different ways can the 4 students pair up?

Answer: _____

Show your work or explain how you got your answer.

The Bookmark Process

- Study content in a book
- Write a story about the content
- Set a Bookmark
  - Separate the content students should know to be in Level 3 from the content that is more than enough

Bookmark Standard Setting

Round 1
- Take test
- Discuss OIB
- Discuss the target student
- Place Bookmarks

Round 2
- Discuss Round 1 Bookmarks
- Place Bookmarks

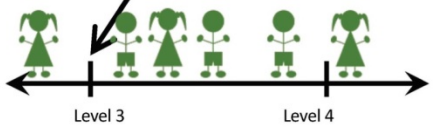
Round 3
- Present Impact Data
- Place Final Bookmarks

Questions?

Bookmark Training

Target Student for Level 3

Describes the knowledge and skills of the student entering the Level 3 Achievement Level

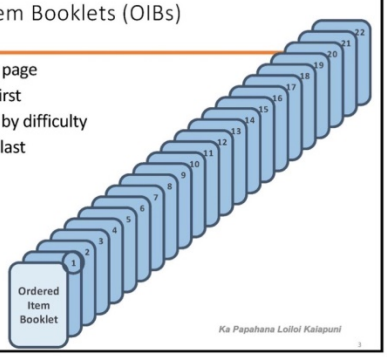


EDMETRIC

Ka Papahana Loiloi Kaiapuni

Ordered Item Booklets (OIBs)

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



EDMETRIC

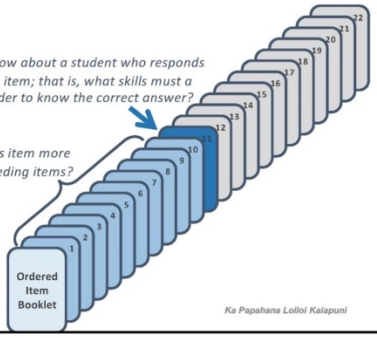
Ka Papahana Loiloi Kaiapuni

Study the Items in the OIB

As you study each item in the OIB, discuss two questions with your fellow panelists:

1. What do you know about a student who responds successfully to this item; that is, what skills must a student have in order to know the correct answer?

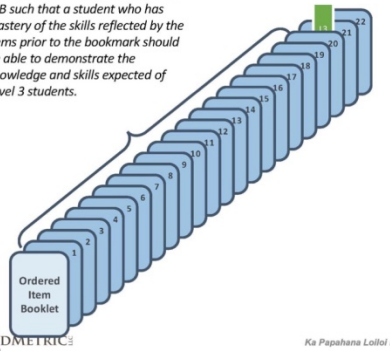
2. What makes this item more difficult than preceding items?



EDMETRIC

Ka Papahana Loiloi Kaiapuni

Place your Level 3 bookmark in the OIB such that a student who has mastery of the skills reflected by the items prior to the bookmark should be able to demonstrate the knowledge and skills expected of Level 3 students.



EDMETRIC

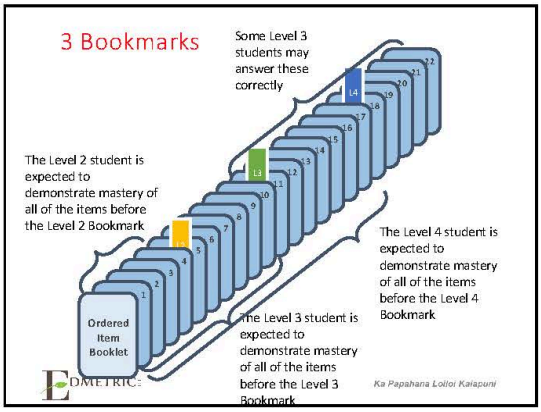
Ka Papahana Loiloi Kaiapuni

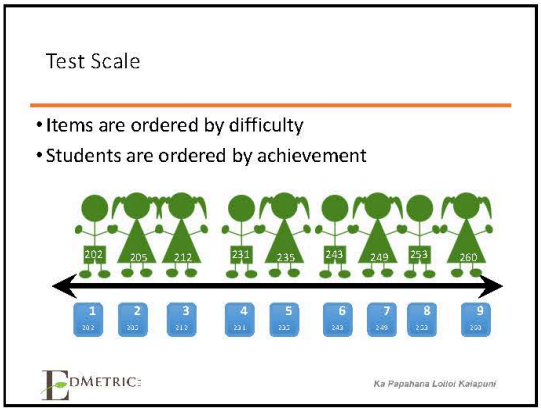
Cut Scores

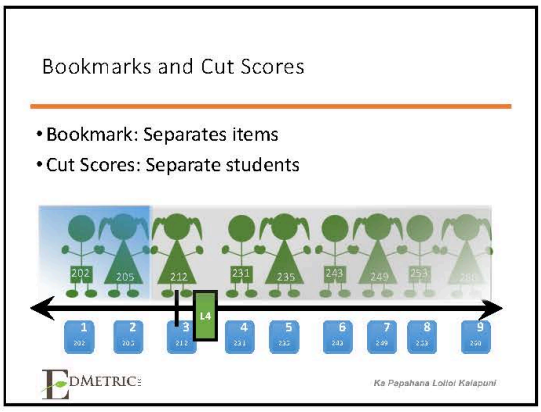
- Three cut scores
  - Level 2, Level 3, Level 4
- Four achievement levels
  - Level 1, Level 2, Level 3, Level 4

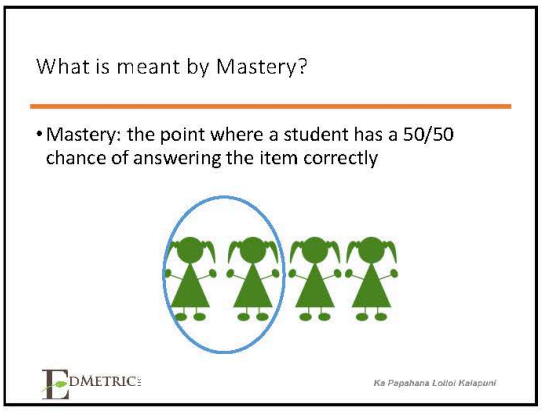
EDMETRIC

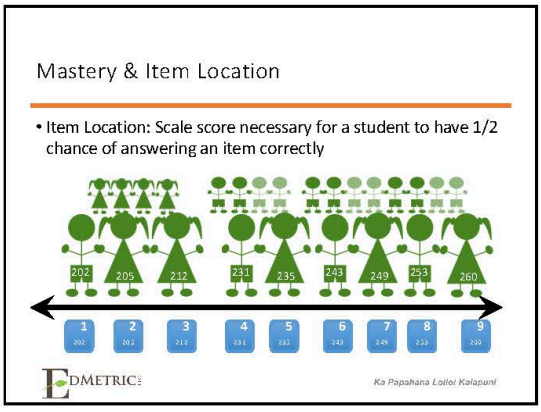
Ka Papahana Loiloi Kaiapuni

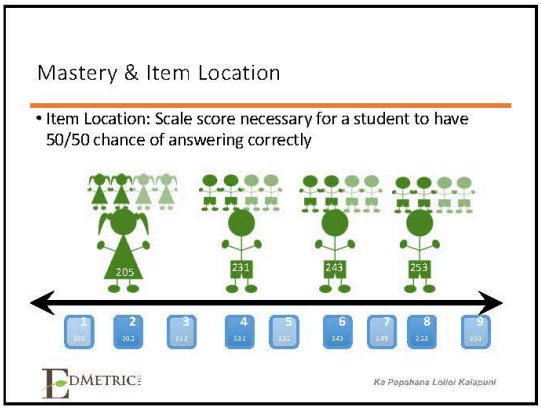

3 Bookmarks
Some Level 3 students may answer these correctly
The Level 2 student is expected to demonstrate mastery of all of the items before the Level 2 Bookmark
The Level 4 student is expected to demonstrate mastery of all of the items before the Level 4 Bookmark
The Level 3 student is expected to demonstrate mastery of all of the items before the Level 3 Bookmark
Ordered Item Booklet
EDMETRIC
Ka Papahana Loiloi Kalapuni


Test Scale
• Items are ordered by difficulty
• Students are ordered by achievement
202 205 212 231 235 243 249 253 260
1 2 3 4 5 6 7 8 9
EDMETRIC
Ka Papahana Loiloi Kalapuni


Bookmarks and Cut Scores
• Bookmark: Separates items
• Cut Scores: Separate students
202 205 212 231 235 243 249 253
L4
1 2 3 4 5 6 7 8 9
EDMETRIC
Ka Papahana Loiloi Kalapuni

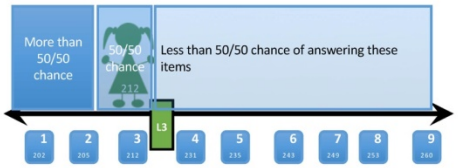

What is meant by Mastery?
• Mastery: the point where a student has a 50/50 chance of answering the item correctly
EDMETRIC
Ka Papahana Loiloi Kalapuni


Mastery & Item Location
• Item Location: Scale score necessary for a student to have 1/2 chance of answering an item correctly
202 205 212 231 235 243 249 253 260
1 2 3 4 5 6 7 8 9
EDMETRIC
Ka Papahana Loiloi Kalapuni


Mastery & Item Location
• Item Location: Scale score necessary for a student to have 50/50 chance of answering correctly
205 231 243 253
1 2 3 4 5 6 7 8 9
EDMETRIC
Ka Papahana Loiloi Kalapuni

Mastery and the Target Student

- Target Student: 50/50 chance of correctly answering the item just before the Bookmark

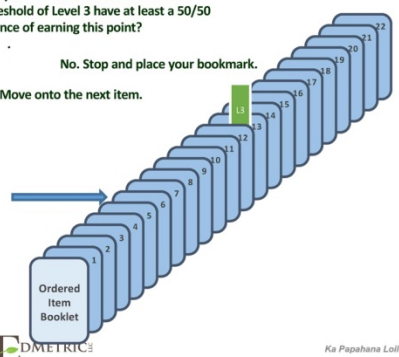


EDMETRICS Ka Papahana Loiloi Kaiapuni

Ask yourself: Would a student at the threshold of Level 3 have at least a 50/50 chance of earning this point?

No. Stop and place your bookmark.

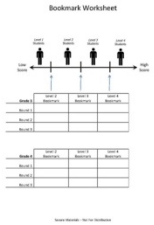
Yes. Move onto the next item.



EDMETRICS Ka Papahana Loiloi Kaiapuni

Bookmark Placement

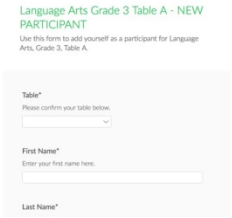
- Individual activity
- Use worksheet
- Enter results
  - Find your Table Links on Google Drive



EDMETRICS Ka Papahana Loiloi Kaiapuni

Bookmark Placement

- Individual activity
- Use worksheet
- Enter results
  - Find your Table Links on Google Drive



EDMETRICS Ka Papahana Loiloi Kaiapuni

Round 2

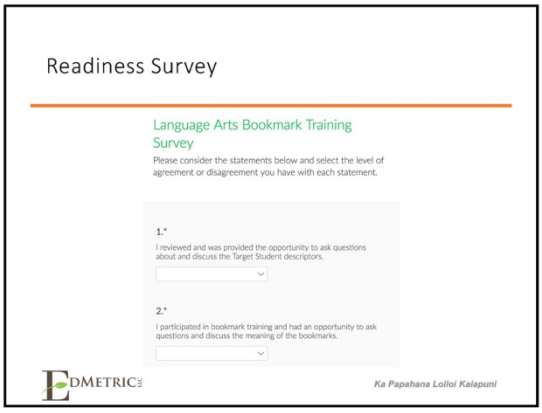
- *Primary Activities*
- *View Round 1 results*
- *Discuss Round 1 results at your table*
- *Place Round 2 Bookmarks*

EDMETRICS Ka Papahana Loiloi Kaiapuni

Questions?

EDMETRICS Ka Papahana Loiloi Kaiapuni

Readiness Survey



Language Arts Bookmark Training Survey

Please consider the statements below and select the level of agreement or disagreement you have with each statement.

1.*

I reviewed and was provided the opportunity to ask questions about and discuss the Target Student descriptors.

2.*

I participated in bookmark training and had an opportunity to ask questions and discuss the meaning of the bookmarks.

EDMETRICS

Ka Papahana Loiloi Kaiapuni

Ka Papahana Loiloi Kaiapuni

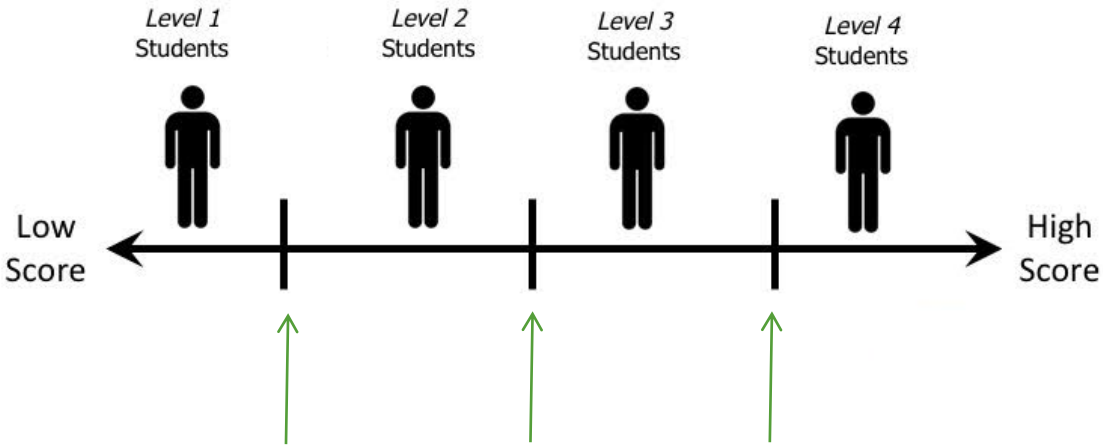
Standard Setting High-Level Agenda[1]

Saturday, July 23: Day 1	
Saturday	**8:30 AM:** Table Facilitator Meeting **9:00 AM:** Opening Session: Welcome and Training **10:00 AM:** Table-level Introductions and Secure Materials Sign-out **10:15 AM:** Q & A for the content standards **10:30 AM:** Break **10:45 AM:** Complete Grade 3 operational form **11:00 AM:** Study ordered item booklet (OIB) **Noon:** Lunch **1:00 PM:** Continue study of the OIB **4:00 PM:** Session Close
Sunday, July 24, 2016: Day 2	
Sunday	**9:00 AM**: UH Staff review Target Student Descriptors **9:30 AM:** Bookmark Training **10:30 AM:** Round 1 Ratings & Break **11:00 AM:** Round 2 **Noon**: Lunch **1:00 PM:** Round 3

[1] *Note: Times are approximate and will be adjusted as needed. Appropriate breaks will be provided throughout.*

	2:00 PM: Review of Grade 4 online operational form **2:30 PM:** OIB Review **3:30 PM:** Break **3:45 PM:** Round 1 Ratings **4:00 PM:** Secure materials collection.
Monday, July 25: Day 3	
Monday	**9:00 AM:** Discussion of Round 1 results and Round 2 ratings **10:30 AM:** Discussion of Round 2 results and Round 3 ratings **11:45 AM:** Final workshop evaluation **Noon:** Lunch **1:00 PM:** Create achievement level descriptors **3:00 PM:** Break **3:50 PM:** Secure materials collection. **4:00 PM**: Close

Bookmark Worksheet



Grade 3	Level 2 Bookmark	Level 3 Bookmark	Level 4 Bookmark
Round 1			
Round 2			
Round 3			

Grade 4	Level 2	Level 3	Level 4

	Bookmark	Bookmark	Bookmark
Round 1			
Round 2			
Round 3			

Kaiapuni Assessment Standard Setting

Post-Opening Session Readiness Survey

Please *consider the statements below* and fill in the bubble for the level of agreement or disagreement you have with each statement. Please bubble *only one* of the four options for each statement.	Strongly Disagree	Disagree	Agree	Strongly Agree
1. The orientation session provided a clear overview of the standard setting process.	①	②	③	④
2. I understand the goals of the standard setting workshop.	①	②	③	④
3. I understand my role in the standard setting workshop.	①	②	③	④
4. The orientation session provided a clear explanation of the development of Kaiapuni assessment	①	②	③	④
5. I understand how the results of the standard setting will be used to support the reporting of Kaiapuni assessment results.	①	②	③	④
6. I understand how to study the items in the ordered item booklet.	①	②	③	④

If you answered Disagree or Strongly Disagree to any of questions 1-8, then please answer the next two Yes/No questions.	No	Yes

7. I would like additional training on studying the ordered item booklet.	①	②
8. I have additional questions on material presented during the opening session that I would like answered before I begin the next task.	①	②

Kaiapuni Assessment Standard Setting

Post-Bookmark Training Readiness Survey

Please *consider the statements below* and fill in the bubble for the level of agreement or disagreement you have with each statement. Please bubble *only one* of the four options for each statement.	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I reviewed and was provided the opportunity to ask questions about and discuss the Target Student descriptors.	①	②	③	④
2. I participated in bookmark training and had an opportunity to ask questions and discuss the meaning of the bookmarks.	①	②	③	④
3. I understand how to place my bookmarks.	①	②	③	④
4. I understand I will have opportunities to change my bookmarks in Rounds 2 and 3.	①	②	③	④

If you answered Disagree or Strongly Disagree to any of questions 1-4, then please answer the next two Yes/No questions.	No	Yes
5. I would like additional training on placing my bookmarks for Round 1.	①	②
6. I have additional questions that I would like to ask before placing my Round 1 bookmarks.	①	②

Kaiapuni Assessment Standard Setting

Evaluation for Grade 3

Please *consider the statements below* and fill in the bubble for the level of agreement or disagreement you have with each statement. Please bubble *only one* of the four options for each statement.	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking.	①	②	③	④
2. The training materials were helpful.	①	②	③	④
3. Taking the student test was helpful and informative.	①	②	③	④
4. My group shared a common understanding of the Target Students.	①	②	③	④
5. Discussing the Target Students helped me place my bookmarks.	①	②	③	④
6. During Round 1, I placed my bookmarks independently.	①	②	③	④
7. I considered the Kaiapuni Standards when I placed my bookmarks.	①	②	③	④
8. The policy definitions were clearly communicated.	①	②	③	④
9. I understood how to place my bookmarks.	①	②	③	④
10. I had enough time to consider my bookmark placement.	①	②	③	④
11. I feel the recommended standards that resulted from this process are reasonable.	①	②	③	④

12. The impact data helped me evaluate my group's final bookmarks.	①	②	③	④
13. I understood how to interpret the impact data.	①	②	③	④
14. The impact data influenced where I placed my final bookmarks.	①	②	③	④
15. I would defend the panel's recommended Level 3 cut scores against criticism that they are too high.	①	②	③	④
16. I would defend the panel's recommended Level 3 cut scores against criticism that they are too low.	①	②	③	④
17. I would defend the panel's recommended Level 2 cut scores against criticism that they are too high.	①	②	③	④
18. I would defend the panel's recommended Level 2 cut scores against criticism that they are too low.	①	②	③	④

Kaiapuni Assessment Standard Setting

Evaluation for Grade 4

Please *consider the statements below* and fill in the bubble for the level of agreement or disagreement you have with each statement. Please bubble *only one* of the four options for each statement.	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking.	①	②	③	④
2. The training materials were helpful.	①	②	③	④
3. Taking the student test was helpful and informative.	①	②	③	④
4. My group shared a common understanding of the Target Students.	①	②	③	④
5. Discussing the Target Students helped me place my bookmarks.	①	②	③	④
6. During Round 1, I placed my bookmarks independently.	①	②	③	④
7. I considered the Kaiapuni Standards when I placed my bookmarks.	①	②	③	④
8. The policy definitions were clearly communicated.	①	②	③	④
9. I understood how to place my bookmarks.	①	②	③	④
10. I had enough time to consider my bookmark placement.	①	②	③	④
11. I feel the recommended standards that resulted from this process are reasonable.	①	②	③	④

Please *consider the statements below* and fill in the bubble for the level of agreement or disagreement you have with each statement. Please bubble *only one* of the four options for each statement.	Strongly Disagree	Disagree	Agree	Strongly Agree
12. The impact data helped me evaluate my group's final bookmarks.	①	②	③	④
13. I understood how to interpret the impact data.	①	②	③	④
14. The impact data influenced where I placed my final bookmarks.	①	②	③	④
15. I would defend the panel's recommended Level 3 cut scores against criticism that they are too high.	①	②	③	④
16. I would defend the panel's recommended Level 3 cut scores against criticism that they are too low.	①	②	③	④
17. I would defend the panel's recommended Level 2 cut scores against criticism that they are too high.	①	②	③	④
18. I would defend the panel's recommended Level 2 cut scores against criticism that they are too low.	①	②	③	④
19. I feel that my grade group as a whole is credible.	①	②	③	④
20. Overall, I believe that my opinions were considered and valued by my group.	①	②	③	④
21. Overall, I valued the workshop as a professional development experience.	①	②	③	④
22. This experience will help me target instruction for the students in my classroom.	①	②	③	④
23. The food and service at the facility met my expectations.	①	②	③	④

Please *consider the statements below* and fill in the bubble for the level of agreement or disagreement you have with each statement. Please bubble *only one* of the four options for each statement.	Strongly Disagree	Disagree	Agree	Strongly Agree
24. The workspace had accommodations appropriate to facilitate our work.	①	②	③	④
25. Participating in the workshop increased my understanding of the Kaiapuni assessment.	①	②	③	④
26. The workshop was well organized.	①	②	③	④

27. What is your current profession?

○ Classroom Teacher
○ School Administrator
○ Higher Education Faculty
○ Other (please specify)

_____

28. How many years have you been in your current profession?

_____

29. Please check all of the following in which you have experience:

○ Special education
○ English language learners
○ Hawaiian language learners
○ Vocational education
○ Adult education
○ Assessment
○ Educational policy

30. In which grade did you work during the workshop?

○ K
○ 1
○ 3
O 5
O 7
O HS

31. What is your gender?

○ Male O Female

32. Are you of Hispanic origin?

○ Yes O No

33. What is your race?

○ Asian/Pacific Islander
○ Hawaiian/Part Hawaiian
○ Black/African American
○ American Indian
○ White
○ Multi-racial

34. Your turn. Do you have any additional comments or thoughts about the workshop?

APPENDIX C. TABLE FACILITATOR MATERIALS

STANDARD SETTING OVERVIEW FOR TABLE FACILITATORS

Thank you for agreeing to serve as a Table Facilitator for the upcoming HLA and Mathematics Standard Setting Workshop. This overview will help you understand a bit about the standard setting process.

What is Standard Setting and Why is it Important?

Educators use assessments every day to determine what their students know and are able to do. Most often, the indicator of how a student is doing – usually a grade in a course or on a test – is determined based on a percentage of points the student earns. Just like educators in the classroom, states use assessments to inform stakeholders about student performance and progress. These assessments, often called large-scale assessments, are based on specific content standards that describe what students should know and be able to do at a specific grade level and in a particular content area. Large-scale assessments are carefully developed to assess students with a wide range of abilities over a very broad content domain. They often include several different item types, as well as items of varying difficulty.

The complexity of large-scale assessments makes it difficult to rely upon a percentage of points earned as an indicator of how well students are doing. Instead, most large-scale assessments transform a student's raw score (or the number of points earned) on the assessment into a scale score. Scale scores are estimated by considering such factors as the difficulty of the items a student answered correctly, and they allow students' scores on the assessment to be compared meaningfully. What scale scores do NOT do, however, is provide a meaningful indicator of student proficiency—that is the goal of standard setting. Standard setting is the process of reviewing content standards and the associated assessment items. Experts, like yourselves, decide how much students should know and be able to do in order to be considered "proficient" in a given grade level/content area, or to meet a specified performance level. This process allows us to determine "how much is good enough".

Standard setting is important because it provides content-based meaning to a numerical test score, and allows for a connection between the idea of "proficiency" to specific content expectations. It is the process that allows educators and stakeholders alike to move from a test score to a definition of the content represented by that test score. For educators, the descriptions of content associated with test scores resulting from standard setting allow assessments to logically impact instruction and vice versa. For other stakeholders, content-based test scores provide a meaningful picture of what students know and can do, and how that content is connected to real world applications.

Who Participates in Standard Setting?

- **Panelists --** The most important participants in standard setting are panels of content experts who have experience with diverse groups of students. Panels typically include classroom teachers for the appropriate grade level(s), curriculum

experts, and educators who work with students with disabilities and second language learners. Depending upon the grade level being assessed, and the purpose of the assessment, panels may also include higher education faculty and/or members of the business community. Panels are often selected to be demographically representative of a state in which the assessment is administered. The expertise of the panels is critically important to setting valid and defensible standards based on the content of the assessment.

- **Table Facilitators –** One of the panelists at each table is identified as a Table Facilitator. Table Facilitators are assigned to a table of panelists and assist in facilitating discussion of assessment items at the table, ensuring that the panelists stay focused on the standard setting task, making certain that all participants have an opportunity to express their opinion, and moving the standard setting process forward within the specified timeframes of the workshop. Table Facilitators also serve as liaisons between the panelists and the Workshop Facilitators.

- **Workshop Facilitators** – Workshop Facilitators provide training and expertise regarding the standard setting process itself.

- **Policy and Assessment Development Specialists** – Policy and assessment development specialists (usually representatives of a State Education Department and/or assessment development entity) serve as resources regarding content standards and policies regarding use of the assessments and their results.

- **Psychometric Specialists –** Psychometric specialists work "behind the scenes" during standard setting to determine the specific scale scores associated with the test content.

What are the Tools of the Bookmark Standard Setting?

Standard setting is conducted using one of several well-documented methodologies. For this standard setting, we will be using a standard setting methodology called the Bookmark Standard Setting Procedure (BSSP). It requires the following tools:

- **Ordered Item Booklets (OIB)** – The OIB is a set of test items selected to be representative of the content measured by the assessment. One item is presented on each page in ascending order of difficulty, based on actual student responses to the items.

- **Item Maps** – An item map is simply a columnar chart for each grade level that includes the OIB page number for each item, the location of the item on the score scale, the score point associated with each item, the item identification number, and the item type. Panelists will provide qualitative information about each item that will be entered into the Item Map during the standard setting workshop.

- **Content Standards** –The content standards provide the framework for the standard setting process. Panelists should familiarize themselves with the content standards prior to the workshop.

- **Achievement Levels and/or Achievement Level Descriptors** – Achievement levels provide a classification system for student performance, and achievement level descriptors provide a qualitative description of what a student at a particular performance level should know and be able to do. Often, these are drafted prior to standard setting, and panelists may revisit them and suggest revisions when standard setting is complete.

How are Standards Set?

To complete the BSSP, panelists will first discuss the content standards to provide a framework for the standard setting process. Next, they will take the operational assessment, just as a student would. After experiencing the assessment through the eyes of a student, participants will study the OIB for the grade level and content area. As the Table Facilitator, you will facilitate this discussion, focusing on the content of each item using the guidelines provided by your Workshop Facilitators. Then, panelists will discuss the characteristics of a student who is at each performance level – the Target Student Descriptors. Finally, the panelists will complete three successive rounds of judgments, placing a "bookmark" within the OIB to delineate between each of the performance levels, based on the content of the items. After each round of judgments, the "behind the scenes" psychometric specialists will determine the scale score associated with the bookmark placements in the OIB, and will provide the median score for the panelists at each table. These scores will represent the "cut scores" between the performance levels – the score a student has to attain in order to be classified in a particular performance level. After each round of judgments, panelists will have an opportunity to discuss their bookmark placements and make adjustments as they see fit. Panelists do not need to reach consensus on their bookmark placements. Additionally, after the second and third rounds of judgments, panelists will have an opportunity to see the percent of students whose scores fall within each performance level based on the cut scores they have recommended – this is called "impact data". Standard setting will culminate with all panelists for the grade level reviewing the median cut scores for the group as a whole to provide a final cut score recommendation.

What Happens After Standard Setting?

The work that you do over the course of the three-day workshop will be critical to the assessment system. We will use these performance levels when we submit data to the Hawaii Department of Education. The standards will also be used to communicate with teachers, parents, and the community about what our students know and can do.

Thank you so much for your help and dedication to the project!

KAIAPUNI ASSESSMENT OF EDUCATIONAL OUTCOMES

GRADE 3 & 4

LANGUAGE ARTS AND MATHEMATICS

Karla Egan

EdMetric, LLC

Note: Times are approximate and will be adjusted as needed. Breaks will be provided throughout.

SATURDAY MORNING, JULY 23, 2016[2]

Morning agenda and goals:

- Introductions
- Distribute and account for secure materials on the forms provided
- Identify a volunteer scribe to take notes on the computer for your table
- Facilitate discussion (Q&A) of the content standards
- Review the operational assessment online. This activity is intended to allow panelists to experience the assessment as a student would.
- Study the ordered item booklet (OIB) at your table. This is intended to help panelists attain a deeper understanding of the content the test measures and the relative difficulty of the items and types.

What you need:

- Panelist materials assignment and tracking form
- Content Standards
- Secure materials packets (printed copies of Target Student Descriptors, OIBs, item maps, and HLA stimulus booklets)
- Your laptop
- Secure thumb drive

Schedule and Description of Activities and Roles

9:00 AM: Opening Session

[2] *Note: Times are approximate and will be adjusted as needed. Appropriate breaks will be provided throughout.*

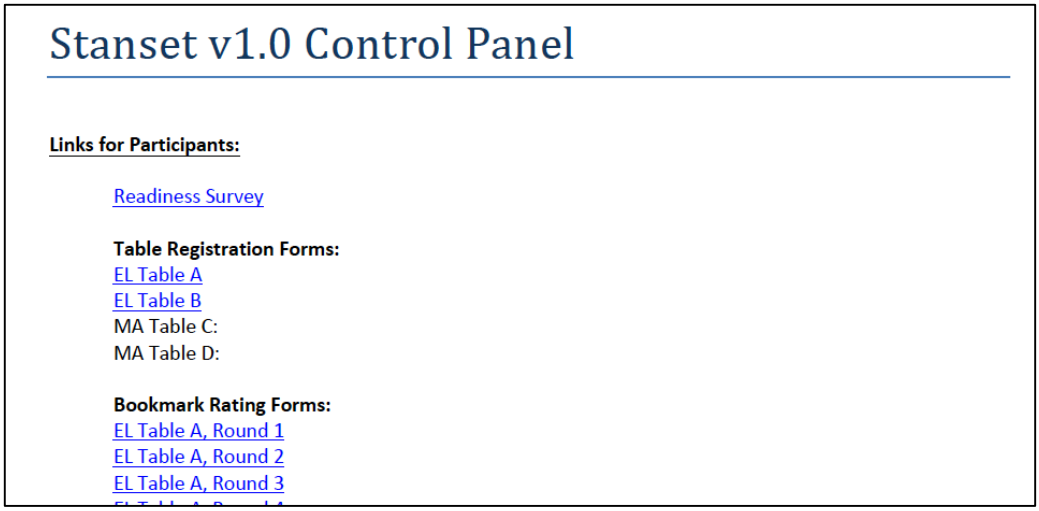


Figure 1. Portion of Standard Setting Control Panel PDF

10:00 AM: Introductions, readiness surveys, and secure materials sign-out

- Readiness surveys. Once panelists are at the table, ask them to complete the readiness survey. They should open the PDF entitled Standard Setting Control Panel using their laptops. This PDF file may be found on the secure thumb drive. They should follow the first link to their readiness survey.
- A packet of secure materials will be provided for each panelist. Each secure piece is color coded and has a Panelist ID number in the top right hand corner. The Table Facilitators should assign secure packets so that they are consecutively numbered at your tables (e.g., Table A assigns sets 1-4 and Table B assigns sets 5-8). Have panelists sign and print their names on the materials tracking form by the ID number of the materials they receive.
- Have panelists put their names on each piece of secure material.
- If necessary, remind panelists to put their cell phones away.
- Introduce yourself and facilitate brief introductions around the table. There will be many opportunities to get to know your fellow panelists so keep initial introductions brief (modeling this with your own brief introduction). One minute per panelist is a good target.
- Ask for a volunteer to act as scribe at the table. The scribe will take notes on the computer about items as panelists study the OIB. You may choose to be the scribe, but you will need to facilitate conversation so it is probably better for someone else to take that on.

10:15 AM: Q & A for the Standards. The Workshop Facilitators and University of Hawaii Staff will direct panelists to the appropriate materials and open the floor up for

questions. This Q & A is intended to answer any questions panelists may have regarding the content standards.

10:30 AM: Break

10:45 AM: Review of Grade 3 online operational form. The UH Staff will demonstrate how to access the online operational forms on your computer. This activity is intended to provide an opportunity to see the assessment in its online form as a student sees it.

- Follow the prompts and let UH Staff know if you have challenges logging in. Move through each item as a student would; perhaps a bit quicker. For example, there is an extended writing task and we don't have an expectation that you write as essay as part of your work.
- During the review, remind panelists that you will have many opportunities to discuss these items. Use this opportunity to experience the assessment through the eyes of a student and not as a means to prompt discussion.

11:00 AM: OIB Review

- Upon return from break, begin studying the OIB.
- Panelists should have their OIBs, item maps, and stimulus booklets out.
- Ask the scribe to open the item map so it appears on the monitor and take notes with input from the table.
- Panelists turn to the first item in their OIBs and locate it on their item maps. Review the item:
    - Prompt with the first question: *What does this item measure? That is, what do you know about the knowledge and skills of a student who responds successfully to this item?*
        - Notes should be informative, succinct responses.
    - Table Facilitator asks the second question: *Why is this item more difficult than the preceding items?*
        - Scribe takes notes. Please remind panelists that this should be a succinct set of notes on each item.
- When discussion ceases to provide new information, move to the next question. Encourage participation from all panelists.
- Repeat for all items in the OIB.
- *Considerations:*
    - Remember, you will see multiple-choice items only once in the OIB. You will see 1-point constructed response items only once in the OIB. You will see items with a maximum score higher than 1 multiple times, once for each non-

zero score point. So, for example, an item scored from 0-2 appears two times. Multiple score points for a single constructed response item may not appear consecutively in the ordered item booklet. For example, you may see the constructed response item for score point 1, followed by several multiple-choice items, followed by the same constructed response item for score point 2. You will be provided with scoring rubrics to help focus your discussion of each score point.

- The first time you see the item it is for score point 1 (e.g., out of 2). Ask and answer the question: What do you know about the knowledge and skills of a student who scores a 1 on this 2-point item?
- The second time you see the item it is for score point 2 (out of 2). Ask and answer the question: What do you know about the knowledge and skills of a student who scores a 2 (a perfect score) on this 2-point item?

- Stimulus booklet (HLA only). Longer stimuli that are common to multiple items are presented in a separate booklet to maintain the flow of the OIB. This booklet is ordered by the title of the passage, which can be found on each associated page in the OIB. To view an item's stimulus, identify the passage title in the OIB and find that passage in the stimulus booklet.
- Monitor time. Encourage thoughtful discussion, and do not rush through items. Remember, you will be discussing the items in this activity, and again after Rounds 1 and 2 before your final round of ratings. You *do not* need to exhaustively dissect every item. It is important to the validity of the process that panelists have enough time and don't feel rushed; however, it is also important to complete the entire process for both grade levels in our 3-day workshop.

- Support equitable and diverse discussion by encouraging all panelists to contribute to the discussion.
  - Caveats
    - Do not spend time critiquing items. While this is natural, this is not an item review workshop. If panelists feel the need to provide item input, have them write their concerns on an index card provided at the table for UH staff review.

Noon: Lunch

- All secure materials are color-coded. Be sure all color-coded materials are left in the room. Panelists should never remove these materials from the room.

For help:

- If you have questions about content, call the UH Staff to your table.
- If you have questions about process or policy, call the Workshop Facilitator to your table. Policy questions will be answered by UH staff, but will be collected by the Workshop Facilitators and answered promptly if vital to proceed, or in a group setting if not vital and if the answer will benefit all panelists.

SATURDAY AFTERNOON, JULY 23, 2016

Afternoon agenda and Goals:

- Review Target Student Descriptors
- Bookmark Training: Receive training to make Bookmark recommendations
- Round 1: Panelists make their first cut score recommendations
- Support systematic collection of secure materials

What you need:

- OIBs, item maps, stimulus booklets
- Target Student Descriptors
- Rating forms
- Readiness surveys
- Post-Its (bookmarks)
- Secure materials collection checklist
- Laptops
- Secure thumbdrives

Schedule and Description of Activities and Roles:

1:00 PM: Continue Study of OIB

3:50 PM: Secure materials collection.

- Table Facilitators follow systematic secure materials collection as described by Workshop Facilitators

4:00 PM: Daily Debrief

- The Table Facilitators are asked to attend the first part of the daily debrief in order to provide feedback:
    - Challenges that should be shared
    - What went well and what could be improved
    - Share any useful information associated with the quality of the workshop, table dynamics, etc.

SUNDAY, JULY 24, 2016

Morning agenda and goals:

- Round 1: Panelists will discuss target students, participate in Bookmark training, and set round 1 bookmarks.
- Round 2: Panelists discuss differences in cut score recommendations within their tables
- Round 3: Workshop Facilitators will share impact data with panelists. Impact data is the distribution of student scores within each level if cut scores were set based on the median of panelists' cut score recommendations following a round of judgments. Panelists review the reasonableness of Round 2 impact data and discuss differences between Table A and Table B cut score recommendations
- View final results
- Distribute, complete, and collect Grade 3 evaluations
- Support systematic collection of secure materials

What you need:

- OIBs, item maps, stimulus booklets
- Readiness surveys (online)
- Rating forms (online)
- Evaluations (online)
- Secure materials collection checklist

Schedule and Description of Activities and Roles

9:00 AM: UH Staff review Target Student Descriptors

9:30 AM: Workshop Facilitator provides Bookmark training

10:25 AM: Bookmark Readiness surveys. (The link will be available in the PDF entitled *Standard Setting Control Panel*)

10:30 AM: Round 1 ratings

- Panelists write their ratings on their paper rating forms independently and without discussion. If there are questions, ask a Workshop Facilitator.
- Ask panelists to enter their ratings into the online survey using the link for the Round 1 rating found in PDF entitled *Standard Setting Control Panel*.
    - Figure 1 shows a PDF form. All panelists will receive a PDF form with links in it to each Round's activities.

- Figure 2 shows the Online Rating Form.
- Have panelists take a short break after providing you their completed rating forms.

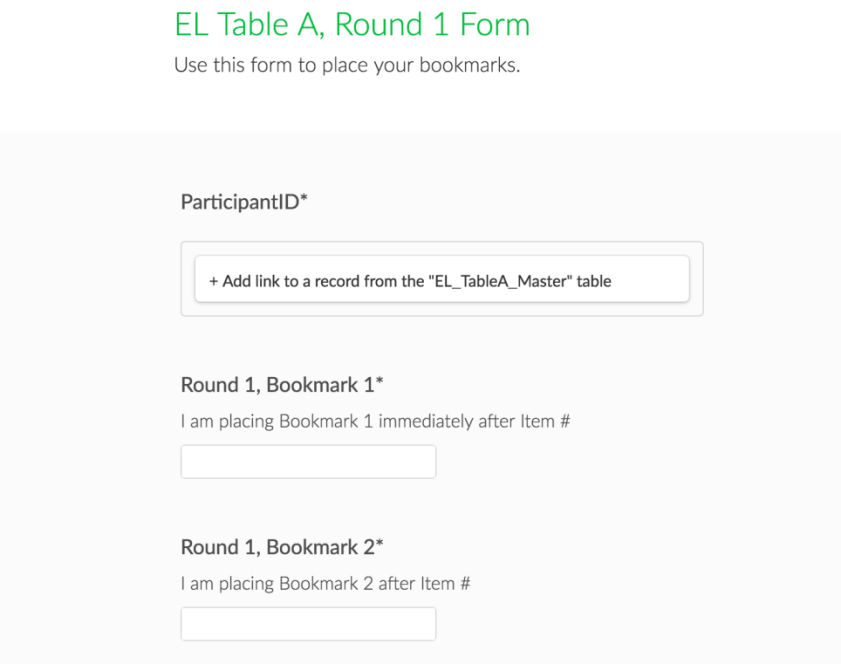


Figure 2. Online Rating Form

11:00 AM: Discussion of Round 1 results and Round 2 ratings

- Listen to the Workshop Facilitator Round 2 Orientation
- Follow steps 1-5.

1. Panelists place green post-its in the OIB representing their Level 3 bookmarks. We start with Level 3 because it is the "anchor level." It will serve as a reference point for Level 2 and Level 4 bookmark placements.
2. Discussion of Round 1 Level 3 bookmarks. Ask panelists to discuss the reasons for their various bookmark placements. One way to do this is to begin by having the panelists with the lowest and highest bookmarks discuss their rationales for their bookmark placement.
    a. Remind panelists that there are no wrong bookmark placements, just differences of opinion. Round 2 allows panelists to reflect on their own and others' bookmark placements.
    b. Remind panelists that they do not need to come to consensus on the placement of their bookmarks
3. Panelists have an opportunity to reset their Level 3 bookmark placements based on their updated perspective.
4. Repeat process for Levels 2 and 4.
5. Have panelists enter the Round 2 bookmarks using the online form.

Noon: Lunch

1:00 PM: Discussion of Round 2 results and Round 3 ratings

- The two tables join together as one group representing the grade. You no longer work as separate, independent tables.
- The Workshop Facilitator will introduce Round 3 activities, which include reviewing:
    - Each table's median bookmark
    - The grade's combined bookmark (median of all panelists at Table 1 and Table 2)
    - The impact data—the percentage of students in the various levels based on the grade's median bookmark
- Observe the impact data. If you are very surprised by the data, call a Workshop Facilitator to your table to help frame the discussion.
- As a single group (Tables 1 and 2 together) discuss the differences between the two tables' median bookmarks
    - Begin with Level 3
        - Discuss differences between each table's Level 3 Bookmarks.
    - Repeat for Levels 2 and 4
- Have panelists enter Round 3 ratings using the online form

1:50 PM: End of activities for Grade 3. Grade 3 Evaluations and collection of secure materials.

- Workshop Facilitators will initiate close of the session.
- Panelists complete Grade 3 evaluations using online form
- Table Facilitators collect evaluations and submit to Workshop Facilitators
- Table Facilitators follow systematic secure materials collection as described by Workshop Facilitators

Note that we have provided detailed directions for one grade level. We will begin with Grade 3. When we have completed all three rounds for Grade 3, we will proceed and repeat this process for Grade 4. Below times for the remaining activities are overviewed.

2:00 PM: Review of Grade 4 online operational form

2:30 PM: OIB Review

3:30 PM: Break

3:45 PM: Round 1 Ratings

4:00 PM: Secure materials collection.

4:15 PM: Daily Debrief

MONDAY, JULY 25, 2016

9:00 AM: Discussion of Round 1 results and Round 2 ratings

10:30 AM: Discussion of Round 2 results and Round 3 ratings

11:45 AM: Final workshop evaluation

Noon: Lunch

1:00 PM: Create achievement level descriptors

3:00 PM: Break

3:50 PM: Secure materials collection.

4:00 PM: Close

Appendix: Additional Details

STUDYING THE ORDERED ITEM BOOKLET

Panelists develop a comprehensive understanding of what each test measures by studying OIBs—a set of test items selected to be representative of the construct measured by each form of the test, ordered by difficulty.

The following three sets of materials are used to study the OIBs

1. Ordered Item Booklet. The OIB has one item per page with the easiest item first and the most difficult item last. Each page includes the item and information such as SLO measured by the item.
2. HLA Stimulus Booklet. In HLA, stimuli that are common to multiple items are presented in a separate booklet to maintain the flow of the OIB. This booklet is ordered by passage title, which can be found on each associated page in the OIB.
3. Item Map. Item maps support the review of the OIB for each grade level. The item maps include the following information, as illustrated in Figure 3:
    - Order of Difficulty: OIB Page number
    - Location: the scale score needed for a student to have a 50/50 chance of answering the item correctly
    - Score Point (1 for multiple choice, the score point considered for multipoint items)
    - Test Item #: the order in which the test was administered
    - Problem ID: the item's identification number
    - Story/Stimulus: Passage
    - Code
    - Content Classification
    - Item Type

Panelists at the standard setting workshop will complete the final two columns of the item map:

- What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?
- Why is this item or score point more difficult than the items that precede it?

Item Map - Mākau ʻŌlelo
Papa 4

Order of Difficulty (OIB Page Number)	Location	Score Point	Test Item #	Problem ID	Story/Stimulus	Code	Content Classification	Item Type	What does this item or score point measure? That is, what do you know about a student who responds successfully to this item or score point?	Why is this item or score point more difficult than the items that precede it?
1	363	1	4	7	Hō Aku Nō!	1.1	3/4O3	STV		
2	389	1	10	13	Nā Hoapili	1.2	3/4O7	SC		
3	402	1	12	14	Hawai'i Pono	2.1	3/4MO1	SC		
4	411	1	16	65	Hawai'i Pono	2.2	4MO1	SC		
5	420	1	33	38	-	3.3	4K18	SC		
6	431	1	6	64	Nā Hoapili	1.2	4O6	FTG		
7	432	1	1	18	Hō Aku Nō!	2.1	3/4MO1	SC		
8	436	1	31	27	-	3.1	4K4	SC		

Figure 3. Sample Item Map

The panelists discuss each item in the OIB. In particular, they will discuss the knowledge, skills, and processes being measured by each item as well as why the item is more difficult than the items that precede it. The designated table note taker will document these comments.

APPENDIX D. RESULTS OF PANELIST EVALUATIONS

Table D.1. Percentage of Standard Setting Panelists who Agree or Strongly Agree with Each Statement

Statement	Language Arts		Mathematics	
	Grade 3	Grade 4	Grade 3	Grade 4
I felt that this procedure was fair and allowed me to recommend cut scores that reflected my thinking.	100.0	100.0	100.0	100.0
The training materials were helpful.	100.0	100.0	92.3	100.0
Taking the student test was helpful and informative.	100.0	100.0	100.0	100.0
My group shared a common understanding of the Target Students.	100.0	100.0	92.3	100.0
Discussing the Target Students helped me place my bookmarks.	100.0	100.0	92.3	91.7
During Round 1, I placed my bookmarks independently.	100.0	100.0	100.0	100.0
I considered the Kaiapuni Standards when I placed my bookmarks.	92.3	100.0	100.0	100.0
The policy definitions were clearly communicated.	100.0	100.0	92.3	100.0
I understood how to place my bookmarks.	100.0	92.3	100.0	100.0
I had enough time to consider my bookmark placement.	100.0	100.0	100.0	91.7
I feel the recommended standards that resulted from this process are reasonable.	100.0	100.0	100.0	100.0
The impact data helped me evaluate final bookmarks.	100.0	100.0	100.0	100.0
I understood how to interpret the impact data.	100.0	100.0	100.0	100.0
The impact data influenced where I placed my final bookmarks.	76.9	69.2	84.6	83.3
I would defend the recommended Level 3 cut scores against criticism that they are too high.	84.6	92.3	100.0	100.0
I would defend the recommended Level 3 cut scores against criticism that they are too low.	69.2	92.3	100.0	100.0
I would defend the recommended Level 2 cut scores against criticism that they are too high.	69.2	84.6	100.0	100.0
I would defend the recommended Level 2 cut scores against criticism that they are too low.	76.9	100.0	100.0	100.0
I would defend the recommended Level 4 cut scores against criticism that they are too high.	83.3*	76.9	100.0	100.0
I would defend the recommended Level 4 cut scores against criticism that they are too low.	100.0*	92.3	100.0	100.0

Statement	Language Arts		Mathematics	
	Grade 3	Grade 4	Grade 3	Grade 4
I feel that my grade group as a whole is credible.**		100.0		100.0
Overall, I believe that my opinions were considered and valued by my group.**		100.0		100.0
Overall, I valued the workshop as a professional development experience.**		100.0		100.0
This experience will help me target instruction in my classroom.		100.0		100.0
The food and service at the facility met my expectations.**		100.0		100.0
The workspace had accommodations appropriate to facilitate our work.**		100.0		100.0
Participating in the workshop increased my understanding of the Kaiapuni assessment.**		100.0		100.0
The workshop was well organized. **		100.0		100.0

*Based on 12 panelist responses **Only asked on the final evaluation because the questions were relevant to the workshop as a whole, not a single grade level.

Table D.2. Number of Panelists disaggregated by Educator role and Content Area

Content Area	Educator Role	Frequency
Language Arts	Classroom Teacher	11
	Other	2
Mathematics	Classroom Teacher	12

Table D.3. Number of Panelists disaggregated by Gender and Content Area

Content Area	Gender	Frequency
Language Arts	Female	13
Mathematics	Female	11
	Male	1

Table D.4. Number of Panelists disaggregated by Grade Level Taught and Content Area

Content Area	Grade Level	Frequency
Language Arts	2nd	2
	3rd	2
	3rd & 4th	5
	4th	3
	M-6	1
Mathematics	2nd	1
	3rd	3
	3rd & 4th	4
	4th	4

Table D.5. Mean, Minimum, and Maximum Number of Years Panelists were in current position, disaggregated by Content Area

Content Area	N	Mean	Std. Deviation	Minimum	Maximum
Language Arts	13	7.46	4.77	1	15
Mathematics	12	12.08	7.62	3	30

Table D.6. Number of Panelists disaggregated by Island, School, and Content Area

Island and School	Language Arts	Mathematics	Total
Hawai'i	2	2	4
Ke Kula ʻO Nāwahīokalaniʻōpuʻu Iki LCPS	1		1
Ka 'Umeke Kā'eo	1		1
Ke kula ʻo ʻEhunuikaimalino		1	1
Ka ʻUmeke Kāʻeo PCS		1	1
Kauai	1		1
KAWAIKINI	1		1
Maui	3	2	5
Paia		1	1
Ke Kula Kaiapuni ʻO Nāhiʻenaʻena	1		1
Pāʻia School	1		1
Kula Kaiapuni ʻo Nāhiʻenaʻena	1		1
Pāʻia		1	1
Molokai	1	2	3
Kualapuu		2	2
Kula Kaiapuni o Kualapuʻu	1		1
Oahu	6	6	12
Anuenue		1	1
Hauula		1	1
Pūʻōhala	2	1	3
Hauʻula	1		1
Ke Kula Kaiapuni ʻo Waiau	1	1	2
Ke Kula Kaiapuni ʻo Nānākuli	1		1
Kula Kaiapuni ʻo Ānuenue	1		1
Ke Kula Kaiapuni o Pūʻōhala		1	1
Ke Kula Kaiapuni o Hauʻula		1	1
Total	**13**	**12**	**25**

Table D.7. Number of Panelists disaggregated by Island, School, and Content Area

Island and School	Language Arts	Mathematics	Total
Hawai'i	2	2	4
Ke Kula ʻO Nāwahīokalaniʻōpuʻu Iki LCPS	1		1
Ka 'Umeke Kā'eo	1	1	2
Ke kula ʻo ʻEhunuikaimalino		1	1
Kauai	1		1
KAWAIKINI	1		1
Maui	3	2	5
Ke Kula Kaiapuni ʻO Nāhiʻenaʻena	2		2
Pāʻia	1	2	3
Molokai	1	2	3
Kula Kaiapuni o Kualapuʻu	1	2	3
Oahu	6	6	12
Ke Kula Kaiapuni ʻo Waiau	1	1	2
Ke Kula Kaiapuni ʻo Nānākuli	1		1
Kula Kaiapuni ʻo Ānuenue	1	1	2
Ke Kula Kaiapuni o Pūʻōhala	2	2	4
Ke Kula Kaiapuni o Hauʻula	1	2	3
Total	**13**	**12**	**25**

Table D.8. Number of Panelists with Teaching Experience in Different Areas, by Content Area

Experience	Language Arts	Mathematics	Total
Hawaiian Language Learners	13	12	25
English Language Learners	1	5	6
Special Education	1	1	2
Policy	1	2	3
Assessment	5	4	9
Adult	3	0	3

Table D.9. Panelists' Comments by Content Area

Language Arts
I now have a clear understanding of how the process works as far as the assessment creation and putting it into place. Also an understanding of how hard the OHE works in partnership with UHM and the kaiapuni teachers statewide.
It gave me a much better understanding of everything going on with the testing situation. Mahalo nui!
It was very valuable. Mahalo nui
Mahalo
Mahalo for this opportunity, I learned a lot about the kaiapuni assessment.
Mahalo nui loa for your awesome leadership and facilitation.
Mahalo nui loa no kēia hoʻonuiʻike!
Mahalo nui!!!
mahalo!!
No
No.
No. Mahalo.
Thank you for bringing us together! It was wonderful to hear and learn from other Hawaiian Immersion/Medium educators. It helps me measure how we as a group are performing.
Mathematics
Everything was well thought out and planned
He hana nui nō, eia naʻe nui nā mea i aʻo ai e hoʻokele hoʻi ai i kaʻu hana ma ka papa.
I truly do mahalo this process it was very informative and systematic.
I would appreciate more training on implementing the standards and developing practice items for our students - How we can better prepare our lower grade students to be successful in language and math skills necessary to meet standards
Learned a lot. I appreciated the opportunity to meet with fellow kumu from various schools.
Mahalo for this workshop and letting us be part of the process.
Mahalo nui ʻia kēia ʻano hui ʻana o nā kumu a me nā kākoʻo ʻōlelo Hawaiʻi. Pono e hālāwai hou.
Mahalo nui iā ʻoukou no ke kono ʻana mai iaʻu i kēia papahana!
Mahalo nui loa.
Mahalo nui no kēia hālawai, ua aʻo au i nā mea he nui mai ia mau kumu kaiapuni ʻē aʻe. Mākaukau au e holomua a hoʻoikaika i ka ʻōlelo Hawaiʻi ma koʻu kula. I look forward to our next Standard Setting hālāwai!
The workshop coordinators and presenters helped to facilitate the process very well. I truly appreciate the process we used and feel that I can leave the workshop with a greater understanding of the assessment, of the process, and of tasks ahead.
This workshop was an amazing experience and was at the perfect time too! I would love to do this again.

APPENDIX E. DETAILED RESULTS OF THE STANDARD SETTING